Bayesian epidemic models for spatially aggregated count data

C. Malesios[1], N. Demiris[2], K. Kalogeropoulos[3] and I. Ntzoufras[2]

[1]*Department of Agricultural Development, Democritus University of Thrace, Greece, Email: malesios@agro.duth.gr*
[2]*Department of Statistics, Athens University of Economics and Business, Athens, Greece*
[3]*Department of Statistics, London School of Economics, London, UK*

Abstract

Epidemic data often possess certain characteristics, such as the presence of many zeros, the spatial nature of the disease spread mechanism, environmental noise, serial correlation and dependence on time varying factors. This paper addresses these issues via suitable Bayesian modelling. In doing so we utilise a general class of stochastic regression models appropriate for spatio-temporal count data with an excess number of zeros. The developed regression framework does incorporate serial correlation and time varying covariates through an Ornstein Uhlenbeck process formulation. In addition, we explore the effect of different priors, including default options and variations of mixtures of g-priors. The effect of different distance kernels for the epidemic model component is investigated. We proceed by developing branching process-based methods for testing scenarios for disease control, thus linking traditional epidemiological models with stochastic epidemic processes, useful in policy-focused decision making. The approach is illustrated with an application to a sheep pox dataset from the Evros region, Greece.

Keywords: Bayesian modelling; Bayesian variable selection; Branching process; Epidemic Extinction; g-prior; Spatial kernel; Disease control.

1 Introduction

Infectious disease outbreaks can have devastating consequences both from the societal and economic perspective. In the present paper we develop

suitable methodology for epidemic spatio-temporal data. Specifically, we extend previous analyses (e.g. [1]), by addressing several important aspects of epidemic modelling. In addition, we provide an epidemic interpretation of our modelling approach, thus linking epidemiological models with stochastic epidemic processes, useful in policy-focused decision making.

Our modelling framework accommodates a number of features: First, we make use of spatial information related to the location of the infected premises. Specifically, we incorporate the spatial coordinates of each infected farm, allowing for the probability of infection between farms to depend upon their distance. A number of spatial transmission kernels are fitted to the spatio-temporal data using Bayesian methods and an extended investigation of their relative importance is performed. Second, we model the inherent serial correlation of the data generating mechanism via a latent Ornstein Uhlenbeck (OU) process evolving around a mean which is regressed upon a number of covariates, including the spatial transmission kernels. The mean of the OU process is allowed to vary across time, adapting to the changes of its predictors. Third, we investigate one of the basic model selection problems in Bayesian regression-type modeling, namely the one of covariate selection. Various methods have been proposed in order to deal with this task. Here we tackle this issue by implementing recently developed variable selection approaches based on hyper-g prior distributions [2] using Gibbs-based variable selection (e.g. [3]). The covariate selection component appears to have received limited attention in the infectious disease literature. In addition to proposing a flexible continuous-time model, suitable for a wide variety of real surveillance data, we describe a simulation-based estimate of the extinction probability which could be of direct use to policy makers. For inference purposes, we generally resort to Markov chain Monte Carlo (MCMC) simulation methods which offer flexibility in the ability to fit complex models of the kind entertained in this paper.

Sheep pox is a highly contagious viral infection that can have devastating consequences [4]. The proposed methodology is applied to a historical dataset described in detail in section 4. To avoid uncertainties related to the exact day of infection - estimated by the local authorities from the clinical signs of the disease - we analyze the aggregated weekly counts of infection. The latter approach shares certain advantages;

a) We avoid part of the measurement error.

b) By creating the (artificial) extra zeros we can infer which factors (e.g.

environmental) assist in creating a disease-free environment through the covariates linked to the excess zeros.

c) Our model may be of use in a number of instances since data of this type are commonplace, for instance in typical surveillance systems.

Note however that the modeling framework provided in this paper is specified in continuous time and can be used even if the data are not aggregated. The continuous time formulation provides additional advantages such as the ability to forecast, interpret and aggregate on different time scales, utilise information observed at different frequencies, allow for delayed impact of time varying covariates etc. Moreover, for the specific application considered here and the case of aggregated counts, it is possible to extract a discrete-time model with complexity that is comparable to standard time series models. Details are provided in the next section.

The data are presented in detail in [5] where a preliminary analysis showed strong support for the Poisson and zero inflated Poisson (ZIP)-based models as opposed to those based on the negative binomial distribution. Hence, in this paper we focus upon the Poisson and ZIP specifications. The following section contains the model formulation while section 3 presents the connection of this class of models with stochastic epidemics, a useful tool for disease control. Section 4 illustrates the results of the analysis and section 5 concludes with some discussion.

2 Bayesian modelling

2.1 Spatio-temporal model

Our response variable consists of 260 weekly observations (infected farms) spanning over a time interval of 5 years. Time is measured in weeks and we let t_i, $i \in \{0, 1, \ldots, 259\}$ correspond to the time after the i-th week. The response vector $\mathbf{y} = \{y_i\,;\, i = 0, \ldots, 259\}$ is also ordered chronologically where each y_i denote the number of sheep pox cases at t_i.

In what follows, the modelling framework is first presented in continuous time and then, given that the data appear in weeks, we extract its form in discrete time. The fitted models are special cases of the specification scheme

described by:

$$\begin{cases} y_i & \sim g(y_i|\Lambda_i, p_i) \\ g(y_i|\Lambda_i, p_i) & = p_i I_{\{y_i=0\}} + (1-p_i) f(y_i|\Lambda_i) \\ \Lambda_i & = \int_{t_{i-1}}^{t_i} \exp(\lambda_s) ds \\ d\lambda_t & = \phi(\lambda_t - \mu_t)dt + dB_t \end{cases} \tag{1}$$

where B_t denotes standard Brownian motion, and μ_t is the piecewise constant deterministic process:

$$\mu_t = \begin{cases} \mu^{(0)} & \text{if } \ 0 \leq t < t_1 \\ \mu^{(1)} & \text{if } \ t_1 \leq t < t_2 \\ \vdots & \\ \mu^{(259)} & \text{if } \ t_{259} \leq t < t_{260} \end{cases}$$

where each $\mu^{(i)}$ corresponding to $t_i \leq t < t_{i+1}$, $i = 0, 1, \ldots, 259$, is given by

$$\mu^{(i)} \ = \ \mathbf{X}_{(i)}\beta + b_i + K\big(\mathbf{d}_i, \boldsymbol{\Theta}_K\big),$$

where $\mathbf{X}_{(i)}\beta + b_i$ is a component comprising of endemic/epidemic covariates and $K\big(\mathbf{d}_i, \boldsymbol{\Theta}_K\big)$ is the rate at which infection is transmitted from an infected farm to a susceptible farm, made precise in (5). Moreover, $I_{\{y_i=0\}}$ is an indicator variable denoting whether the response is positive or not, θ_i is the instantaneous rate of the process at the i-th time point, p_i denotes the proportion of excess zeros for t_i and β is a vector of dimension $\dim(\beta)$ with the coefficients of the fixed-effects covariates; the corresponding covariate values for t_i are denoted by the row vector $\mathbf{X}_{(i)}$, also of dimension $\dim(\beta)$. Note that the diffusion path integral $\int_{t_{i-1}}^{t_i} \exp(\lambda_s) ds$ does not have a closed form solution and therefore the model in (1) is intractable. One option is to adopt a data augmentation scheme, introducing a fine partition to the path of λ_t in the spirit of [6]. An alternative is to proceed via the following approximation:

$$\int_{t_{i-1}}^{t_i} \exp(\lambda_s) ds \approx (\exp(\lambda_{t_{i-1}}) + \exp(\lambda_{t_i}))\delta_i/2, \quad \delta_i = t_i - t_{i-1}.$$

The former option offers the benefit of controlling the approximation error due to time discretisation. Nevertheless, this comes at the expense of a

substantial increase in the computational cost and complexity of the implementation algorithm. Hence, we proceed by adopting the latter approximation.

In our dataset $\dim(\beta) = 10$, and therefore $\boldsymbol{\beta}=(\beta_0, \beta_1, \beta_2, \beta_3, ..., \beta_9)^T$ is the vector of regression coefficients for the intercept (β_0), the covariates describing the number of villages infected in the previous week (β_1), rainfall (β_2), average temperature (β_3), maximum temperature (β_4), minimum temperature (β_5), average humidity (β_6) and seasonal effects: spring (β_7), summer (β_8), autumn (β_9). In addition, $\boldsymbol{\Theta}_{\mathbf{K}}$ is the parameter vector of the transmission kernel function $K(\cdot)$ and b_i reflects independent yearly random effects with $\mathbf{b} = (b_0, b_1, b_2, b_3, b_4)^T$. Note that Poisson regression is recovered when $p_i = 0$ and f denotes the Poisson probability mass function.

The instantaneous λ_t is an OU process evolving around μ_t, which in turn is determined by the potentially time varying covariates $\mathbf{X}_{(i)}$. Its transition density is available in closed form allowing us to write (for all i)

$$\lambda_{t_{i+1}}|\lambda_{t_i} \sim N\left(\mu^{(i)} + \left(\lambda_{t_i} - \mu^{(i)}\right) e^{-\phi\delta_i},\ \frac{1 - e^{-2\phi\delta_i}}{2\phi}\right). \tag{2}$$

The OU process reflecting λ_t need not be stationary. In fact, every change in the covariates provides a shock to the system, to which the latent process λ_t adapts through a transient OU process with rate of convergence driven by the parameter ϕ. Large values of ϕ imply that the effect of the covariates on λ_t is imminent, whereas values close to zero point towards a delayed impact. This formulation is therefore substantially different than that of [1] and [7],under which the model would have been written as

$$\begin{cases} y_i & \sim g(y_i|\lambda_i, p_i) \\ g(y_i|\lambda_i, p_i) & = p_i I_{\{y_i=0\}} + (1 - p_i) f(y_i|\lambda_{t_i}) \\ \lambda_{t_i} & = \mu^{(i)} + \theta_1 \lambda_{t_{i-1}} + \epsilon_{t_i}, \end{cases} \tag{3}$$

where ϵ_{t_i} is modeled via a stationary, zero mean, OU process in the spirit of [8]. In fact, the model in (1) may actually be seen as the model used in [9], adapted to the context of this paper. Comparisons between the two approaches can be found in the supplementary material Table S3 supporting the non-stationary version for the particular application of this paper. Overall, this OU formulation has the following features:

a) Its mean reverting nature is essential in order to capture the effect of the several time-varying predictors, such as the spatial transmission kernels

and climate effects, and introduce serial correlation. For example an increase in the temperature will result in an increase of $\mu^{(i)}$ which, according to (2), is the mean of the intensity process as time increases. Hence, the intensity process will adapt to the increased temperature at a rate controlled by the parameter ϕ. Note that this is not the case for the model in (3) where $\mu^{(i)}$ is the constant in the AR(1) equation.

b) As δ_i or ϕ increases we get that $\lambda_{t_{i+1}}|\lambda_{t_i} \sim N\big(\mu^{(i)}, (2\phi)^{-1}\big)$, in other words the autoregressive part vanishes and the λ_{t_i}'s become independent with finite variance.

c) It introduces an additional source of variability that can be linked with the inherent environmental noise of the system or it can also absorb potential model mis-specification.

d) As it is based on a continuous time process, it can be used to different types of datasets; e.g. irregularly spaced observations.

e) Finally, incorporating the above aspects comes at no additional complexity in terms of computation or implementation since the presence of a tractable solution simplifies the calculations significantly. In fact, the amount of complexity is comparable to using an $AR(1)$ model.

An inherent feature of epidemic data - such as those analyzed here - is the excessive number of zero counts (excess zeros). This is often due to endemic conditions or diseases that are hard to detect. To account for this, we now complete our model formulation by using a similar structure for the zero-inflation probability p_i as the one for λ_i in (1), writing

$$\log\left(\frac{p_i}{1-p_i}\right) \;=\; \mathbf{X}_{(i)}\beta^z + b^z_{t_i} + K(\mathbf{d}_i, \mathbf{\Theta}^z_K)\;; \tag{4}$$

where the superscript z denotes the parameters used for modeling the zero-inflation probability with similar role as β, b_{t_i}, $\mathbf{\Theta}_K$, respectively. The latter equation for p_i involves a Kernel K, as was done with the positive response part, anticipating here a negative association with the probability of zero cases occurring as time increases from the latest infection.

2.1.1 Spatial kernels

The literature contains several attempts to capture the spatial structure of transmission of an animal disease, including foot-and-mouth disease (FMD) which had economically devastating consequences to livestock [10]. Thus, the use of simulation modeling for estimating the spread of highly contagious livestock diseases and for conducting risk assessment for various control measures has become common in recent years (e.g. [11], [12], [13]).

The $K\big(\mathbf{d}_i, \mathbf{\Theta}_K\big)$ term is an infection kernel used to model the spatial component of disease propagation, where $\mathbf{d}_i = \{\mathsf{d}_{k\ell} : k \in \mathcal{S}_i, \ell \in \mathcal{I}_{i-j}\}$, is the set of all the Euclidean distances between uninfected farms $k \in \mathcal{S}_i$ at time i and previously infected farms $\ell \in \mathcal{I}_{i-j}$ within the typical incubation period for sheep pox. Here, $\mathcal{S}_i$ and $\mathcal{I}_i$ denote, respectively, the set of susceptible and infected farms at time point i. The Euclidean distance $d_{k\ell}$ is calculated by $d_{k\ell} = \sqrt{(u_k - u_l)^2 + (v_k - v_l)^2}$ with (u_k, v_k) denoting the geographical coordinates of farm k measured in kms according to global positioning system (GPS). The geographical coordinates were then used for calculating the matrix containing all pairwise distances.

We model this risk factor as a parametric function assuming a 3-week incubation period [14] by including kernels, $K(\cdot)$, of the form:

$$K\big(\mathbf{d}_i, \mathbf{\Theta}_K\big) = \begin{cases} \frac{1}{|\mathbf{d}_i|} \sum\limits_{k \in \mathcal{S}_i} \sum\limits_{\ell \in \mathcal{I}_{i-j}} \mathcal{K}(\mathsf{d}_{k\ell}, \mathbf{\Theta}_K) & \text{if at least one } \; y_{i-j} > 0 \;\; (j = 1, 2, 3) \\ \mathcal{K}\big(\mathsf{d}_{min}, \mathbf{\Theta}_K\big) & \text{if all } \; y_{i-j} = 0 \;\; (j = 1, 2, 3) \end{cases} \tag{5}$$

where $|\mathbf{d}_i|$ is the cardinality of $\mathbf{d}_i$.

The pre-specified constant d_{min} denotes the minimum distance beyond which infections cannot occur (see, e.g., [15]). In our analysis we set $\mathsf{d}_{min} = 250km$, a distance sufficiently higher than the largest observed distance of 69 kms occurred in the Evros prefecture. For specifying the transmission kernel we have resorted to a variety (see Table 1) of relevant functions.

Table 1 near here

Note that the term $\mathbf{d}_i$ depends on y_{i-j} $(j = 1, 2, 3)$; i.e. the presence of many infections will affect the average distances between infected farms. However, if the $\mathbf{\Theta}_K$ parameters (a, b, c) are restricted on the positive real line, $K\big(\,\cdot\,\big)$ is decreasing in $d_{k\ell}$ and therefore on $\mathbf{d}_i$, hence $K\big(\,\cdot\,\big)$ is bounded

above. This ensures that the model is well defined and doesn't explode. In continuous time this is due to the fact that the stochastic differential equation for λ_t, and therefore also θ_t, has a unique weak solution for all y_i; typically this requires the drift and volatility function to be locally Lipschitz and with a linear growth bound. In discrete time one can note that the model is very similar to the class of models considered in ([16], equation 11) that are known to be geometrically ergodic for $\phi > 0$.

Figure S5 near here

A graphical illustration of the model assumptions is depicted in Figure $S5$ via a schematic representation of the proposed model.

2.1.2 Intensity Decomposition

An interesting interpretation of such models can be found in [17] and references therein where the model components are appropriately split, disentangling the *epidemic and endemic* aspects of disease dynamics. In particular, one can imagine that environmental and seasonal covariate information (i.e. rainfall, average temperature, maximum temperature, minimum temperature, average humidity and seasonal effects) may well relate to the endemic part of the disease while farm-to-farm contacts as described by the spatial component $K\big(\mathbf{d}_i, \boldsymbol{\Theta}_K\big)$ and the covariate of the number of villages infected in the previous week are concerned with epidemic spread. Our model can naturally be adapted to this framework via an additive decomposition of μ_t, the mean driving the instantaneous log rate of infection λ_t. The trajectory of μ_t can be split into its endemic and epidemic parts as follows:

$$\mu_t = \boldsymbol{\Theta}_{endemic} + \boldsymbol{\Theta}_{epidemic}$$

where $\boldsymbol{\Theta}_{endemic}$ ($\boldsymbol{\Theta}_{epidemic}$) denotes the time-dependent endemic (epidemic) component.

This decomposition enhances our ability to inform control strategies since a large epidemic component (relative to the endemic component) would suggest imposing restrictions associated with the spatial allocation of farm structure in the region of interest, whereas the opposite results may indicate that most of infections are due to external factors and thus are less sensitive to such control measures [18].

2.2 The variable selection component

The variable selection problem arises naturally when one wishes to simplify the model structure by eliminating predictors with negligible effects; see [19] for a review.

Here, we are interested to identify important effects and eliminate non-significant covariates both for reasons of interpretation and prediction. When variable selection is ignored, it is common to experience the well known problems of using over-fitted models; see, for example, in [20] (Section 5.1) for details.

Bayesian variable selection typically involves the introduction of a vector of binary indicators $\gamma \in \{0,1\}^{\dim(\beta)}$ which contains each possible combination of covariates to be included in the model. Then, MCMC methodology can be used to estimate the posterior distribution of γ. The exploratory results of [5] suggest that only a few of the variables under consideration, such as the number of villages infected in the previous week and certain meteorological/environmental variables should be included in the final model.

2.2.1 Hyper g-prior setup.

In the present analysis, we use the hyper-g prior introduced by [2] and implemented by [21] in generalized linear models (GLMs). Over the last decade, hyper-g priors [2] have become a common approach for comparing and averaging models within the Bayesian framework due to several reasons. First, the hyper-g prior is a natural descendant of the classic Zellner's (1986) g-prior approach [23] inheriting also its important properties — for example the g-prior leads to BIC-based variable selection and also has an intuitive interpretation based on the use of additional imaginary data points from the null model weighted by the dispersion parameter g. The idea of the hyper-g prior is relatively simple. We introduce an additional hierarchical level to the model formulation by specifying a prior on the shrinkage parameter $g/(g+1)$. Then we estimate the dispersion parameter g (which is responsible for the sensitivity of posterior model probabilities and Lindley-Bartlett paradox [24–26]) by the resulting posterior distribution. The induced procedure leads to a default Bayesian method which is robust over a wide range of values of the hyper-parameters. Moreover, it has been proven to posses important consistency properties such as model selection consistency (posterior probability of the true model goes to one as the sample size increases),

information consistency (posterior probability of the true model goes to one as the error variance reduces) and prediction consistency (the posterior mode of the predicted values converge to the true expected predicted values as the sample size increases). These desirable properties, along with the simple computation within the regression framework, have established the method as a common approach for Bayesian variable selection and averaging.

In our implementation, all covariates are first centered. Moreover, following [22], we consider a slightly modified version of the hyper-g prior which depends on the intercept β_0, that is:

$$f(\beta_{\setminus 0}|\beta_0, \sigma^2) \sim Normal\Big(\mathbf{0}, ge^{\beta_0}(\mathbf{X}_{\setminus 0}^T\mathbf{X}_{\setminus 0})^{-1}\Big), \tag{6}$$

$$f(\beta_0) \sim Normal(0, 10^4), \text{ and } \frac{g}{1+g} \sim \ Beta\left(1, \frac{\alpha}{2} - 1\right).$$

where $\beta_{\setminus 0}$ is the vector β excluding β_0, $\mathbf{X}_{\setminus 0}$ is the data matrix $\mathbf{X}$ without the column that corresponds to the intercept β_0 and e^{β_0} denotes a rough estimate of λ_i under the above prior setup. This prior can be seen as the power prior of [24] for the parameters $\beta_{\setminus 0}$ conditioned upon the constant β_0 and setting the power equal to g. In the power prior setup, if we assume responses equal to e^{β_0} (which corresponds to the expected value under the null model) and covariates with values equal to the observed ones, then this prior accounts for n/g additional data points from the null model.

The above hyper g-prior is slightly modified from the original version in the sense that the covariate effects a-priori depend on the constant parameter. By this way, the corresponding power-prior interpretation is more sensible since we assume that all data are equal (fully supporting the constant model) but without any specific value to be chosen a-priori (in contrast to [22] who proposed to choose the zero value).

Hence, hyper-parameter g controls the weight of the data in the posterior distribution. Following [2], we use a Beta hyper-prior for the shrinkage weight $g/(g+1)$ which introduces an additional hierarchical layer in our model enabling us to estimate its posterior distribution by using information from the data. This approach avoids sensitivity problems created with the direct specification of g and has better consistency properties than the usual g-prior [2].

We primarily focus on the hyper-g prior using the value $\alpha = 4$ which corresponds to the uniform prior on the shrinkage parameter but, as we demonstrate later in the paper, the results are fairly robust for different

choices of $\alpha \in (2, 4]$. The same hyper-g prior structure has been considered for the covariate effects β^z of the zero-inflation probability component p_i.

2.2.2 Implementation details, sensitivity and comparisons.

Bayesian variable selection is notorious for its sensitivity to the choice of prior, particularly the prior variances of β or its multiplier g in (6). This is due to the well known Bartlett-Lindley paradox; see for details in [25] and [26]. To this end, we have performed a number of comparisons and sensitivity analyses to test for the robustness of our results. Specifically, we have compared our results with the Zellner's g-prior [23] with $g = n$ and $g = p^2$ denoted as $ZG(n)$ and $ZG(p^2)$, respectively; n is the sample size and p the number of covariates. The prior distribution for the vector of covariate parameters under the Zellner's g prior approach involve the multivariate Gaussian distribution, with zero-mean vector and a prior variance of the form: $\frac{g}{\tau}(\mathbf{X}^T\mathbf{X})^{-1}$, where τ is a precision parameter. Additionally, an empirical normal prior with an approximate unit information interpretation was used (denoted by EUI).

Concerning the comparison with other priors, the $ZG(n)$ can be thought as the default choice in Bayesian variable selection since it has a unit information interpretation and its results correspond asymptotically to those obtained via BIC [27]. Following the modification of [2], and in order to have prior structure equivalent to the hyper-g prior, we use (6) with $g = n$. Similarly, we have also considered as an alternative the (modified) Zellner's g-prior (6) with $g = p^2$ which corresponds to the risk inflation criterion of [28]; see also [29] for a related comment.

The empirical independent prior with approximate unit interpretation (EUI) suggested by [30] is used as a rough yardstick. It is comprised by independent normal prior distributions for each β_j, i.e. $\beta_j \sim N(0, n\sigma^2_{\beta_j})$ with σ_{β_j} set to the posterior standard deviation of each β_j of the full model with flat priors. This setup obviously uses information from the data to specify the prior variance but its multiplication with the sample size n makes this effect minimal and approximately equivalent to one data-point in a similar manner to the $g = n$ choice.

Finally, for the hyper-g hyper-prior, we perform variable selection using a variety of values for hyper-parameter α and we graphically examine the robustness of the posterior inclusion probabilities for each covariate. We expect that results will be robust as previously reported by [31].

2.2.3 Prior distribution on model space.

Concerning the prior specification of model indicators γ, we primarily use the uniform prior on model space with $\gamma_j \sim \text{Bernoulli}(0.5)$. We also compare our results with the recently used beta-binomial prior on the model space where $\gamma_j \sim \text{Bernoulli}(p)$ with $p \sim \text{Beta}(1, 1)$; see for example [32]. The latter is very useful in large scale problems (with large p) due to its shrinkage effect yielding parsimonious model structures. Moreover, it allows for additional prior variability and robustness [33].

2.2.4 Prior specification for the remaining parameters.

For the $\mathbf{\Theta_K}$ parameters, a weakly informative normal prior with zero mean and large variance (equal to 10^4) was used. Random yearly effects are assumed to follow a $N(0, \sigma_b^2)$ density with $\sigma_b \sim U(0, 100)$. Regarding the priors on ϕ and the kernel parameters $\Theta_K = (\alpha, b, c)$, they are all half normals with large variances and therefore less likely to have an impact on the results. Specifically for the Θ_K parameters, the restriction on the positive real line ensures the well-definition of the model. Other non-informative priors have been also tested for σ_b giving similar results, indicating the robustness of different priors on the standard deviation of the random effects.

3 Epidemic control

This section is concerned with the connection of our model, which can be seen as a typical epidemiological model, to stochastic epidemic models, often considered as invaluable tools for disease control. One of the primary objectives of modelling the spread of an infectious disease is the ability to evaluate disease severity through key measures such as its reproduction number R_0, often interpreted as the average number of infections produced by a single infective during their infectious period. This facilitates for disease control via appropriate prophylactic measures. Specifically, one can evaluate the corresponding extinction probability q as well as the effect of control strategies in achieving the *sine qua non* target of reducing R_0 below unity, thus securing that major outbreaks cannot occur. Here, we link our models to stochastic epidemics via considering the corresponding branching process [34]. Subsequently we exploit this connection by exploring alternative, covariate-based,

scenarios for the probability of hypothetical sheep pox outbreaks going extinct in the region of Evros.

A branching process represents an accurate approximation to a stochastic epidemic model (i) at the early stages of an outbreak when the number of infected individuals is much smaller than the population size and (ii) at the onset of disease re-emergence in the context of endemic diseases. Assuming constant offspring mean ξ is probably reasonable in these two scenarios. Note that our epidemic model does not assume independent disease reproduction as do the branching process models. However, the coupling of the two models appears reasonable in the two scenarios given above and can be made rigorous using techniques described in [35].

A general family for the offspring distribution Z of branching processes is given by the power series family where:

$$P(Z = r) = \alpha_r \frac{\xi^r}{A(\xi)}, \quad A(\xi) = \sum_{r=0}^{\infty} \alpha_r \xi^r,$$

with $\xi = exp(\lambda)$ being the canonical parameter and $\alpha_r \geq 0$. Then the probability, $q(\xi)$, of an epidemic going extinct is the smallest non-trivial root in the interval [0,1] of the equation:

$$A(q\xi) = qA(\xi), \tag{7}$$

see for example [36] and [37]. For $\alpha_r = (r!)^{-1}$ we obtain the Poisson distribution whence $q(\xi)$ can be numerically calculated as the smallest non-trivial root of $\exp(q\xi) = q\exp(\xi)$.

We proceed by exploring the effect of particular covariates upon $q(\xi)$. In particular, we utilise three distinct values for each covariate (minimum, median and maximum) keeping the other covariates fixed at their median values and for each covariate combination we simulate from the posterior density of $q(\xi)$ by sampling from the posterior of the β's and solving $\exp(q\xi) = q\exp(\xi)$ for each set of samples. Regarding the spatial kernel term, we keep the best selected spatial kernel term as was chosen among the candidate functions of Table 1, fixed at the median distance between the susceptible and the previously infected farms. In addition, the random yearly effects term was kept at its mean level. Note that by using the posterior output we preserve the correlation structure of the posterior density, an important aspect when estimating non-linear functionals such as $q(\xi)$. For the ZIP model where $E(Y_i) = (1-p)\xi + p \cdot 0$ it is possible for non-negligible p to have a scenario

where solving (7) results in a q such that $q(\xi)+p > 1$. We truncate such values by adjusting the extinction probability via $Pr(\text{extinction}) = 1 \wedge (q(\xi) + p)$.

We also use a recent result due to [38] to estimate the expected time, say $E(A_Q)$, the outbreak has Q infected farms via:

$$E(A_Q) = \frac{\xi^{Q-1}}{Q(1 \vee \xi)^Q}, Q = 1, 2, ...$$

This gives a somewhat complementary measure of disease propagation which must be interpreted with respect to the length of the incubation period. The following section illustrates the application of the model and the control methods to real data.

4 Application to sheep pox data

The data refer to the sheep pox epidemic in the Evros Prefecture of North-eastern Greece, made available by the Veterinary Directorate of N. Evros Prefecture (VDNEP). The epidemic began on December 1994 and ended in December 1998, infecting 249 premises. The overall number of dead animals was approximately $35,500$. The data comprised of temporal information such as the day of culling, daily records of infected herds, detection time of the virus and the putative infection time. The virological tests used for sheep pox detection were PCR and ELISA. For the purposes of our analysis, a case is a herd infected with sheep pox virus, i.e. a farm for which there is at least one infected animal during a particular week. Table S1 in the supplement presents the aggregated monthly number of sheep pox cases during the 1994-98 period in the Evros Prefecture, Greece. We first illustrate the variable selection procedure through detailed comparisons and extensive sensitivity analyses (Section 4.1.1). Then we investigate the form of the spatial kernels (Section 4.1.2). Having selected our model, we disentangle the endemic and epidemic components in Section 4.2 and estimate the extinction probabilities under several covariate scenarios in Section 4.3.

4.1 Model building and choice of variables

For all MCMC runs we used an output of ten thousand iterations produced from chains with total length equal to 105,000 iterations and after using a

burn-in of 5,000 and a thinning lag of 10 iterations. The analyses were conducted using the WinBUGS software ([39]). The codes and data are available in http://utopia.duth.gr/malesios/ARTICLES/code.pdf. The comparison of the spatial kernels was based upon the mean deviance ($\overline{D}$) due to the well-known asymptotic equivalence of AIC (and DIC) with cross-validation in model selection ([40]).

In the following analysis, we compared various models in order to assess the impact of spatial kernels and the Struthers and McLeish-like OU structure [9]. Specifically, the introduction of the spatial component results in a substantially improved fit reducing the mean deviance from 336.9 to levels ranging from 255.5 to 265.5 depending on the choice of kernel. The adopted OU structure of (1) also results in improved fit compared to Taylor-based alternatives as in (3) where the corresponding mean deviance is around 300; see also Table S3 of the supplementary material. As discussed in Section 2.1 the main difference between the two types of models is that the one of this paper allows for a delayed effect, which is estimated by the data, at which the covariates affect the mean of the intensity process. Hence. in our view, this seems to be the most plausible explanation for the improved fit. In what follows, we present more details for each modelling component.

4.1.1 Covariate selection.

The posterior inclusion probabilities of each covariate for the ZIP model under the hyper-g prior specification with $\alpha = 4$ are summarized in Table 2.

Table 2 near here

These results are obtained using the uniform prior on model space, however the corresponding posterior results under the beta-binomial prior on the model space give the same ordering for the posterior inclusion probabilities of each covariate (see Table S2 in the supplementary material), inflating them upwards for all covariates. Table 3 summarizes posterior results of the coefficients of the a-posteriori supported covariates and the additional model parameters (ϕ and random effects variance) for the selected model.

Table 3 near here

Posterior inclusion probabilities suggest the inclusion of covariates x_6 and x_9, corresponding to the average humidity and the autumn season and, potentially, the selection of covariates x_3, x_4 and x_8 (average and maximum temperature and summer season, respectively) regarding the infection rate. On the other hand covariates x_1 and x_3 (i.e. number of villages infected in the previous week and average temperature) are selected for the prediction of excess zeros. In particular, the importance of villages infected in the previous week on the excess zeros is intuitively natural since more infected villages imply a smaller chance of excess zeros. In summary, it appears that a combination of temperature, humidity and seasonality seems largely responsible for explaining disease occurrence. This is intuitively reasonable and represents a common finding for animal diseases.

In the subsequent analyses, we retain only covariates with posterior inclusion probabilities higher than a threshold. A natural choice for this threshold is the value of 0.5 which is recommended by the theoretical work of [41] and has better predictive properties than the usual choice of the maximum a-posteriori model for specific cases in normal linear regression models. Here, we decided to raise slightly this threshold to the value of 0.6 in order to alleviate the problem of inflated (towards 0.5) posterior inclusion probabilities of non-important effects reported for hyper-g priors (see, for example, [31]). This way, we also avoid the inclusion of variables with high uncertainty concerning their importance. (Note that covariate selection was based upon the utilization of the best fitted - among the tested - spatial kernels, as described in the following section 4.1.2).

Sensitivity analysis. We performed sensitivity analyses using the hyper-parameter values $\alpha \in \{2.01, 2.1, 2.5, 3.0, 3.5, 3.9, 3.99\}$ for the hyper-g prior setup. The results are summarized in Figure S1 in the supplementary material which present posterior inclusion probabilities under the hyper-g prior, for each covariate regressed on the rate of infection. The corresponding results for the covariates related to the probability of excess zeros are depicted in Figure S2 in the supplement. These analyses have been conducted using the uniform prior on model space. The results obtained using the beta-binomial prior (not shown), although quantitatively different to those of the uniform prior, display similar ordering of the importance of the covariates. Comparing the outcomes of the analysis one may deduce that the results are reasonably robust, especially for the covariates associated with excess zeros.

Liang et al. [2] recommended using the values of $\alpha = 3$ and $\alpha = 4$ "...although any choice $2 < a \leq 4$ may be reasonable". Generally, hyper-g priors are relatively robust to the choice of α (e.g. [31]) and this was also the case in the present study; see Figure S1 for the sensitivity of posterior inclusion probabilities over different values of α where any plausible α value indicates the same set of covariates as important. We have a-priori decided to use the value of $\alpha = 4$ in order to express our relative indifference concerning the shrinkage parameter since this choice imposes a $U(0, 1]$ prior for $g/(g+1)$. Sensitivity analysis confirmed that this choice had minimal effect on the posterior selection of covariates. Thus, for the remaining of this paper, we focus on the variables selected by the hyper-g prior with $\alpha = 4$ as reported in Table 2.

Comparisons with other prior setups. Figure 2 presents a summary of the results on the comparison between the various choices of prior for covariate selection.

Figure 2 near here

Four different priors are compared, notably EUI, $ZG(n)$, $ZG(p^2)$ and hyper-g priors. In particular, Figure 2 depicts posterior inclusion probabilities for each covariate of the infection rate for the hyper-g prior and compares these values with the other choices. The corresponding results for the excess zeros are shown in Figure S3 in the supplementary materials.

We observe that the EUI is systematically more parsimonious in comparison to the other methods with the exception of the support on X_8. For the two (with $g = n$ and $g = p^2$) g-prior and the hyper-g prior setups, the inclusion probabilities are identical for all covariates except for X_7. For X_7, the inclusion probability under the hyper-g is lower (≈ 0.5) than the corresponding ones for the other two set-ups. The picture of the inclusion probabilities of the zero-inflated component is more clear. The hyper-g prior seems to have inflated the inclusion of all covariates towards 0.5 with the exception of X_4 and X_6 which are a-posteriori not supported by this method; see Figure S3 in the supplementary materials.

4.1.2 Spatial kernels.

We proceed with fitting different kernel forms to the ZIP models in order to assess their relative importance in disease spread. The posterior estimate for

the K parameter (i.e. α) is also presented.

Figure 3 presents posterior density strip plots for the deviance D of the time-varying kernel models; see Table S3 in the supplementary materials for a comparison between the Struthers and McLeish [9] and the Taylor et al. [8] OU formulations. We additionally include the posterior mean deviance (denoted by $\overline{D'}$) for the model without a spatial component.

Figure 3 near here

For all the fitted kernels, the posterior distribution of the corresponding coefficients are placed well away from zero and the model fit is improved indicating the key importance of the spatial component in describing the progression of the epidemic. There are small differences in the model fit when different kernels are employed. The exponential kernel (B) ([11]) yields the best fit, followed by (D). On the other hand, the kernels (A), (C) and (E) gave a slightly worse fit. The better performance of the exponential-based kernel may be an indication of the importance of short distances on disease spread when compared to the fat-tailed functions (i.e. kernels A and E) which place more mass in the tails of the kernel functions.

The large overlap of the deviance distributions suggests that we cannot select a particular kernel with confidence. Hence, we choose for the remainder of the analysis the exponential decay with a single parameter (kernel C) due to parsimony considerations. The parameter α is associated with the effect of distance with high α values indicating disease transmission to shorter distances. The high estimate of α ($\alpha = 6.749$) suggests that long-distance transmission was unlikely in this outbreak. In summary, the association between spatial information and progression of the sheep pox disease can be expressed by the following form:

$$\mathcal{K}(d_{k\ell}) = 6.749 \exp\left(-6.749 \ \mathsf{d}_{k\ell}\right).$$

Figure 4 depicts the fit of the best model, indicating reasonably good agreement with the observed infectious disease counts.

Figure 4 near here

4.2 Endemic/epidemic decomposition

The results regarding the endemic/epidemic decomposition presented in Section 2.1.2 are given in Table S4 in the supplement where the instantaneous

mean (μ_t) of the log rate of infection (λ_t) is decomposed to its endemic ($\boldsymbol{\Theta}_{endemic}$) and epidemic ($\boldsymbol{\Theta}_{epidemic}$) components.

Specifically, we report the mean μ_t along with the corresponding 95% credible intervals under various scenarios for a hypothetical outbreak. The results indicate that disease spread is likely to increase for lower levels of the endemic/epidemic components. Figure S4 in the supplementary materials demonstrates the endemic/epidemic decomposition over the 5-year period (1994-1998) of the sheep pox epidemic. This graph may assist in illustrating the relative importance of the epidemic spatial component over the endemic part of the model and vice versa during the progress of the disease spread.

4.3 Extinction Probabilities

Here we present results based on the approach introduced in Section 3, investigating the effect of each important covariate on a hypothetical future epidemic outbreak in the Evros region. Specifically, we combine parameter estimates from the historical sheep pox epidemic data and current farm locations in the region to calculate the probability of an epidemic going extinct. The findings are summarized in Table 4 and present extinction probabilities obtained for the minimum, maximum and median value of each covariate, keeping the other covariates fixed at their median value (the covariate values of humidity and distance used for the current analysis refer to year 2012 and are: min=14, max=100, median=65 for humidity; min=1, max=99, median=31 for distance).

Table 4 near here

It appears that a large epidemic may occur when the levels of humidity are low (q=0.036). Also small distances between infected farms are decreasing the probability of an epidemic going extinct (q=0.597). Results of this kind provide an indicator towards potential disease re-emergence. Therefore, monitoring these covariates may be useful for surveillance purposes.
The results for the expected 'Q-occupation times', $Q \in \{1, 2, 3, 4, 5, 6\}$, are summarized in Table S5 (see also Figure S5) in the supplement, and are typical of a supercritical branching process (i.e. $\lambda_t > 1$). Indeed, in the few occasions where $\lambda_t < 1$ we expect only a few farms to get infected thus $E(A_Q)$ is relatively large. In contrast, for fixed λ_t (being a specific combination of humidity and seasonality) $E(A_Q)$ decreases with Q reflecting the standard behaviour of the final size distribution of epidemic models (e.g. [42]).

5 Discussion

In the present paper we proposed a general infectious disease modelling framework which incorporates several common features of epidemic data. This was achieved by extending current spatio-temporal models via different variants of the O-U process. Particular attention was also paid to covariate selection by investigating recently developed priors which have not hitherto been used in models of high complexity. More elaborate versions of the kernel distance, effectively reweighing the contribution of each week based upon some sort of 'infectivity function' are possible and trivial to incorporate. However, the lack of more detailed data would imply identifiability issues so we decided not to pursue this matter further.

Perhaps more importantly, we linked our epidemiological models with stochastic epidemic processes through an approximate branching process representation. Although the two models have distinct characteristics, they can be thought of as different instances of a certain class of counting processes. Thus, one can calculate crude extinction probabilities under different scenarios providing a link to policy decisions targeted at disease control. This complements a byproduct of our model, the identification of 'disease-free' status, stemming from particular regions of the covariates which link to the modelling component that is concerned with excess zeros.

Regarding covariate effects we found that an increase in humidity levels yields a decrease in sheep pox occurrence, while the average temperature has the opposite effect. The significance of humidity is perhaps an indication for the relative importance of airborne spread [43]. The spatial component of our model uses the distances among farms in order to quantify the pressures caused by currently infected farms on the remaining susceptible farm population in the region. For this particular dataset it appears that the disease is transmitted mainly by short-range interactions. Spatially-explicit evidence represents an important component of our model and substantially improves model fit. The absence or possible aggregation of such evidence, a plausible feature of some surveillance systems, would result in simpler fitting of the corresponding models at the expense of potentially biased results. In this paper we used standard Bayesian model determination techniques. An interesting alternative is based upon the prequential principle [44] and this is the subject of current work.

Acknowledgements

We are grateful to Zafeiris Abas and Omiros Papaspiliopoulos for helpful discussions and to two reviewers and the associate editor for their detailed comments which substantially improved the presentation of the paper. The first two authors acknowledge financial support from the Research Centre of Athens University of Economics and Business, in the framework of the project entitled "*Original Scientific Publications*".

References

1. Choi YK, Johnson WO, Jones G, Perez A, Thurmond MC. Modelling and predicting temporal frequency of foot-and-mouth disease cases in countries with endemic foot-and-mouth disease. *Journal of the Royal Statistical Society A* 2012 **175(2):** 619–636.

2. Liang F, Paulo R, Molina G, Clyde MA, Berger JO. Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association* 2008 **103(481):** 410–423.

3. Dellaportas P, Forster JJ, Ntzoufras I. On Bayesian model and variable selection using MCMC. **S**tatistics and Computing 2002 **12:** 2736.

4. Garner MG, Sawarkar SD, Brett EK, Edwards JR, Kulkami VB, Boyle DB, Singh SN. The extent and impact of sheep pox and goat pox in the state of Maharashtra, India. *Tropical Animal Health and Production* 2000 **32(4):** 205–223.

5. Malesios C, Demiris N, Abas Z, Dadousis K, Koutroumanidis T. Modelling Sheep pox Disease from the 1994-1998 Epidemic in Evros Prefecture, Greece. *Spatial and Spatio-temporal Epidemiology* 2014 **11:** 1–10.

6. Roberts GO, Stramer O. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika* 2001 **88(3):** 603–621.

7. Branscum AJ, Perez AM, Johnson WO, Thurmond MC. Bayesian spatiotemporal analysis of foot-and-mouth disease data from the Republic of Turkey. *Epidemiology and Infection* 2008 **136:** 833–842.

8. Taylor JMG, Cumberland WG, Sy JP. A stochastic model for analysis of longitudinal AIDS data. *Journal of the American Statistical Association* 1994 **89:** 727–736.

9. Struthers CA, McLeish DL. A particular diffusion model for incomplete longitudinal data: application to the multicenter AIDS cohort study. *Biostatistics* 2011 **12(3):** 493–505.

10. Keeling MJ. Models of foot-and-mouth disease. *Proceedings of the Royal Society, B, Biological Sciences* 2005 **272(1569):** 1195-1202.

11. Keeling MJ, Woolhouse ME, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, Cornell SJ, Kappey J, Wilesmith J, Grenfell BT. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* 2001 **294:** 813–817.

12. Tildesley MJ, Savill NJ, Shaw DJ, Deardon R, Brooks SP, Woolhouse ME, Grenfell BT, Keeling MJ. Optimal reactive vaccination strategies for a

foot-and-mouth outbreak in the UK. *Nature* 2006 **440:** 83–86.

13. Chis-Ster IC, Ferguson, NM. Transmission parameters of the 2001 Foot and Mouth epidemic in Great Britain. *PLoS ONE* 2007 **6:** e502.

14. World Organisation for Animal Health (OIE). 2010 Terrestrial Animal Health Code, Chapter 14.10. Available at: `http://web.oie.int/eng/normes/MCode/en_chapitre_1.14.10.pdf` (accessed on 15 July 2015).

15. Deardon R, Brooks SP, Grenfell BT, Keeling MJ, Tildesley MJ, Savill NJ, Shaw DJ, Woolhouse EJ. Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica* 2010 **20:** 239–261.

16. Fokianos K, Tjøstheim D. Log-linear Poisson autoregression. *Journal of Multivariate Analysis* 2011 **102:** 563–578.

17. Meyer S, Elias J, Höhle M. A space-time conditional intensity model for invasive meningococcal disease occurence. *Biometrics* 2012 **68:** 607–616.

18. Brown PE, Chimard F, Remorov R, Rosenthal JS, Wang X. Statistical inference and computational efficiency for spatial infectious-disease models with plantation data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2013 doi:10.1111/rssc.12036.

19. Clyde M, George EI. Model uncertainty. *Statistical Science* 2004 **19(1):** 81–94.

20. Yan X. Linear Regression Analysis: Theory and Computing. *World Scientific Publishing Company Pte Limited; ISBN: 9789812834119.* 2009

21. Bové DS, Held L. Hyper-g priors for generalized linear models. *Bayesian Analysis* 2011 **6(3):** 387–410.

22. Ntzoufras I, Dellaportas P, Forster JJ. Bayesian variable and link determination for generalized linear models. *Journal of Statistical Planning and Inference* 2003 **111(1-2):** 165–180.

23. Zellner A. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *In Bayesian inference and decision techniques: Essays in honor of Bruno de Finetti, eds. P.K. Goel and A. Zellner, Amsterdam: North-Holland/Elsevier* 1986 233–243.

24. Ibrahim JG, Chen MH. Power Prior Distributions for Regression Models, *Statistical Science* 2000 **15:** 46–60.

25. Lindley DV. A Statistical Paradox. *Biometrika* 1957 **44:** 187–192.

26. Bartlett MS. Comment on D.V. Lindley's Statistical Paradox. *Biometrika* 1957 **44:** 533–534.

27. Kass RE, Wasserman L. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* 1995 **90:** 928–934.

28. Foster DP, George EI. The risk inflation criterion for multiple regression. *The Annals of Statistics* 1994 **22:** 1947–1975.

29. Fernandez C, Ley E, Steel MF. Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 2001 **100:** 381–427.

30. Ntzoufras I. *Bayesian Modeling Using WinBUGS*. Wiley Series in Computational Statistics, 2009 Hoboken, USA.

31. Dellaportas P, Forster JJ, Ntzoufras I. Joint Specification of Model Space and Parameter Space Prior Distributions. *Statistical Science* 2012 **27:** 232-246.

32. Chipman H, George E, McCulloch RE. The Practical Implementation of Bayesian Model Selection (with discussion), *IMS Lecture Notes - Monograph Series*, 2001 **38:** 65 – 116

33. Wilson MA, Iversen ES, Clyde MA, Schmidler SC, Schildkraut JM. Bayesian model search and multilevel inference for snp association studies. *The Annals of Applied Statistics* 2010 **4(3):** 1342–1364.

34. Jagers P. *Branching Processes with Biological Applications*. Wiley 1975

35. Ball F, Donnelly P. Strong approximations for epidemic models. *Stochastic Processes and their Applications* 1995 **55(1):** 1–21.

36. Guttorp P. *Statistical inference for branching processes.* New York: Wiley 1991.

37. Farrington CP, Kanaan MN, Gay NJ. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics* 2003 **4(2):** 279–295.

38. Ball F, Britton T, Neal, P. On expected durations of birth-death processes, with applications to branching processes and SIS epidemics. *Journal of Applied Probability* 2016.

39. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* 2000 **10:** 325–337.

40. Stone M. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society, Series B* 1977 **39:** 44–47.

41. Berger JO, Barbieri MM. Optimal predictive model selection. *The Annals of Statistics*, 2004 **32(3)**: 870897.

42. Andersson H, Britton T. *Stochastic epidemic models and their statistical analysis.* Springer Lecture Notes in Statistics, 151 2000

43. Bhanuprakash V, Moorthy ARS, Krishnapa G, Sirinivasa GRN, Indrani BK. An epidemiological study of sheep pox infection in Karanataka State, India. *Revue scientifique et technique (International Office of Epizootics)* 2005 **24(3):** 909–920.

44. Dawid AP. Present position and potential developments: Some personal views: Statistical theory: The prequential approach (with Discussion). *Journal of the Royal Statistical Society. Series A* 1984 **147(2):** 278–292.

45. Szmaragd C, Wilson AJ, Carpenter S, Wood JLN, Mellor PS, Gubbins S. A modeling framework to describe the transmission of bluetongue virus within and between farms in Great Britain. *PLoS ONE* 2009 **4(11):** e7741.

Table 1: Summary of transmission kernel functions included in spatio-temporal models.

Notation	$\mathcal{K}(\mathsf{d}_{k\ell}, \boldsymbol{\Theta}_K)$	$\boldsymbol{\Theta}_K$	Reference
A	$b\Big(1+\frac{\mathsf{d}_{k\ell}}{a}\Big)^{-c}$	(a,b,c)	[13]
B	$b\exp\big\{-\big(\frac{\mathsf{d}_{k\ell}}{a}\big)^{c}\big\}$	(a,b,c)	[11]
C	$a\exp\left(-a\,\mathsf{d}_{k\ell}\right)$	a	[45]
D	$\frac{\alpha}{\sqrt{\pi}}\exp\left(-a^2\mathsf{d}_{k\ell}^2\right)$	a	[45]
E	$\frac{a}{4}\exp\Big(-a^{\frac{1}{2}}\mathsf{d}_{k\ell}^{\frac{1}{2}}\Big)$	a	[45]

Supplementary Material for the paper

"Bayesian epidemic models for spatially aggregated count data"

CHRISOVALANTIS MALESIOS, NIKOLAOS DEMIRIS,
KONSTANTINOS KALOGEROPOULOS & IOANNIS NTZOUFRAS

May 17, 2017

Table 2: Posterior inclusion probabilities (γ and γ^z) under the uniform prior on model space for the hyper-g-prior ($\alpha = 4$) variable selection approach applied to ZIP model with flat prior on β_0 (in bold inclusion probabilities above 50%).

	Hyper-g-prior ($\alpha = 4$)		
Parameter	γ	Parameter	γ^z
β_1	**0.514**	β_1^z	**0.969**
β_2	0.404	β_2^z	0.406
β_3	**0.573**	β_3^z	**0.617**
β_4	**0.551**	β_4^z	0.276
β_5	0.376	β_5^z	**0.589**
β_6	**0.757**	β_6^z	0.143
β_7	**0.509**	β_7^z	**0.505**
β_8	**0.539**	β_8^z	**0.589**
β_9	**0.874**	β_9^z	0.492
$\bar{D}$	261.5 (235.8-288.9)		

Table 3: Posterior medians and corresponding 95% credible intervals of the ZIP model.

Parameter	Estimates
β_6 (humidity)	-0.023
	(-0.047,-0.005)
β_9 (autumn)	1.359
	(0.183,2.904)
β_1^z (villages infected in previous week)	-1.405
	(-2.483,-0.575)
β_3^z (average temperature)	0.044
	(0.011,0.127)
ϕ	0.353
	(0.292,0.843)
$\sigma_{b_i}^2$	0.444
	(0.011,6.56)

Table 4: Estimated average extinction probabilities (q) along with corresponding 95% credible intervals based on the branching process approximation.

	humidity	distance
min	0.036 (0.031-0.04)	0.597 (0.576-0.619)
max	0.965 (0.956-0.974)	0.731 (0.711-0.751)
all covariates at median values	0.654 (0.0.628-0.675)	

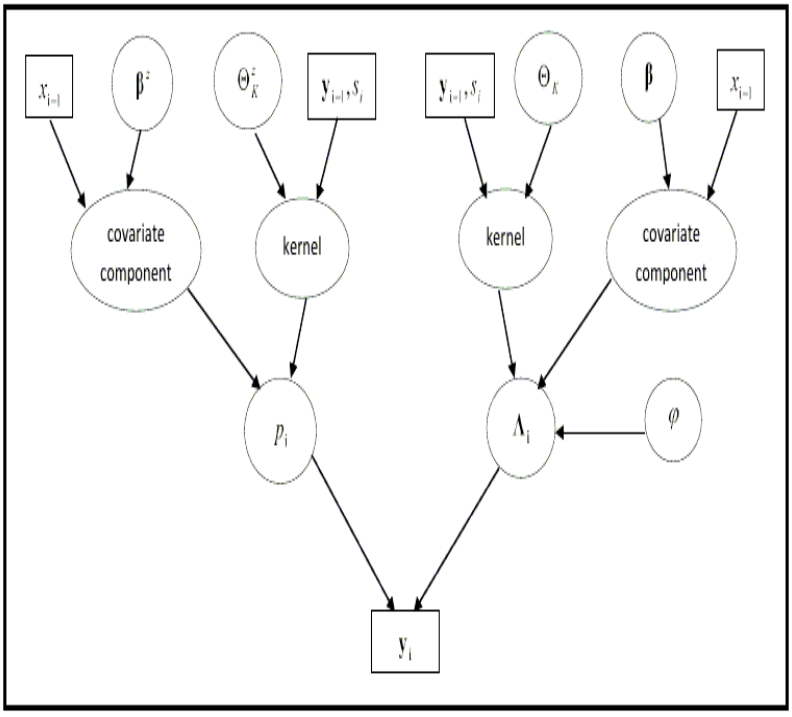

Figure 1: A schematic representation of the model.

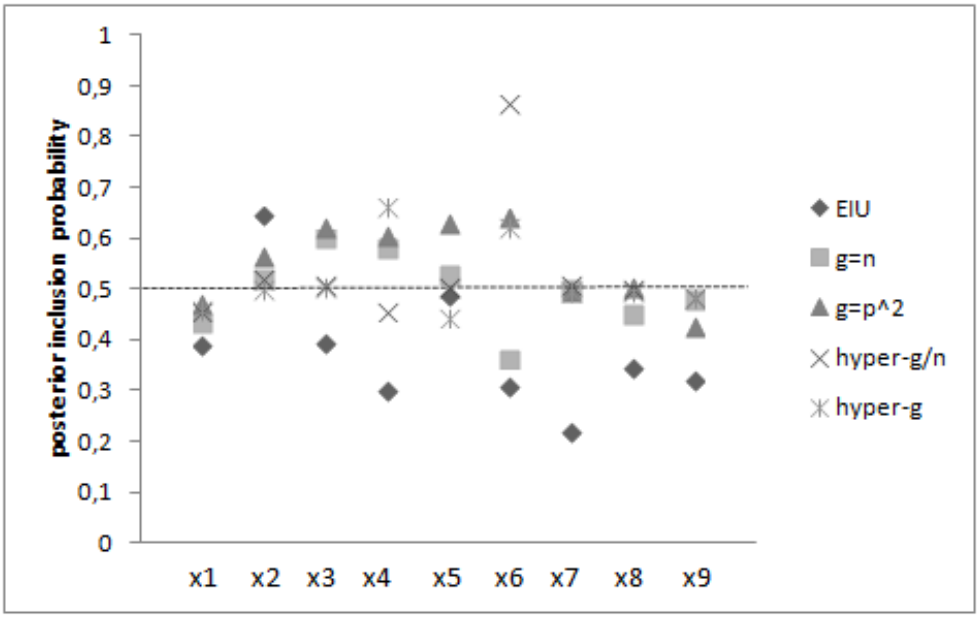

Figure 2: Inclusion probabilities for each covariate of infection rate λ_t of the hyper-g and comparison with other choices (uniform prior).

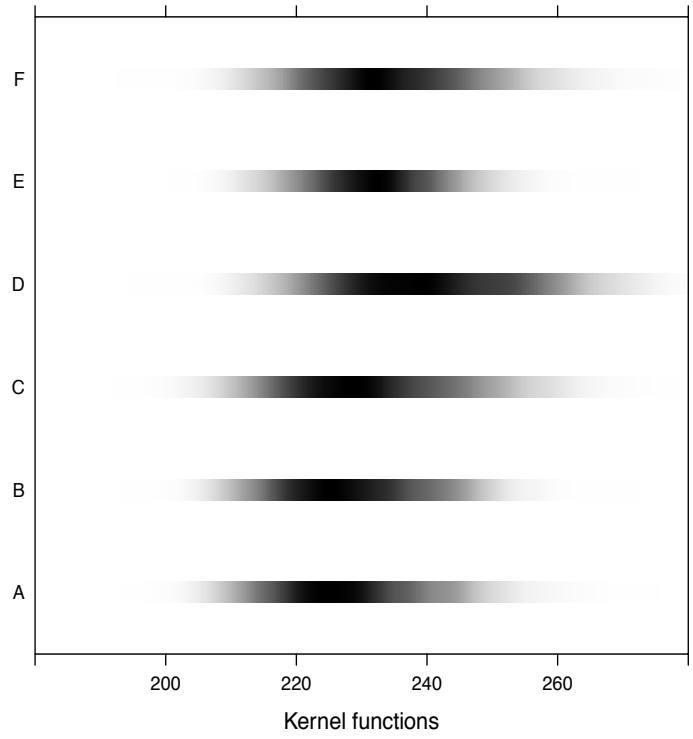

Figure 3: Posterior density strip plots of deviance D for the six spatio-temporal models.

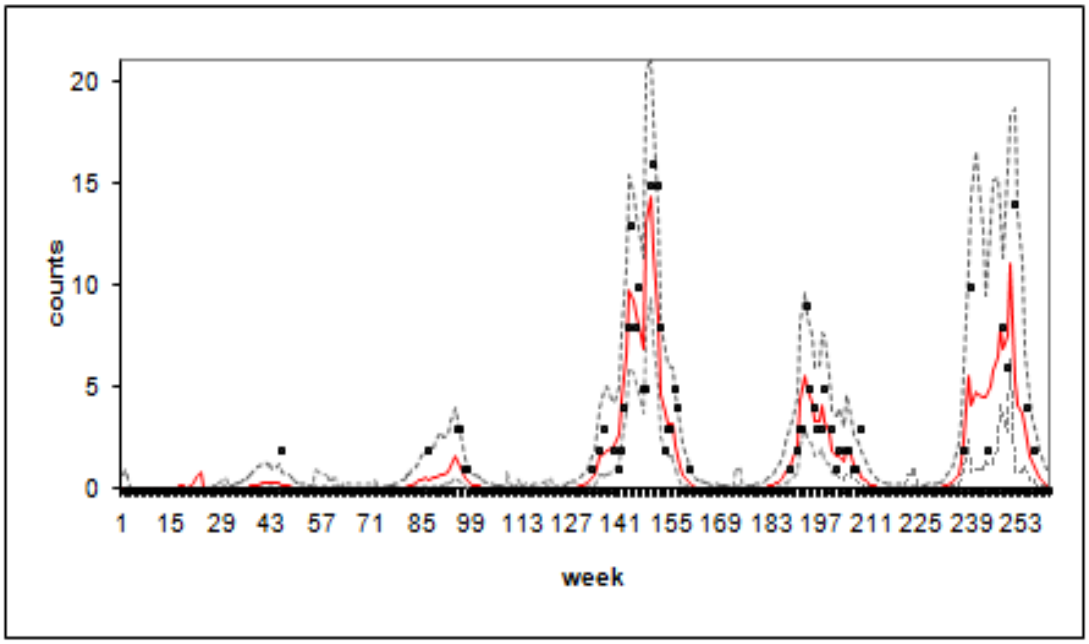

Figure 4: Predicted vs observed numbers of disease occurrence for the ZIP model (dashed lines represent the 95% credible intervals).

Month/year	11/94	10/95	11/95	7/96	8/96	9/96	10/96	11/96	12/96	1/97	TOTAL
Number of cases	2	6	2	6	3	28	34	49	15	1	
Month/year	8/97	9/97	10/97	11/97	12/97	7/98	8/98	9/98	10/98	11/98	249
Number of cases	11	20	10	7	5	12	2	14	14	6	

Table S5: Number of sheep pox cases during period 1994-98 in Evros Prefecture, Greece by month

	Hyper-g-prior ($\alpha = 4$)		
Parameter	γ	Parameter	γ^z
β_1	**0.699**	β_1^z	**0.983**
β_2	**0.593**	β_2^z	**0.609**
β_3	**0.712**	β_3^z	**0.736**
β_4	**0.661**	β_4^z	**0.597**
β_5	**0.537**	β_5^z	**0.651**
β_6	**0.794**	β_6^z	0.347
β_7	**0.664**	β_7^z	**0.664**
β_8	**0.689**	β_8^z	**0.751**
β_9	**0.915**	β_9^z	**0.641**
$\bar{D}$	261.5		

Table S6: Posterior inclusion probabilities (γ and γ^z) under the beta binomial prior on model space for the hyper-g-prior ($\alpha = 4$) variable selection approach applied to ZIP model with flat prior on β_0 (in bold inclusion probabilities above 50%).

	min	median	max
$\mathbf{\Theta}_{endemic}$	0.656	0.244	0.166
	(0.65-0.662)	(0.233-0.256)	(1.149-0.183)
$\mathbf{\Theta}_{epidemic}$	6.081	5.612	5.584
	(5.867-6.295)	(5.42-5.803)	(5.394-5.775)

Table S8: Endemic/epidemic decomposition of μ_t.

	SM OU process	Taylor et al. OU process
Kernel	$\overline{D}$	$\overline{D}$
A Chis-Ster and Ferguson (2007)	265.5	310.1
B Keeling et al. (2001)	255.5	303.4
C Szmaragd et al. (2009)	261.5	302.6
D Szmaragd et al. (2009)	255.7	298.1
E Szmaragd et al. (2009)	261.6	305.1
	$\overline{D'}$	
	336.9	

Table S7: Goodness-of-fit statistics for the time-varying spatial models.

	$E(A_Q)$											
	humidity						max temperature					
	Q						Q					
	1	2	3	4	5	6	1	2	3	4	5	6
min λ_t	0.188	0.094	0.063	0.047	0.038	0.031	1	0.401	0.214	0.129	0.083	0.055
median λ_t	1	0.433	0.288	0.216	0.173	0.144	1	0.432	0.288	0.216	0.173	0.144
max λ_t	1	0.236	0.074	0.026	0.01	0.004	0.568	0.284	0.189	0.142	0.114	0.095

Table S9: Expected time for exactly Q infected farms, $E(A_Q)$, for the various levels of covariates, based on the branching process approximation.

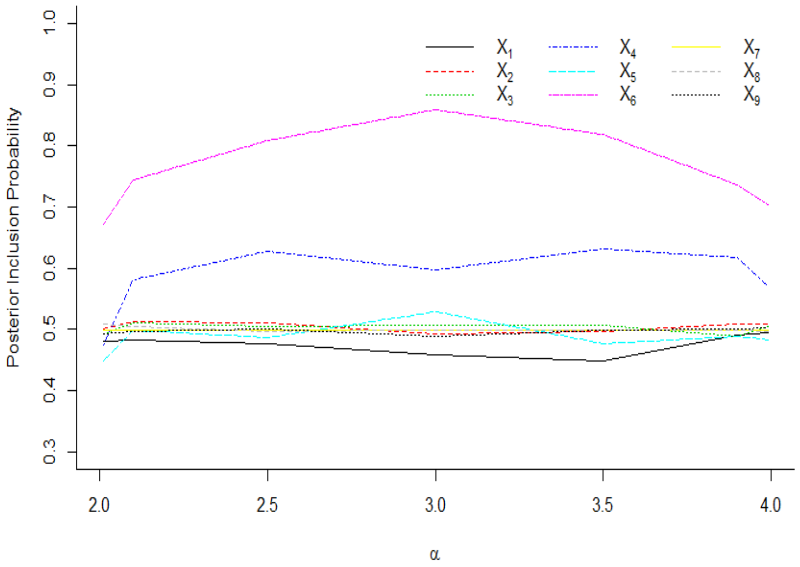

Figure S5: Sensitivity analysis of posterior inclusion probabilities for each covariate of infection rate λ_t of the hyper-g (uniform prior on model space).

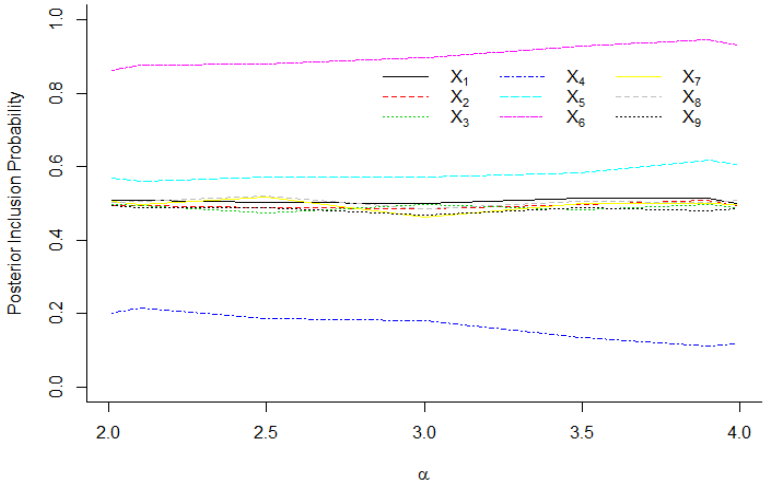

Figure S6: Sensitivity analysis of posterior inclusion probabilities for each covariate of excess zeros of the hyper-g (uniform prior on model space).

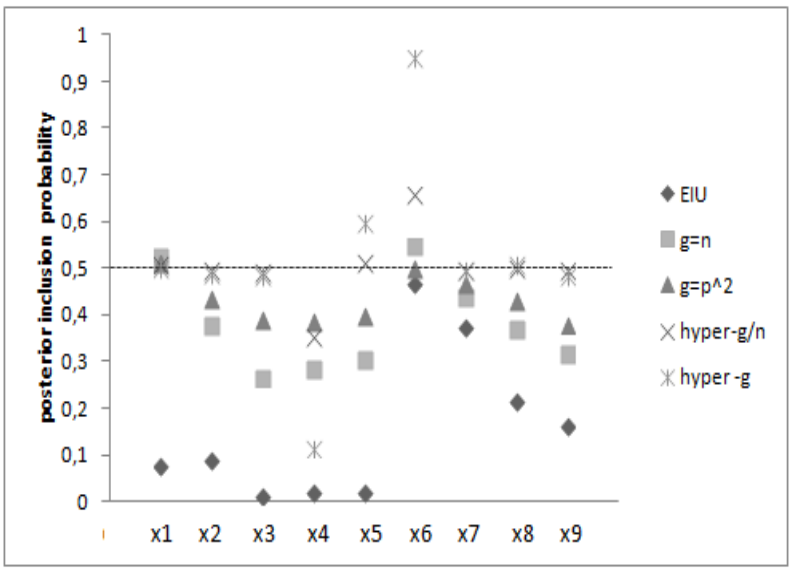

Figure S7: Inclusion probabilities for each covariate of excess zeros of the hyper-g and comparison with other choices (uniform prior).

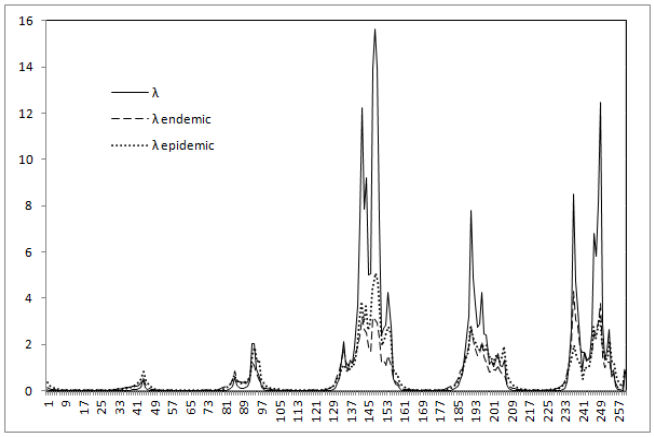

Figure S8: Epidemic and endemic decomposition of μ_t during the 1994-98 sheep pox epidemic in Evros Prefecture, Greece.

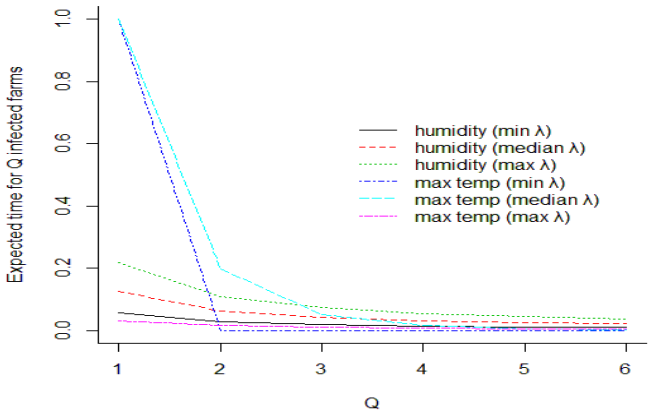

Figure S9: Expected time for exactly Q infected farms based on the branching process approximation.