Modeling spatial risk of the Foot-Mouth-Disease epidemic in South Korea

EunHye Yoo\textsuperscript{1} and JiYoung Lee\textsuperscript{2}

\textsuperscript{1}Department of Geography, University of Buffalo (SUNY), USA
\textsuperscript{2}Department of Geoinformatics, Seoul University, S. Korea

*Corresponding author: eunhye@buffalo.edu

Abstract: The 2010/2011 food and mouth (FMD) disease epidemic in South Korea spread nationwide and resulted in substantial economical damage. Its rapid transmission throughout the country despite the government’s control policy argues for the need to improve our understanding of its spatial dimensions. The spatial point patterns of FMD incidences revealed that FMD incidences formed statistically significant spatial clusters during the early phase of epidemic but its significance and the spatial dimensions dynamically evolved over the course of epidemic. We developed a log-Gaussian Cox process (LGCP) model to quantify and to make inferences on the spatial distribution of FMD risk, while accounting for both known risk factors and unexplained variation. The FMD risk surface estimated from the LGCP model captured the spatial heterogeneity of FMD risk driven by environmental and landscape factors and unknown risk variations at both local and regional scales. This study demonstrated that point pattern analyses and statistical models enabled investigators to improve our understanding on the spatial dynamics of FMD, which may provide useful information for decision makers to make efficient and effective control strategies for future outbreaks.

Keywords: foot and mouth disease, spatial point pattern analysis, log-Gaussian Cox process

I INTRODUCTION
Foot and mouth disease (FMD) is considered by far the most serious infection between livestocks due to its rapid transmission (Keeling, 2005). In South Korea (S. Korea) the outbreaks of FMD during November 2010 and April 2011 spread nationwide and resulted in a total of 3.48 millions susceptible animals being culled (Park et al., 2013). The spatial pattern analysis of infectious diseases may reveal the shape of disease and guide how to control their future spread. However, most studies on the recent FMD incidences in S. Korea have overlooked the potential association between the observed FMD cases and spatial risk factors, and merely focused on descriptive statistics (Moutou and Durand, 1994; Alexandersen et al., 2003). In this study, we aim to fill the gap of FMD literature by examining the spatial patterns of infection risk and their dynamic changes. We will make an inference on the biological processes underlying the observed FMD cases using a log-Gaussian Cox process (LGCP) model.

II MATERIAL AND METHODS
2.1 Data
The official public health records of the FMD epidemic in S. Korea were used, which include the location of farms on which FMD infection was confirmed, the reported date of infection, the type of livestock infected. The types of livestocks infected include cattle, pig, deer, and goat, but we focused on cattle and pigs in this paper. Assuming that the spatial variation of FMD prevalence is affected by both the environmental factors and the spatial configurations of
livestock farms, we incorporated 2010 national census data on the number of domestic animals and farms. Figure 1(A) shows the elevation, (B) and (C) are the spatial distribution of livestock density and the average temperature of December of 2010 and January of 2011. The driving distance between each grid cell at the same resolution to the index farm in Andong is illustrated in Figure 1(D). The index farm was considered as an initial seed of infection because there was hardly any FMD cases reported in S. Korea prior to this epidemic in 2010 (Park et al., 2013).

Figure 1: Spatial distributions of risk factors; (a) Elevation (in meter), (b) Density of livestock farm holdings per non-overlapping 5 × 5 km\(^2\) areas, (c) Average temperature in winter months (in Fahrenheit), (d) Driving distance from the point source of infection (in kilometer).

### 2.2 Spatial point pattern analysis

The K function describes the extent to which there is spatial dependence in the arrangement of georeferenced point locations of events in the study area (Gatrell et al., 1996). This function can be estimated from an observed event distribution and its behavior in a particular situation is assessed. One of the standard reference distribution is homogeneous poisson model. Given that FMD is infectious disease which forms a spatial cluster, homogeneous model is least sensible to the observed FMD cases. As an alternative, we estimated the inhomogeneous K-function \( \hat{K}_{inhom}(r) \) (Baddeley et al., 2000) where the spatially varying intensity is estimated from the observed FMD cases. To make an inference about observed point patterns under the consideration of non-stationarity, we simulated envelops with the inhomogeneous K function. An observed process is assumed to be different than the null process model, if the estimated K function at a specific distance lie outside the envelop (Moller and Waagepetersen, 2003).
2.3 A log-Gaussian Cox Process model

A special case of Cox processes where log-intensity is a Gaussian process is a log-Gaussian Cox Process (LGCP) model (Møller et al., 1998; Diggle, 2013). We modeled the spread of FMD in S. Korea as a poisson process that was partially driven by observed risk factors. This conceptualization enabled us to link the observed global heterogeneity in the dispersion of FMD to the spatial patterns of environmental factors. On the other hand, we viewed the unexplained spatial variability of the point process as outcomes of infectious cases transmitting the disease to nearby susceptible. Assuming that FMD incidence locations \( \{ s_i, i = 1, \ldots, n \} \) are independently distributed conditional on a Gaussian spatial random field \( U(s) \), the LGCP model for FMD is written as

\[
\begin{align*}
\{ s_i; i = 1, \ldots, n \} | U(s) & \sim \text{Poisson}(\Lambda(s)) \\
\log \Lambda(s) &= \mu + X(s) \beta + U(s) \\
\text{Cov}\{U(s), U(s')\} &= \sigma^2 \rho(\| s - s' \|/\theta)
\end{align*}
\]

the vector of covariates \( X(s) = [X_1(s), \ldots, X_4(s)] \) consists the elevation \( X_1 \), the density of livestock holdings \( X_2 \), average temperature of winter months \( X_3 \) at the event location \( s \), and \( X_4 \) denotes a driving distance from the index farm. We included \( X_4 \) as a risk factor for FMD risk, because the initial FMD case on November 28th, 2010 served as a source of infection in subsequent cases and the risk of infection decreases as susceptible is further away. Based on related studies (Ferguson et al., 2001; Grubman and Baxt, 2004), which showed that cold temperatures aided the virus to persist outside the host for a longer time period, we incorporated the spatial variability of winter temperature as one of known risk factors in our LGCP model. The unexplained spatial variation \( U(s) \) in FMD risk is a stochastic component (the latent Gaussian process) whose spatial structure was modeled as a function of the separation vector between any two incidence locations with a correlation function specified with two parameters — a range \( \theta \) and a sill \( \sigma^2 \). In geostatistical jargon, a range refers to the maximum distance at which spatial dependence exists and a sill is a maximum variance of the process.

The target of inference of the LGCP model in Equation 2 and 3 includes \( \{ \beta, \theta, \sigma, U(s) \} \). Møller et al. (1998) originally used the method of minimum contrast estimation, which is often computationally formidable (Diggle, 2013). We addressed this computational challenge by using two approximations: the Gaussian Markov Random Field (GMRF) approximation to the Matern correlation (Banerjee et al., 2004) and integrated nested Laplace approximation (INLA) for the marginal posterior distributions. More specifically, we used the \textit{lgcp} function implemented in \textit{geostatsp} library (Brown, 2015) where a computational grid that consisted of cells 5 × 5 km\(^2\) in dimensions is superimposed over the study region \( A \). The FMD risk surface \( \Lambda(\cdot) \) was defined over a collection of cells \( g_l, k = 1, \ldots, L \) that form a disjoint partition of \( A \). Let \( Y(g_l) \) be the case count within the \( l \)th cell, where the count follows a Poisson distribution with a FMD risk \( \Lambda_l \). The risk is defined as product of an offset \( \delta_l \) and the intensity function (or spatial random function) \( \Lambda_l \), which is a function of known risk factors and a latent Gaussian process as

\[
\begin{align*}
Y_l & \sim \text{Poisson}(\delta_l \Lambda_l) \\
\log(\Lambda_l) &= \mu + X_{1l} \beta_1 + X_{2l} \beta_2 + X_{3l} \beta_3 + X_{4l} \beta_4 + U_l
\end{align*}
\]

The spatial random effects \( U(s) \) is approximated by \( U_l \) using GMRF where \( s \in g_l \) with Gaussian priors with a zero mean, \( \theta \in [5, 20] \) km for range of the spatial random effects, and \( \sigma \in [1, 4] \) for its standard deviation.
III RESULTS

The FMD epidemic in 2010/2011 has shown that more than 90% of FMD cases occurred at five provinces throughout the epidemic and over 60% of the total number of livestock farms in S. Korea reside within these five provinces. We further examined their relationships in the LGCP model by taking the total number of the farms in each town and treated them as population at risk. The observed FMD epidemic curve in Figure 2(a) showed that the major peaks reached on the week 4 and week 7.

![Figure 2: Temporal and spatial profile of FMD incidences in S. Korea](image)

We defined four time periods of the FMD epidemic to represent the natural phases of the epidemic: weeks 1-2 were the period of rapid spread followed by a sharp decrease (Period 1); in weeks 3-5 there was a medium- to long-distance spread combined with localized transmission from newly affected farms (Period 2); in weeks 6-8 the epidemic peaks (Period 3); in weeks 9-22 eradication was achieved (Period 4). The locations of FMD incidences are shown in the map of Figure 2(b) with different symbols per period. We characterized their spatial patterns using the inhomogeneous K-function, and the results are summarized in Figure 3. Spatial clustering is commonly found during the first three periods, although the spatial extents at which the clustering is observed are different with varying statistical significances. K-function for Period 1 in Figure 3(A), for example, shows statistically significant clustering across all distances, but the spatial cluster exists up to 20 km in Period 2. The K-function estimates for Period 3 (Figure 3(C)) shows statistically significant spatial clustering at medium- to long-distance. Lastly, the K-function in the last period in Figure 3(D) showed regularity without statistical significance.

Bayesian inference on the LGCP model parameters were conducted using the R-INLA package in the statistical computing language R (R Core Team, 2015). The posterior distribution of spatial intensity of FMD and spatial random effects are shown in Figures 4: The lower and upper bounds of the posterior mean of spatial intensity of FMD cases are shown in Figures 4(A), (C), and (B), respectively. The corresponding posterior distribution of spatial random effects are shown in Figures 4(D), (F), and (E). The posterior distribution of parameters are also summarized in Table 1. The posterior mean of parameter estimates $\hat{\beta}$ are similar to $\hat{\beta}^*$. 
except the coefficient of elevation $\hat{\beta}_1$ which was not statistically significant. It is clear that the influence of temperature appears to be relatively substantial compared to other risk factors in LGCP model, too. Both the FMD risk map in Figure 4 and the non-zero estimate of intercept $\hat{\mu}$ indicate that the risk of FMD was not constant over the study region but instead was spatially varying. Lastly, the posterior mean of range parameter ($\hat{\theta}$) is approximately 14.49 km with lower and upper bounds of 9.94 km and 20.12 km, respectively.

Table 1: Posterior distribution of LGCP model parameters

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>0.025 quantile</th>
<th>0.975 quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.25</td>
<td>2.99</td>
<td>6.38</td>
<td>18.15</td>
</tr>
<tr>
<td>Elevation</td>
<td>-0.08</td>
<td>0.08</td>
<td>-0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>Livestock holdings</td>
<td>0.30</td>
<td>0.08</td>
<td>0.14</td>
<td>0.47</td>
</tr>
<tr>
<td>Average temperature</td>
<td>-3.84</td>
<td>0.83</td>
<td>-5.48</td>
<td>-2.21</td>
</tr>
<tr>
<td>Driving distance</td>
<td>-1.10</td>
<td>0.23</td>
<td>-1.56</td>
<td>-0.66</td>
</tr>
<tr>
<td>Range</td>
<td>14,492.42</td>
<td>2,602.11</td>
<td>9,944.41</td>
<td>20,120.24</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>1.12</td>
<td>0.94</td>
<td>1.32</td>
<td></td>
</tr>
</tbody>
</table>

IV DISCUSSION AND CONCLUSIONS
We described the FMD transmission process in S. Korea during 2010-2011 using spatial point pattern analyses. The application of inhomogeneous K function to the farm level FMD cases...
across four time periods suggested the presence of dynamically evolving the spatial patterns of FMD transmission over the course of the epidemic. We found that clusters at short scales during Period 2 and 3 were not statistically significant, which contrasts to the common assumption that contagious diseases form a spatial cluster. The absence of short scale clusters (less than 5 km) in both periods might be due to the enforcement of the government’s control policy including culling and the movement restriction (Park et al., 2013).

Inhomogeneous K function analysis enabled us to account for the non-stationarity of the disease spread, but the spatial intensity was purely based on the observed FMD cases. For a further investigation of FMD transmission with respect to the behaviors and appearance of intensity function, we developed a LGCP model where spatially varying intensity of the disease was estimated via a log-linear function of environmental and socioeconomic factors. We also took a Bayesian approach to explicitly take into account parameter uncertainty. We found that LGCP models for FMD can be used as a tool for improving our understanding of the spread of FMD and optimizing disease control. For instance, our analyses demonstrated that temperature plays a critical role in determining the spread of FMD — probably because the virus tends to persist in lower temperatures. As shown in the map of FMD intensity in Figure 4(A)-(C), the area with high intensity is centered at the source of infection where the onset of FMD epidemic started. On the other hand, the maps of unexplained risk in Figure 4(D)-(F) illustrate the areas with high to low risk whose shapes and sizes vary over the study area.

The Bayesian estimates of the LGCP model parameters and prediction of random effects enabled us to assess the spatial variability of FMD incidences risk and to identify the spatial clustering under the explicit consideration of population at risk. The non-zero estimate of spa-
tial variability confirmed that FMD incidences formed a surprising tendency to cluster together even after the spatially inhomogeneous livestock farm density was taken into account. Perhaps LGCP may not be the only means of testing the hypothesis of spatial independence or interactions between events as variant K-functions can provide similar outcomes. However, LGCP models provide statistically rigorous inference and estimation and allow investigators to estimate the second order properties of the underlying process, that is, the spatial dependence of events. Furthermore, the LGCP inference using INLA approximation has relatively tractable moment properties, which allowed a simple method of parameter estimation and a flexible specification of the space-time covariance structure. We simplified the mean structure as a set of spatial covariates but this approach is naturally extendable to incorporate additional covariate information.

The proposed analysis/modeling framework is general and can be applied to other infectious diseases or FMD cases in other countries. The proposed LGCP model can fully account for the temporal aspect of FMD incidences by incorporating relevant temporal covariates and specifying spatio-temporal covariance structure (Diggle, 2013). In doing so, for example, we might have been able to take into account the changes in the susceptible population due to the culling policy. We expect that the longer time series of FMD incidence will improve our understanding (given situation) and allows us to evaluate a wide range of control strategies using the model.

References


