New model diagnostics for spatio-temporal systems in epidemiology and ecology
Lau, Max Siu Yin; Marion, Glenn; Streftaris, George; Gibson, Gavin Jarvis

Published in:
Interface

DOI:
10.1098/rsif.2013.1093

Publication date:
2014

Document Version
Peer reviewed version

Link to publication in Heriot-Watt University Research Portal

Citation for published version (APA):
1 Reconstruction of the Epidemic Process

In this section, we provide the theoretical justification for the use of latent residuals in reconstructing the epidemic process. As we shall see, three independent random draws of $U(0, 1)$ variates are involved in constructing three different aspects of the epidemic process. The residuals are named as Exposure Time Residuals (ETR), Infection-Link Residuals (ILR) and Latent Time Residuals (LTR) in the main text. They correspond respectively to the times of sequential exposure, infectious contact of an exposure event (infection link) and the sojourn time in class E.

1.1 The Epidemic Model

We first recap the epidemic process. Using the same notation defined in the main text, in the spatio-temporal context the transition from $S$ to $E$ (i.e., an infection event) in time interval $(t, t + dt)$ is characterized by the following probability

$$P(S(t + dt) = s - 1|S(t) = s) = \left\{ \sum_{j \in \xi_S(t)} (\alpha + \beta \sum_{i \in \xi_t(t)} K(d_{ij})) \right\} dt + o(dt). \tag{1}$$

On the other hand, the transition of an exposed individual $k$ from class $E$ to class $I$ in the time interval $(t, t + dt)$ is governed by the following probability,

$$P(k \in \xi_t(t + dt)|k \in \xi_E(t), k \notin \xi_t(t)) = h_T(t) dt + o(dt), \tag{2}$$

where $h_T(\cdot)$ is the hazard rate function corresponding to the waiting time $T$ from class $E$ to class $I$. The transition from class $I$ to class $R$ should follow an analogous form given a waiting time distribution between these two classes.

1.2 Construction of the Exposure Times and the Infection Links

Sellke’s construction [1] requires a threshold to be assigned to each individual in the population and therefore censoring of the threshold (or the residual) of a non-exposed individual is necessary [2]. To avoid the censoring issue in using Sellke thresholds and to construct a test specifically for testing the fit of the spatial kernel, we consider a construction by defining a population-level threshold (in contrast to the individual-level threshold in the Sellke construction) and explicitly construct the infection links within the competing-risk framework (i.e., a failure event could only occur in one and only one of the possible ways to fail). We first define the accumulated infectious challenge $Q(t)$ in the population by time $t$ as

$$Q(t) = \int_0^t \sum_{j \in \xi_S(y)} \{\alpha + \beta \sum_{i \in \xi_t(y)} K(d_{ij})\} dy. \tag{3}$$

Also we denote the time of $k^{th}$ infection event by $t_k$, where $t_0 = 0$, and conditional on the $(k - 1)^{th}$ infection, we define the additional infective pressure $A_{k-1}(t)$ in the population between $t_{k-1}$ and time $t$ as

$$A_{k-1}(t) = \int_{t_{k-1}}^t \sum_{j \in \xi_S(y)} \{\alpha + \beta \sum_{i \in \xi_t(y)} K(d_{ij})\} dy$$

where $k \in \{1, 2, 3, \ldots\}$ and $t \geq t_{k-1}$. We also define the threshold $r_{1k}$ (i.e., ETR) for the $k^{th}$ infection event as

$$r_{1k} \sim \text{Exp}(1).$$

Note that $r_{1k}$ can be transformed to $U(0, 1)$ by the cumulative distribution function of the exponential distribution. It is also noted that the threshold we defined is on the population-level (i.e., aggregating all infective challenges in the population) instead of assigning an individual threshold for each susceptible in the population. Lastly we let $h_G(\cdot)$ be the hazard function of the random threshold $r_{1k}$. We then prove the mechanism defined below is equivalent to the mechanism specified by equation (1).
Proposition 1.1 A threshold model, which states that the \( k^{th} \) infection event would occur at time \( t \), where \( t \geq t_{k-1} \), only if \( A_{k-1}(t) \geq r_{1k} \), is equivalent to the mechanism specified by equation (1).

Proof Denoting \( F \) as the history of the change of infectiousness before time \( t \), the probability of having the \( k^{th} \) infection event at time interval \((t, t + dt)\), where \( t \geq t_{k-1} \), from the threshold model defined in proposition 1.1 is given by the following,

\[
P(\{A_{k-1}(t) \leq r_{1k} \leq A_{k-1}(t) + dA_{k-1}(t) \} | r_{1k} > A_{k-1}(t), t_{k-1}, F_t) = \frac{h_G(A_{k-1}(t))dA_{k-1}(t) + o(dA_{k-1}(t))}{dQ(t) + o(dt) + \sum_{j \in \xi(t)} (\alpha + \beta \sum_{i \in \xi_j(t)} K(d_{ij})) dt + o(dt).}
\]

The second last equality holds as

\[
A_{k-1}(t) = Q(t) - Q(t_{k-1}).
\]

Further, conditional on the occurrence of the \( k^{th} \) infection event at time \( t_k \), we can construct the corresponding infection link within the competing-risk framework. We first denote \( p_{ij} \) as the probability of individual \( i \) infecting individual \( j \) where \( i \in \xi(t_k) \) and \( j \in \xi(t_k) \). By noting that \( p_{ij} = z_{ij}dt \), where \( z_{ij} = \alpha \) when considering the primary infection and \( z_{ij} = \beta K(d_{ij}) \) for \( i \in \xi(t_k) \) and \( j \in \xi(t_k) \), it can be readily seen that the total probability of having this infection event is in fact the sum of all \( p_{ij} \) and this sum is the same as the transmission probability in equation (1) and hence the equivalence is not altered in constructing the infection link corresponding to this infection event. The actual infection link is then determined by a random draw from \( U(0, 1) \) and the values of \( z_{ij} \): we first sort all the \( p_{ij}' \) \( \leq \sum_{i,j} z_{ij} \) in ascending order and denote them as \( p_{(1)}, \ldots, p_{(m)} \) where \( m \) is the total number of the possible links; we then draw a random number, \( r_{2k} \), from \( U(0, 1) \) (i.e., the ILR), if \( r_{2k} \in [p_{(m)}', p_{(m+1)}'] \) \( n^{th} \) link is realized as the actual infection link.

1.3 Construction of the Sojourn Time

We now propose a threshold model for the construction of the sojourn time in class E (i.e., the latent period) and prove its equivalence with the mechanism defined in equation (2). Define the pressure \( Q_w(t) \) exerted on an individual corresponding to the \( k^{th} \) infection event by time \( t \) by the following expression,

\[
Q_w(t) = \int_0^t h_T(y)dy
\]

where \( t = 0 \) corresponds to the exposure time of the individual. Also define the threshold \( r_{3k} \) (i.e., LTR) such that

\[
r_{3k} \sim \text{Exp}(1).
\]

Similarly, \( r_{3k} \) can be transformed to \( U(0, 1) \) by the cumulative distribution function of the exponential distribution. Finally define \( h_{Q_w}() \) to be the hazard function of the (random) threshold \( r_{3k} \).

Proposition 1.2 A threshold model, which states that the transition of an exposed individual \( k \) from class E to class I would occur at time \( t \) only if \( Q_w(t) \geq r_{3k} \) (i.e., \( t = \inf \{ t' \mid Q_w(t') \geq r_{3k} \} \)), is equivalent to the mechanism specified by equation (2).

Proof The transition probability from class E to class I during time interval \((t, t + dt)\) from the threshold model defined in proposition 1.2 is given by the following,

\[
P(Q_w(t) < r_{3k} \leq Q_w(t) + dQ_w(t) | k \in \xi_E(t), r_{3k} > Q_w(t)) = h_{Q_w}(Q_w(t))dQ_w(t) + o(dQ_w(t)) = dQ_w(t) + o(dt) (\because h_{Q_w}() = 1) = h_T(t)dt + o(dt).
\]

Denote \( F_{Q_w}() \) and \( F_T() \) as the cumulative distribution functions for the threshold and for the sojourn time in class E respectively. It should be noted that \( F_T(t) = F_{Q_w}(Q_w(t)) \) where \( F_{Q_w}(Q_w(t)) \sim U(0, 1) \). Therefore, the sojourn time can be obtained by computing \( F_T^{-1}(r') \) where \( r' \sim U(0, 1) \).

It can be readily seen that this approach can be extended to transition between two classes in which the forms of sojourn time distributions are explicit (e.g., sojourn time in class I).
1.4  Rationale for Ordering the Infection Links

The ordering of infection links for an exposure event according to the strength of the links $p_{ij}$ is not necessary when one only wishes to generate stochastic realisations of the epidemic using a functional-model representation. However, the operation is crucial to ensuring that the imputed $\hat{r}_2$ may be informative regarding possible mis-specification of the kernel. In particular, our goal is to distinguish between the comparative goodness-of-fit of radially symmetric kernels that differ in terms of their behaviours at short and at long distances. When the ordering operation is used, the imputed residuals corresponding to infection links that represent very long or very short range transmissions are located at the extremes of the unit interval. A mis-specified kernel, which misrepresents the propensity for short- or long-range transmission, may therefore be expected to cause the distribution of imputed $\hat{r}_2$ to deviate from $U(0, 1)$ by exhibiting a concentration, or a scarcity, of residuals at the extremes of the unit interval, dependent on the nature of the mis-specification. Our results presented in the main text suggest that this hoped-for sensitivity is indeed achieved.

To illustrate the point further we consider the case where an exponentially-bounded kernel is fitted to data generated using a power-law kernel (corresponding to Scenario I in section 2.1 of the main text). Figure S1 is a schematic representation of the relative strength of interaction of these kernels. The strength of infection links are represented by the segments length, and the monotonically decreasing nature of these kernels means that short range links are associated with stronger links than longer range interactions. Figure S1 shows the tendency of the exponential kernel to underestimate the interaction strength at short and long distances; as a result, the lengths of these infection links (which correspond to short and long distances transmission) are reduced when an exponential kernel is fitted. Should these links are imputed as the active links for exposure events, then the corresponding imputed residuals will be located closer to the extremes of the unit interval, than had the correct kernel been used, leading to a concentration of residuals at the extremes (also see Figure 2 in the main text) in this case. Note that the above ordering operation is specifically aimed at comparing radially symmetric kernels with different tail properties, a common goal in epidemic studies. Alternative ordering schemes may be considered if our prior knowledge suggested a different form of mis-specification. For example, if transmission were to occur preferentially in certain directions, for example due to prevailing wind or other effects, then an ordering operation that took account of the direction, as well as the length, of the I-S links may be advisable. Such developments are beyond the scope of this paper.

![Figure S1](image.png)

Figure S1: A schematic representation of the relative strength of interaction of these kernels. The strength of infection links are represented by the segments length.

2  Statistical Inference

2.1  Likelihood

We consider an epidemic from a spatio-temporal $S-E-I-R$ model. Consider a population with size $N$. Assume that individuals in the population are all susceptible at time 0 which is the time of introduction of the force of primary infection, and assume that the epidemic is to be observed up to time $t_{\text{max}}$. Let $E = (E_1, E_2, \ldots, E_N)$ be a vector of the exposure times of $N_E$ individuals, $I = (I_1, I_2, \ldots, I_N)$ be a vector of the times of becoming infectious of $N_I$ individuals; $R = (R_1, R_2, \ldots, R_N)$ be a vector of the times of recovery of $N_R$ individuals. Also,
we let $\xi_U$ be the set of indices of the individuals remaining in class $S$ at the end of observation period $t_{\text{max}}$; and, we let $\xi_E$, $\xi_I$ and $\xi_R$ be the set of indices of the individuals who have gone through class $E$, class $I$ and class $R$ by $t_{\text{max}}$ respectively. Also, we let $\xi_E i$ be the set corresponding to the exposed individuals who have not been infectious up to time $t_{\text{max}}$ and $\xi_I R$ be the set corresponding to the infectious individuals who have not recovered up to time $t_{\text{max}}$. Similar to previous sections, we let $\alpha$ be the primary infection rate and $\beta$ be the infection rate in a (susceptible-infectious) contact, and let $K(d_{ij}; \kappa)$ be the kernel function of the Euclidean distance between individual $j$ and $i$ (i.e., $d_{ij}$). The latent period (i.e., the sojourn time in class $E$) is assumed to be a two-parameter distribution $F_w(\cdot)$ characterized by parameters $\mu$ and $\sigma^2$ (the density function is denoted by $f_w(\cdot)$); similarly the infectious period (i.e., the sojourn time in class $I$) has density and distribution functions denoted as $f_w(\cdot)$ and $F_w(\cdot)$ respectively. Finally, let $\theta = (\alpha, \beta, \mu, \sigma^2, \gamma, \eta, \kappa)$ be vector of parameters in the model. As a result, we can express the likelihood function given the times of events as

\[
L(\theta; E, I, R) = \left\{ \prod_{j \in \xi_E} \left\{ \alpha + \sum_{i \in I_{E_j}^i} \beta K(d_{ij}; \kappa) \right\} e^{-q_j^i} \right\} \times \prod_{j \in \xi_U} e^{-q_{r_j}} \\
\times \prod_{j \in \xi_I} f_w(I_j - E_j; \mu, \sigma^2) \times \prod_{j \in \xi_R} f_w(R_j - I_j; \gamma, \eta) \\
\times \prod_{j \in \xi_{E \cap I}} (1 - F_w(t_{\text{max}} - E_j; \mu, \sigma^2)) \times \prod_{j \in \xi_{I \cap R}} (1 - F_w(t_{\text{max}} - I_j; \gamma, \eta))
\]

where $I(t)$ is number of individuals in class $I$ at time $t$, and $\xi_E^{-1}$ is the set of individuals in class $E$ excluding the index case, and $I_{E_j}^i$ is the set of individuals in $I$ class (and not yet recovered) before time $t$. Also,

\[
q_j^i = \int_{t=0}^{c_j} \left\{ \alpha + \sum_{i \in I_{E_j}^i} \beta K(d_{ij}; \kappa) \right\} dt,
\]

and

\[
q_{r_j} = \int_{t=0}^{t_{\text{max}}} \left\{ \alpha + \sum_{i \in I_{E_j}^i} \beta K(d_{ij}; \kappa) \right\} dt.
\]

### 2.1.1 Adaptation to Giant Hogweed Data

The likelihood function described above deals with a general S-E-I-R model which is used in our simulated example. In application to the giant hogweed data, we have instead fitted a S-I model. The likelihood function can be readily adapted by first omitting the terms corresponding to the sojourn time distributions. Two extra components are also modelled. First, the relative suitability of the sites is incorporated; second, as it is known that the survey was not extensive during the period when the snapshot was taken, we also estimate the probability that a colonized site reported at the third snapshot was actually non-reported at the second snapshot and denote it by $p$. Also, we remark that we have considered the likelihood for the observations at the last two snapshots (1987 and 2000) by conditioning on the occurrence of observations at the first snapshot (1970). The likelihood becomes

\[
L(\theta; \mathbf{I}) = \left\{ \prod_{j \in \xi_{E_j}^{-1}} b_j \times \left\{ \alpha c_j + \sum_{i \in I_{E_j}^i} \beta c_j K(d_{ij}; \kappa) \right\} e^{-q_j^i} \right\} \times \prod_{j \in \xi_U} e^{-q_{r_j}}
\]

where $c_j$ is the suitability of the site $j$; $b_j = 1$ if the colonized site is first reported at the second snapshot and $b_j = p$ if the colonized site is first reported at the third snapshot.

### 2.2 Estimation

We estimate the parameters in a Bayesian framework by considering the joint posterior distribution of the model parameters given the data, which is given by

\[
\pi(\theta | y) \propto L(\theta; y) \pi(\theta),
\]
where $\pi(\theta)$ is joint prior distribution of the parameters and $L(\theta; y)$ is the likelihood function with data $y$ (observed and unobserved data). Markov Chain Monte Carlo (MCMC) methods are employed to estimate the joint posterior. In particular we use the (single-step) Metropolis-Hastings algorithm [3] and update the model parameters sequentially. We assume uniform priors for the parameters. To allow cryptic exposures in the simulation study, following Gibson & Renshaw [4], we adapt the reversible-jump algorithm [5] to the compartmental model setting. More details are given below.

### 2.2.1 Single-step MH algorithm for model parameters

**I Update $\alpha, \beta, \kappa, \sigma^2, \gamma, \eta,$ and $p$ sequentially**

(a) Propose a new parameter value, $\alpha'$, by performing a random-walk on the corresponding current value of the parameter, $\alpha$. Specifically,

$$\alpha' = \alpha + N(0, 1)$$

where $N(0, 1)$ is a random variate drawn from the standard normal distribution. If $\alpha' < 0$, it is rejected and the current value is retained. Note that a scaling constant can be multiplied with $N(0, 1)$ to control the step size and facilitate the convergence.

(b) Accept the proposed $\alpha'$ with probability

$$\frac{L(\theta'; z)}{L(\theta; z)}$$

where $\theta'$ denotes the vector of parameters with $\alpha$ replaced by $\alpha'$. Note that since we have used uniform priors and a symmetric distribution $N(0, 1)$ for the random walk, the acceptance probability reduces to the ratio of likelihoods.

(c) If $\alpha'$ is accepted, replace the current value by $\alpha'$, otherwise retains the current value.

(d) Apply the same algorithm to the remaining parameters sequentially.

**II Update the exposure times $E_j$**

(a) Randomly choose an exposure, $j$, and draw a new exposure time $E'_j$ uniformly between $(0, t_{\text{max}})$, where $t = I_j$ if $j$ has become infectious, otherwise $t = t_{\text{max}}$.

(b) With current data $z$, accept the proposed new exposure time with probability

$$\frac{L(\theta'; z')}{L(\theta; z)}$$

where $z'$ denotes the data with the current exposure time $E_j$ replaced by $E'_j$.

(c) If accepted, replace $E_j$ with $E'_j$, otherwise retain the current value.

### 2.2.2 Reversible jump algorithm for cryptic exposures

Sites that have been exposed but have not yet become infectious are referred to as cryptic exposures. In the simulation study in which a S-E-I-R model is fitted, we allow (unobserved) cryptic exposures and ‘swap’ of sites between the set $\xi_{E \setminus I}$ and $\xi_U$. These operations involve changes of model dimension, which requires the reversible jump algorithm. Adapted from Gibson & Renshaw [4], we apply two operations an addition and a deletion on the set of $\xi_{E \setminus I}$. At each iteration during the MCMC run, each operation is equally likely to be applied.

**I Addition of a cryptic exposure**

(a) Randomly choose a site from $\xi_U$ and move it to the set of $\xi_{E \setminus I}$ and $\xi_E$. Uniformly draw an exposure time $E'_j$ between $(0, t_{\text{max}})$ for this newly added cryptic exposure.

(b) Denote $n_u$ and $n_{E \setminus I}$ as the number of sites in current sets $\xi_U$ and $\xi_{E \setminus I}$ respectively. Accept the proposed new sets and new exposure time with probability

$$\frac{L(\theta'; z')}{L(\theta; z)} \times \frac{n_u \times t_{\text{max}}}{1 + n_{E \setminus I}}$$

where $z'$ denotes the data with the changed sets ($\xi_{E \setminus I}$ and $\xi_E$ and $\xi_U$) and with the current exposure time $E_j$ replaced by $E'_j$. 

II Deletion of a cryptic exposure

(a) Randomly choose a site from $\xi_{E \setminus I}$ (also from $\xi_E$) and move it to the set of $\xi_U$; delete the corresponding $E_j$ accordingly.

(b) Accept the proposed new sets with probability

$$\frac{L(\theta; z')}{L(\theta; z)} \times \frac{n_{E \setminus I}}{(1 + n_u) \times t_{max}}$$

where $z'$ denotes the data with the changed sets ($\xi_{E \setminus I}$ and $\xi_E$ and $\xi_U$) and with the current exposure time $E_j$ being deleted.

2.3 Hypothesis Testing

For the hypothesis testing we use the Anderson-Darling test [6] which has the test statistics

$$A = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \left[ \ln Y(i) + \ln(1 - Y(n-i+1)) \right],$$

where $n$ is the sample size and $Y(i)$ is the $i^{th}$ largest sample. The null distribution of $A$ [7] is used to compute the appropriate p-value using a package [8] available in the statistical software R. In view of our aim of detecting an anticipated mis-match in the tails of the distribution of imputed residuals, we do not adopt the commonly used K-S test which is known to be insensitive to the tail of the distribution.

2.4 Normalization of Spatial Kernel

As the secondary transmission rate and the kernel parameter are often highly correlated, normalization of the spatial kernel is implemented for reducing this correlation. Following [9, 10], we normalize the kernel by dividing $K(d_{ij}; \kappa)$ in the likelihood function by the normalization coefficient

$$S = \sum_{k \neq i, 1 \leq k \leq N}^{N} K(d_{ik}; \kappa).$$

Different normalization schemes may also relate to different assumptions over the dispersal mechanisms [9]. It is noted that in our setting above the spatial kernel describes how much more likely the colonizations of nearby sites are as compared with the colonizations of distant sites (in other words, there is competition among all possible infection links, which conforms to our model assumption), and this is referred to as relative density dependence [9].

3 Model Specifications for Simulated Example & Parameter Estimates

3.1 Model Specifications

Consider a population with size $N = 1000$ whose spatial coordinates are simulated uniformly on a square area $2000 \times 2000$. Assume all individuals in the population are susceptible at time 0 which is also the time of introduction of the force of primary infection $\alpha$, and assume that the epidemic is to be observed up to time $t_{max} = 50$. For spatial kernels, we use an exponentially-bounded kernel (used to simulate the epidemic) $\exp(-\kappa_1 d)$ where $d$ is the distance between the infectious and susceptible, and we have used a Cauchy-form kernel $1/(1 + d/\kappa_2)$ and a power-law kernel $d^{-\kappa_3}$ as the incorrect kernels. The (correct) latent period distribution used to simulate the epidemic is Gamma distribution with mean $\mu$ and variance $\sigma^2$; in fitting an incorrect latent period, an exponential distribution with mean $\mu$ is considered. A Weibull distribution, with scale parameter $\gamma$ and shape parameters $\eta$, is assumed for the infectious period. The locations and times of transitions from class S to class E are assumed to be unobserved, and the transitions from class E to I and the transitions from class I to R are assumed to be observed. However, we allow cryptic exposures (i.e., exposures that have not yet become infectious and remain undetected). Uniform priors, which should be constrained to bounded regions to ensure a proper posterior distribution, are specified for all parameters.
Figure S2: Posterior distributions of model parameters for models fitted to the simulated data (Replicate 1). Dotted lines are the actual values of the parameters used for simulating the epidemic.

3.2 Parameter Estimates

Figure S2 shows the posterior distributions of the model parameters for Case A to Case D from Replicate 1.
4 Model Assumptions for Giant Hogweed Data & Parameter Estimates

4.1 Model Specifications
S-I models are fitted to the giant hogweed data. Similar to the simulated example, we fit an exponentially-bounded kernel \( \exp(-\kappa_1 d) \) (kernel A) and a Cauchy-form kernel \( 1/(1 + d/\kappa_2) \) (kernel B) and a power-law kernel \( d^{-\kappa_3} \) (kernel C) and compare their performance. The second snapshot at 1987 is known to be incomplete due to the insufficient efforts of surveying the sites. We therefore also estimate the probability, \( p \), that a site has been colonized by 1987 but remained unreported given that it is reported at last snapshot (2000). Improper flat priors are used for all parameters.

4.2 Parameter Estimates
Figure S3 shows the posterior distributions of the model parameters from model M1 (kernel A) and model M2 (kernel B) and model M3 (kernel C) with the consideration of suitability.
5 Suitability of Sites

In the consideration of the suitability of sites, we adopt the mean relative suitability estimated from a previous study [10] in which an extensive range of covariates such as average temperature and altitude of sites was taken into account. To reinforce the confidence over the specification of the suitability, we subdivide the sites into three classes (assuming common suitability in each class) according to the estimated suitability from the analysis [10] and estimate the common suitability in each class.

Sites are classified into three classes (Less Favorable, Favorable and Highly Favorable, with suitability \(s_1\), \(s_2\) and \(s_3\) respectively) according to the corresponding suitability estimated from the previous study [10]. We denote \(c_j\) as the estimate of suitability of site \(j\) given in [10]. If \(0 < c_j \leq 0.25\), the site is classified as Less Favorable; if \(0.25 < c_j \leq 0.5\), it is classified as Favorable; if \(0.5 < c_j \leq 1.0\), it is classified as Highly Favorable. We estimate \(s_1\) and \(s_2\) as the suitability relative to \(s_3 = 1\).

We consider fitting a model with kernel A (a ‘better’ kernel as we have shown) and we do not constrain the domain of parameters \(s_1\) and \(s_2\) (i.e., improper flat priors are used). Figure S4 shows the posterior distributions of parameters \(s_1\) and \(s_2\). From the figure, it is evident that \(s_1 < s_2 < s_3 = 1\) and it indicates that the previous estimates of suitability [10] are reliable and broadly consistent with our estimates, which reinforces the confidence of adopting these earlier estimates [10] for our model.

6 Posterior Predictive Checking Based on Spatial Autocorrelation Analysis

In this section, we consider posterior predictive checks based on spatial autocorrelation coefficients which measure spatial dependency among observations, we specifically consider two common measures Moran’s I and Geary’s \(c\) indexes [11]. The epidemic is simulated with kernel \(K(d, \kappa) = d^{-2.8}\). We respectively fit a correct kernel \(K(d, \kappa_2) = d^{-k_2}\) (Model 1) and an incorrect kernel \(K(d, \kappa_1) = \exp(-\kappa_1 d)\) (Model II) to the simulated data. Predictive distributions of the spatial autocorrelation coefficients from Model I and Model II, at three different time points within the observation period are compared to the corresponding measures computed from the actual (simulated) data.

We divide the 2000×2000 square area into \(n = 100\) equally-sized square sub-regions and count the number of sites \(x_i\) in class I or class R in sub-region \(i\) at time points considered for the computation of Moran’s I and Geary’s \(c\) indexes

\[
I = \frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{17}
\]

\[
c = \frac{n-1}{2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - x_j)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{18}
\]

where \(\bar{x}\) is the mean of \(x_i\) over \(n\) sub-regions and \(w_{ij}\) is the spatial weight between sub-region \(i\) and \(j\). There are many ways to define \(w_{ij}\), and here we use the common binary weights in which \(w_{ij} = 1\) if \(i\) is a neighbour of \(j\),
otherwise $w_{ij} = 0$. The definition of neighbour is also open to variations, and here we define that any sub-regions whose centroids are within two sub-region width (400) from the centroid of sub-region $i$ are considered to be the neighbour of $i$. Both indices are computed by using a package `spdep` [12] available in the statistical software R.

Model I represents a long-tail dispersal mechanism and Model II represents a localized dispersal mechanism - this is also illustrated by Figure S6 - and $\pi(P(r_2) < 0.05|y)$ (i.e., the primary measure of degree of model mis-specification in utilizing $r_2$) shows strong evidence against Model II (90%) and no evidence against Model I. Figure S5 shows the predictive distributions of these two indices (at three different time points) obtained from simulating (1,000) epidemics respectively from Model I and Model II with the model parameters drawn from their respective posterior distributions. It can be seen from the figure that the posterior predictive distributions of the spatial autocorrelation indices from both models are broadly consistent with the observed values (i.e., all of the 95% two-sided intervals contain the actual value). This shows that posterior predictive checks based on these indices when only partially observed epidemic are available could be insensitive to the specification of the spatial kernel. Posterior predictive checking has to be computed ‘offline’ - for example, one needs to obtain the posterior distribution or point estimates of model parameters first and compute the required summary statistics based on simulation techniques. Our method, instead, can be easily embedded along with the estimation and by default takes the full posterior distribution of model parameters into account. It is also noted from above that summary statistics based on these spatial autocorrelation measures are subject to variations in definitions which might lead to different conclusions.

**Figure S5:** Posterior predictive distributions of Moran’s $I$ and Geary’s $c$ indexes obtained by simulating 1,000 epidemics from Model I and Model II respectively at time points $T_1 = 25$, $T_2 = 35$ and $T_3 = 45$. The vertical lines represent the observed values computed from the ‘actual’ epidemic.
Figure S6: Estimated spatial kernels from fitting Model I and Model II with kernel parameters set to posterior means. Transmissibility is expressed relative to the amplitude of the respective kernel at $d = 60$ to highlight the difference between two kernels at both short and long distances.

7 Other Supplementary Figures

Figure S7 shows how the Great Britain is subdivided into 65 ring regions for the comparisons of colonized sites within the ring regions predicted by competing models (also see main text).
Figure S7: The partition of Great Britain according to intersection with 65 concentric annuli. Each annulus is centred on the black dot and has width 10km.

Figure S8: Distributions of subsets of imputed ILR which lead to p-values less than 0.05 from M2.
References


