Spatial and Spatio-temporal Epidemiology 41 (2022) 100505

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Original Research

Spatially-heterogeneous embedded stochastic SEIR models for the 2014–2016 Ebola outbreak in West Africa

Kaitlyn Martinez [a,b,*], Grant Brown [c], Stephen Pankavich [a]

[a] Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO, USA
[b] Information Systems and Modeling Group (A-1), Analytics, ALD Global Security, Los Alamos National Laboratory, Los Alamos, NM, USA
[c] Department of Biostatistics, University of Iowa, Iowa City, IA, USA

ARTICLE INFO

Keywords:
Spatial SEIR models
Approximate Bayesian computation
Ebola
Epidemics

ABSTRACT

The dynamics of human infectious diseases are challenging to understand, particularly when a pathogen spreads spatially over a large region. We present a stochastic, spatially-heterogeneous model framework derived from the foundational SEIR compartmental model. These models utilize a graph structure of spatial locations, facilitating mobility via random walks while progressing through disease states, parameterized by the net probability flux between locations. The analysis is bolstered by Approximate Bayesian Computation, by which epidemiological and mobility parameter distributions are estimated, including an empirically adjusted reproductive number, while model structure proposals are compared using Bayes Factors. The utility of this novel class of models is demonstrated through application to the 2014–2016 Ebola outbreak in West Africa. The flexibility of such models, whose complexity may be adjusted as desired, and complementary methods of analysis enable the exploration of various spatial divisions and mobility schema, while maintaining the essential spatiotemporal disease dynamics.

1. Introduction

The data used in epidemic models often features both spatial and temporal variation, but difficulties arise in the incorporation of spatial phenomena within a dynamical model. Analysts therefore often make the assumption of spatial homogeneity throughout the region (or regions) under study, but this is unreasonable when populations are not "well-mixed" and contacts between individuals depend upon underlying spatial structures within their environment, such as transportation routes (Brauer, 2017). Unfortunately, this assumption is often imposed despite its shortcomings because the analysis of spatial models that utilize partial differential equations (PDEs) is particularly challenging, especially when a variety of compartments or stochastic components must be included. Small outbreaks are especially challenging to model and analyze, as similar spatiotemporal conditions can give rise to the progression of an epidemic and alternatively, little to no spread of the disease. Stochastic models aid in such efforts, as this natural uncertainty is encoded directly by a stochastic process (Allen, 2008). While such processes are often employed in spatially-homogeneous models, stochastic models that meaningfully describe the spatial aspects of infection dynamics remain an active and challenging area of research. This is of particular importance, as stochastic models allow practitioners to account for uncertainty in disease spread in addition to parameter estimation. In particular, quantifying knowledge of epidemic parameters (e.g., latent period, variation in infection/contact rate, drivers of infection) can be of critical importance to public health responses, as well as planning efforts for future outbreaks. Stochastic models are especially amenable to analysis via Bayesian inference methods, which allow modelers to include prior information about known infection parameters. Simulation techniques are heavily utilized both in Bayesian and non-Bayesian settings. For example, Approximate Bayesian Computation (ABC) methods (Beaumont, 2010; Beaumont et al., 2009) permit Bayesian inference based primarily on forward stochastic simulations and allow for the comparison of candidate epidemic models through various means, including the application of Bayes Factors (Kass and Raftery, 1995). As with other epidemic parameters, stochastic methods also enable the estimation of various reproductive numbers, as the basic reproductive number employed in deterministic methods does not have a single unambiguous counterpart (Brown et al., 2016).

Consider the Ebola outbreak which spread throughout West Africa between 2014 and 2016 (Center for Disease Control, 2014). While a large number of people were infected during this epidemic, it began and

* Correspondence to: Mail Stop D415, Information Systems and Modeling Group (A-1), Analytics, ALD Global Security, Los Alamos National Laboratory, Los Alamos, NM, USA.
E-mail address: kaitlynm@lanl.gov (K. Martinez).

https://doi.org/10.1016/j.sste.2022.100505
Received 12 January 2021; Received in revised form 3 December 2021; Accepted 21 March 2022
Available online 7 April 2022
1877-5845/Published by Elsevier Ltd.

was propagated due to relatively rare infection and interaction events — both in terms of biological disease processes and the movements of individuals. In an epidemiological sense, the epidemic thus exhibited traits of a "small" outbreak, presenting challenges for both the homogeneous mixing stochastic models and their deterministic counterparts. This outbreak confounded international experts at the time due to its transmission uncertainty (Ladner et al., 2015) and spatial heterogeneity (Dilorenzo, 2014a,b). We propose a novel, spatially heterogeneous, stochastic model to mitigate these challenges.

In the next section, we begin by describing the characteristics of Ebola Virus Disease generally and in West Africa between 2014 and 2016 and providing a survey of previous methods used to model this particular outbreak. Additionally, we provide a general summary of models that describe spatial heterogeneity or stochastic elements in disease transmission. Next, we propose a spatially-heterogeneous extension to a stochastic process embedded within an Susceptible–Exposed–Infectious–Removed (SEIR) compartmental model. This new model employs a graph structure of spatial locations connected by edges over which individuals can relocate via random walk while simultaneously transitioning between disease states. We then describe the statistical methods employed herein, including the Bayesian framework and implementation through ABC, as well as techniques for reproductive number estimation and model comparison. Finally, we demonstrate the utility and efficacy of the proposed class of models by designing and analyzing collections of models with various spatial mobility structures for the Ebola outbreak in West Africa.

2. Background

2.1. Ebola virus disease characteristics

Ebola virus disease (EVD) is a viral hemorrhagic fever affecting humans and primates that is caused by four of the six different species of Ebola virus, and is native to tropical regions of sub-Saharan Africa (The World Health Organization, 2016a; WHO Ebola Response Team, 2014). The Ebola virus first spreads from a natural animal reservoir to a proximal human population. The precise natural reservoir is unknown; however, it has been shown that bats are asymptomatic carriers and have been implicated in earlier outbreaks of EVD (Center for Disease Control, 2015). After its initial introduction from the reservoir, the virus spreads within a human population via interaction with infected bodily fluids such as blood, vomit, feces, mucus, urine, and semen, either during direct physical contact with an infectious individual or indirect contact with contaminated surfaces and materials. Symptoms typically manifest 2–21 days after exposure and usually last 6–16 days before recovery or death (The World Health Organization, 2016a). Recovered individuals develop non-permanent antibodies which protect them from the virus for up to ten years. The symptoms of EVD begin as extreme flu-like symptoms, and progress to include rash, vomiting, and diarrhea, internal and external bleeding, and severe dehydration (The World Health Organization, 2016a). The mortality rate of EVD ranges from 25% to 90%, depending on the availability of supportive treatment focusing on rehydration and blood pressure stabilization (The World Health Organization, 2016b).

The largest and most deadly outbreak of EVD to date officially began on December 26, 2013 in West Africa, causing 28,646 cases and 11,323 deaths worldwide (Center for Disease Control, 2014). The epicenter of this epidemic occurred in the three African countries of Guinea, Sierra Leone, and Liberia. Many factors, including the extreme poverty, inadequate and dysfunctional healthcare systems, distrust of government, and the introduction of cases into dense urban areas, contributed to the massive spread of the virus and challenges in containing it (WHO Ebola Response Team, 2014). Additionally, traditional religious burial practices involving widespread, direct contact with deceased individuals were identified as a significant transmission pathway.

2.2. Previous models and methods

Due to the complexity of transmission pathways and the heterogeneity of the spatial spread, modeling efforts for this particular outbreak have been quite diverse. Attempts to model the epidemic began in real time as a response measure to the crisis, and have continued into the present to better prepare for future outbreaks. A number of previous models for Ebola are compartmental in nature, breaking the disease process into a set of discrete states or "compartments". These include both deterministic mathematical models (Mamo and Koya, 2015; Rivers et al., 2014) and stochastic models (Lekone and Finkenstädt, 2006), ranging from a standard susceptible–exposed–infectious–removed (SEIR) compartmental structure to a broad range of disease state subdivisions of the population. These additional disease states include infectious classes representing other pathways of disease spread (e.g. hospitalized, deceased), distinctions between reported and unreported cases, and further divisions of the incubation period or the particular type of care or hospitalization (Wong et al., 2017; Diaz et al., 2018). Generally, as the complexity of the compartmental model is increased, additional assumptions regarding the homogeneous spatial mixing of populations are necessary. This assumption was typically addressed not via implementation of spatial heterogeneity, but through stratified population structures that further split populations into more refined categories of age, risk, or community structures (Kiskowski, 2014; Fast et al., 2015; Agusto et al., 2015). Other efforts addressed the spatial heterogeneity by building models at the finest spatial resolution available — districts in Guinea, counties in Liberia, and provinces in Sierra Leone (Santermans et al., 2016; Pell et al., 2016a). Despite this high resolution of spatial discretization, models were constructed for an individual location, i.e. a single region or country, ignoring the inherent interactions amongst spatial locations. Most existing models, even highly-complex compartmental ones, struggle to address the spatial heterogeneity inherent within the epidemic, and specifically fail to capture the spread of Ebola within Guinea (Rivers et al., 2014; Diaz et al., 2018). This is particularly problematic, as the outbreak began in Guinea and, for the first 12 weeks, is observed only within Guinea's borders.

Many models that have incorporated interactions between spatial locations in both mathematical and statistical studies consider only the mobility of the pathogen and not of individuals themselves, wherein infection can spread across borders, while individuals remain static in their locality. Gravity models constitute one such example and have been implemented at a variety of spatial scales to explore the spread of this Ebola epidemic (Kramer et al., 2016; Valdez et al., 2015; D'Silva and Eisenberg, 2017). More generally, in the statistical epidemiology community, the gravity model concept is generalized to characterize any spatial auto-correlation structure which could explain the spread of disease using spatial or conditional auto-regressive models (SAR/CAR) (Porter and Oleson, 2014). One of the few modeling frameworks that facilitates the movement of individuals as a source of disease spread is that of Agent-based Models (ABMs) (Merler et al., 2015). ABMs are often computationally constrained, however, and tend to focus on relatively small populations as a result (Hunter et al., 2017; Perez and Dragicevic, 2009). In general, the development of robust stochastic patch models that facilitate statistical parameter estimation and data fitting with the addition of mobility parameters (Pell et al., 2016b; McCormack and Allen, 2007; Lahodny and Allen, 2013) remains a challenge. The fully stochastic inferential modeling framework presented herein enables the incorporation of individual as well as pathogenic mobility while retaining a simple disease structure that fosters both parameter identification and the feasible use of model selection techniques, as we shall demonstrate through application to the Ebola outbreak in West Africa that began in 2014.

3. Methods

3.1. Spatially heterogeneous embedded stochastic SEIR model

We consider a graph G with k nodes, each representing a spatial location. Population mobility on this graph is implemented via transition probabilities defined on the edges according to a Markov probability matrix

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}$$

where p_{ij} represents the probability of transitioning from node i to node j. Individuals can then move between locations via a random walk on this graph. This structure is a Markov chain (Modica and Poggiolini, 2012), as an individual's movement at the next time step only depends on their current position. If $p_{ij} = p_{ji} = 0$ then no edge connects node i to node j. While it is feasible to allow G to be a directed graph and enforce unidirectional movement along an edge, we assume that there is some level of continuous movement in both directions along a defined edge. Thus, for every edge either $p_{ij} = p_{ji} = 0$ or $p_{ij}, p_{ji} \neq 0$. With the edges and transition probabilities defined, a single flux term $\rho_{ij} = p_{ij} - p_{ji}$, which we will refer to as the overall probability flux, is defined for each existing edge. This quantity is necessary as the individual transition probabilities are not structurally identifiable (Bellman and Åström, 1970), and are also often not identifiable in practice and the former has been directly verified via simulation.

The overall probability flux is intuitively understood as the net exchange of individuals between two spatial locations and will be practically identifiable for a general Markov random walk on G. Additionally, our utilization of a single flux term only requires the identification of $2k(k-1)$ parameters rather than $4k(k-1)$. Using this parameterization, we fix the less dominant probability of the pair (p_{ij}, p_{ji}) and add the estimated probability flux to uniquely determine the second in the pair when necessary.

This random structure assumes a constant probability of individuals moving between two locations, which is valid assuming constant mobility. However, such an assumption may fail to be realistic. In the context of the spread of an infectious disease, one type of non-constant movement, known as a "sparking" event, can be extremely important to understanding the spatial spread of the disease. Sparking events are characterized by situations in which two locations with typically negligible mutual mobility experience a rare transmission event that introduces new cases to a previously unaffected region. Such events have a large impact on the scope and timing of an epidemic, and can be vital to understanding the spatial spread of the disease.

To facilitate the incorporation of such sparking events, the proposed model will utilize a matrix $\boldsymbol{\mathcal{I}}$ of spatial and temporal indicator functions. We first define a time interval T centered at a fixed time $t^* > 0$ and possessing a specified duration of 2τ, so that $T = [t^* - \tau, t^* + \tau]$. Then, the temporal indicator function on this interval is denoted by

$$I_T(t) = \begin{cases} 0, & t \notin T \\ 1, & t \in T \end{cases}$$

and for $p^*_{ij} \in (0,1)$ the corresponding spatio-temporal matrix $\boldsymbol{\mathcal{I}}$ is defined entrywise by

$$\boldsymbol{\mathcal{I}}_{ij}(t) = \begin{cases} 0 & (i,j) \notin \mathcal{H} \\ p^*_{ij} I_T(t) & (i,j) \in \mathcal{H} \end{cases} \tag{1}$$

where $\mathcal{H} \subset \mathcal{G}$ is a specified edge subset of the graph. When added to $\boldsymbol{P}$, this matrix will implement the short term probability p^*_{ij} of population mobility between nodes i and j with $(i,j) \in \mathcal{H}$ that otherwise possess no mobility.

With movement between nodes determined by the probabilities $\hat{p}_{ij}$ in $\hat{\boldsymbol{P}} = \boldsymbol{P} + \boldsymbol{\mathcal{I}}$ at each time step, we can track individuals that move between nodes from each compartment of the disease process. Consider individuals in a particular disease class X that is mobile. Then $x_{ij}(t)$, the number of individuals in compartment X leaving node i for node j at time t, can be drawn from multinomial distributions over all k locations. Hence, we let

$$[x_{i1}(t), x_{i2}(t), \ldots, x_{ik}(t)] \sim \text{Multi}\left(\chi_i(t), [\hat{p}_{i1}, \hat{p}_{i2}, \ldots, \hat{p}_{ik}]\right)$$

where $\chi_i(t)$ is the number of individuals within compartment X and location i that are available to move at time t. To understand the number of mobile individuals at each node and within each disease class requires a simplifying assumption. To help us preserve positivity, we will assume that within any time step, a person first completes the disease process stochastic step, e.g. moving from the susceptible to the exposed state, prior to being considered mobile if appropriate. This imposes a reality in which an individual cannot simultaneously move locations and disease states. The multinomial draws produce k vectors for the ith spatial location, generating matrices of mobile X individuals given by

$$\boldsymbol{M}_X = \begin{pmatrix} x_{11}(t) & x_{12}(t) & \ldots & x_{1k}(t) \\ x_{21}(t) & x_{22}(t) & \ldots & x_{2k}(t) \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1}(t) & x_{k2}(t) & \ldots & x_{kk}(t) \end{pmatrix}.$$

We then calculate the net total individuals in disease class X entering (or remaining within) spatial location i from all other locations, denoted $N_i^X(t)$, by summing the columns of $\boldsymbol{M}_X$, namely

$$N_i^X(t) = \sum_{\ell=1}^{k} x_{\ell i}(t).$$

With population movement defined in this way, the standard stochastic spatially homogeneous SEIR model (Lekone and Finkenstädt, 2006) can be extended to a spatially heterogeneous model describing the spread of an SEIR disease process over k nodes, given by

$$\left.\begin{aligned} S_i(t+\Delta t) &= N_i^S(t) - B_i(t) \\ E_i(t+\Delta t) &= N_i^E(t) + B_i(t) - C_i(t) \\ I_i(t+\Delta t) &= N_i^I(t) + C_i(t) - D_i(t) \\ R_i(t+\Delta t) &= D_i(t). \end{aligned}\right\} \tag{2}$$

This model assumes there are four disease states: susceptible, exposed, infectious, and removed (recovered, deceased, and immune) populations at time t and spatial location i denoted by $S_i(t), E_i(t), I_i(t)$, and $R_i(t)$, respectively. The susceptible population is, in this case the entire population of the country (or other relevant spatial region), as no vaccine or direct treatment methods to prevent the spread of the disease were available prior to the onset of the epidemic. The newly exposed, $B_i(t)$, infectious, $C_i(t)$, and removed, $D_i(t)$, populations at time t and location i are defined by

$$\begin{aligned} B_i(t) &\sim \text{Bin}(S_i(t), P_i(t)) \\ C_i(t) &\sim \text{Bin}(E_i(t), p_C) \\ D_i(t) &\sim \text{Bin}(I_i(t), p_D) \end{aligned}$$

where

$$P_i(t) = 1 - \exp\left[\frac{-\beta_i(t)}{N_i(t)} I_i(t) \Delta t\right], \tag{3}$$

$$p_C = 1 - \exp(-\alpha \Delta t), \quad \text{and} \quad p_D = 1 - \exp(-\gamma \Delta t).$$

The binomial probabilities for each transition population depend on the biological parameters in the foundational deterministic SEIR model (Brauer and Castillo-Chávez, 2001) within which the stochastic model is embedded. In general, members of the susceptible population may contact an infectious individual with probability μ_1, and each interaction has an associated probability of transmission, given by μ_2. These probabilities are then combined to form $\beta = \mu_1 \mu_2$, which represents the characteristic infection rate. We will consider the situation in which

$\beta = \beta_i(t)$ is a function of time and spatial location. The exposed population represents individuals experiencing the biological latent period (on average, $\frac{1}{\alpha}$ days) during which they have been infected but present no symptoms and are not infectious. After this time, exposed individuals transition into the infectious compartment, from which they are ultimately removed after an average period of $\frac{1}{\gamma}$ days via either recovery or death. Both α and γ are fixed, as the latent and infectious periods typically do not change over the course of an epidemic. The net mobile susceptible, exposed, and infectious individuals are defined by

$$N_i^S(t) = \sum_{\ell=1}^{k} s_{\ell i}(t), \qquad N_i^E(t) = \sum_{\ell=1}^{k} e_{\ell i}(t), \qquad N_i^I(t) = \sum_{\ell=1}^{k} \imath_{\ell i}(t).$$

Additionally, the multinomial draws for mobile individuals are given by

$$[s_{i1}(t), s_{i2}(t), \dots, s_{ik}(t)] \sim \text{Multi}(S_i(t) - B_i(t), [\hat{p}_{i1}, \hat{p}_{i2}, \dots, \hat{p}_{ik}])$$
$$[e_{i1}(t), e_{i2}(t), \dots, e_{ik}(t)] \sim \text{Multi}(E_i(t) + B_i(t) - C_i(t), [\hat{p}_{i1}, \hat{p}_{i2}, \dots, \hat{p}_{ik}])$$
$$[\imath_{i1}(t), \imath_{i2}(t), \dots, \imath_{ik}(t)] \sim \text{Multi}(I_i(t) + C_i(t) - D_i(t), [\hat{p}_{i1}, \hat{p}_{i2}, \dots, \hat{p}_{ik}])$$

This structure assumes that all individuals possess the same mobility characteristic probabilities. Depending on the disease, this assumption should be addressed considering that some diseases render infectious individuals much less mobile or completely immobile. Of course, the mobility of removed individuals does not impact the system, as the SEIR framework described above assumes immunity (or death), rendering mobility of recovered individuals irrelevant to the disease process; hence, we will not include such a feature in the model.

3.2. BayesIan inference for model selection and parameter estimation

The above framework defines a rich class of models from which an appropriately calibrated model for a particular disease data set can be identified. There are two primary model design decisions that must be made. First, the number of spatial locations k will be informed by the data; however, it is valuable to explore a variety of possible spatial divisions, i.e. values of k, as inherent spatial divisions may be uncovered that are not apparent from the data. The second decision regards the mobility structure between the k spatial locations. Once $k \geq 3$ is determined there exist $\frac{k(k-1)}{2}$ possible connections between nodes, resulting in $\sum_{i=1}^{k} \binom{k}{i} 2^{\binom{i}{2}}$ different spatial connectivity structures (Baylis et al., 2003). Additionally if "sparking" events are suspected the model space expands for each "sparking" event explored. Selection of the appropriate model can be determined primarily by the modeler or via a systematic approach. Independent of the choice of model structure, most of the inherent parameters are unknown and must be estimated.

Given the potentially large space of models to consider, Bayesian inference provides a natural approach to deal with model uncertainty in addition to parameter estimation. In addition to providing a mechanism for model comparison, parameter estimation via Bayesian Inference formally quantifies parameter uncertainty rather than simply providing point estimates. Additionally, prior knowledge about the biological process or the model can be directly encoded in the form of informative priors, which can help constrain highly flexible, dynamic models for epidemics to biologically plausible configurations. The most common method for Bayesian Inference is Markov Chain Monte Carlo (MCMC) (Gilks et al., 1995). However, MCMC presents difficulties as models become more complex, as it requires repeated evaluation of a complex likelihood and can explore large parameter spaces slowly. In contrast, Approximate Bayesian Computation (ABC) can be applied to any process with a defined forward model, without requiring evaluation of the likelihood (Beaumont, 2010). ABC provides approximate posteriors based on simulation, by comparing replicate data sets Y^* generated by candidate parameter values θ^* to the observed data Y. Many extensions of ABC exist, including sequential techniques designed to make more efficient parameter proposals (Beaumont, 2010). In this work, we apply the ABC-SMC method detailed in Beaumont et al. (2009).

All ABC methods require a metric to compare simulated and observed data, which generally relies on the use of summary statistics and selected norms. The choice of summary statistics and norms depends on the nature and dimension of the problem at hand and is an ongoing area of research (Beaumont et al., 2009; Beaumont, 2010). In this case, with epidemic data there were various dynamics to consider when selecting summaries and norms. It is highly unlikely that the data is error free due to the measurement challenges that arise early in clinical surveillance of an ongoing epidemic. In fact, it is nearly guaranteed that cases will be unreported or recorded incorrectly, and the precise timing of cases is always suspect. Comparison of replicate data sets in an ABC context should therefore avoid over-reliance on granular spatiotemporal indices. While simple and often effective, point-wise norms like a standard sum of squares or $\mathcal{L}_2$ error metric over the entire spatial time series could inflate the effects of reporting artifacts if an appropriately tailored data model is not available. Here, we focus on three particular measures which are distinctive and determine the qualitative structure of the epidemic: infection intensity, timing, and duration. These characteristics are defined at each location $i = 1, \dots, k$ by:

1. The number of cases observed during the peak week, given by $PC_i = \max_{t \geq 0} W_i(t)$, where $W_i(t) = \sum_{s=t_0}^{t_1} C_i(s)$ is the weekly number of cases and $t_0 = 7\left\lfloor \frac{t}{7} \right\rfloor$ and $t_1 = t_0 + 6$ are the first and last days of the week containing day t of the outbreak.
2. The week in which the maximum number of cases are observed (i.e., the peak week), defined as $PW_i = \frac{1}{7} \min\{t \geq 0 : W_i(t) = PC_i\}$.
3. The end of the epidemic can be defined independent of location by $T = \max\{t \geq 0 : C_i(s) > 0 \text{ for all } s \in [0, t] \text{ for any } i = 1, \dots, k\}$.

Using these quantities, two relative norms which are spatial in nature $\mathcal{N}_i^n$, $n = 1, 2$ and one relative non-spatial norm $\mathcal{N}^3$, are defined, by which the sample data, denoted by $*$, can be compared to the established data set, namely

$$\mathcal{N}_i^1 = \frac{|PC_i - PC_i^*|}{PC_i^*}, \qquad \mathcal{N}_i^2 = \frac{|PW_i - PW_i^*|}{PW_i^*}, \qquad \mathcal{N}^3 = \frac{|T - T^*|}{T}^*. \tag{4}$$

Thus, for each simulated epidemic a total of $2k + 1$ norm values $\mathcal{N}_i^n$ for $i = 1, \dots, k$ and $n = 1, 2$ and $\mathcal{N}^3$ can be calculated and used to determine the acceptance or rejection of parameter proposals in the ABC-SMC algorithm.

In addition, an ABC-SMC framework allows for the direct comparison of candidate models via the computation of approximate Bayes Factors (Kass and Raftery, 1995). These terms measure the relative posterior evidence in favor of one model over another. It can be shown that when using ABC-SMC the Bayes Factor is merely the ratio of probabilities of each model given the data x or

$$B_{ij} = \frac{P(m_i|x)}{P(m_j|x)}.$$

These values can be well-approximated by the acceptance rates for each model. This property of ABC methods allows different models to be easily compared and ranked in relation to one another, at any tolerance value ϵ for which an acceptance rate is defined, by using an established categorization for the values of Bayes Factors. With this, not only can various model pairs be ranked, but we can also precisely determine the degree to which one model, relative to another, is more effective at producing acceptable simulations at a particular threshold.

It is also valuable to calculate and analyze the basic reproduction number and other associated quantitative threshold values. The basic reproduction number is intuitively defined as the expected number of secondary infected cases directly generated by the introduction of

a new primary infected case (Brauer and Castillo-Chávez, 2001; Van den Driessche and Watmough, 2002). For the classical, non-spatial, deterministic SEIR model, with constant infection rate, the associated mathematical quantity is defined by $R_0 = \frac{\beta}{\gamma}$. Instead of this value, we will calculate and analyze an empirically adjusted reproduction number for stochastic compartmental models, denoted by $R^{(ea)}$, which measures the number of expected secondary infections per infectious individual in a particular location and time (Brown et al., 2016). For the proposed model, an empirically adjusted reproduction number for each spatial location i and time t is defined as:

$$R_i^{(ea)}(t) = \sum_{\tau=t}^{T_i} G_i(\tau) e^{-\gamma(\tau - t)} \tag{5}$$

where

$$G_i(t) = \begin{cases} \dfrac{S_i(t) P_i(t)}{I_i(t)}, & \text{if } I_i(t) > 0 \\ 0, & \text{if } I_i(t) = 0. \end{cases}$$

This metric is easily calculated in the proposed framework, and captures epidemic thresholding behavior in a spatiotemporally heterogeneous way.

4. Results & discussion

As an application of the newly defined class of embedded stochastic models and the Bayesian analysis techniques for both parameter estimation and model selection, we explore modeling the Ebola epidemic in West Africa between 2014 and 2016. The particular spatio-temporal dynamics demonstrated over the course of this epidemic present many modeling challenges. Specifically, the diverse pathways of transmission add high levels of uncertainty to the time course of infection. This uncertainty has previously been addressed via complex compartmental structures incorporating numerous infection and transmission parameters. However, this particular epidemic spread throughout three separate countries in a distinctly heterogeneous way. As a result, it is possible that some of the difficult-to-model characteristics of this epidemic arose in part from spatial heterogeneity. The flexibility and breadth of the embedded stochastic model class defined in the Methods section will aid in examining various hypotheses about the nature of the spatial spread of disease in this particular epidemic. We explore and compare a range of simplified compartmental models, with a shared spatial structure and differing mobility frameworks constrained only by the geography of the region, to examine the role that spatial heterogeneity plays in the infection spread. Using a simple SEIR compartmental structure constrains the biological parameter space to more easily accommodate the additional mobility parameters. Furthermore, this structure facilitates examination of the hypothesis that spatial heterogeneity contributed more to epidemic dynamics than previously explored. Finally, while the class of models allows for highly complex and large spatial structures with many nodes, increasing spatial complexity possesses a high computational cost. The ultimate goal is to identify and elucidate the simplest spatial model that achieves a desired level of accuracy and provides reasonable explanations for the spread of the epidemic.

4.1. Data: Ebola in West Africa between 2014 and 2016

The data used for this study was produced by the World Health Organization (WHO), which tracked the Ebola epidemic in the West Africa nations of Sierra Leone, Liberia, and Guinea between 2014 and 2016. The data is publicly available and can be downloaded from the WHO (The World Health Organization, 2020b). The sum of confirmed and probable cases obtained from the Patient Database (PD) was used in lieu of the counts from the Situation Report (SR) because the SR does not measure the initial trajectory of the epidemic particularly well, as the epidemic had not gained enough attention, and so was not described by a weekly situation report until later in its time course. The raw data is discretized to 15 counties in Liberia, 33 prefectures in Guinea, and 16 districts in Sierra Leone which can be aggregated to coarser spatial resolutions. The availability of weekly, high spatial resolution data for each country facilitates the exploration of a variety of spatial structures to determine an appropriate and computationally feasible representation of the spatial heterogeneity. While the epidemic curves in Liberia and Guinea are typical of epidemic data, the super exponential rise in the Liberia data indicates reporting errors (The World Health Organization, 2020a).

4.2. Model space and selection

Within the framework proposed in Methods, a vast array of model structures exist to describe this epidemic. Though a basic, country-level implementation of the model would consider $k = 3$, with three spatial locations mimicking the existing country boundaries, this approach was insufficient to capture the early dynamics of the epidemic in Guinea (see Supplemental Information). Given the heterogeneity within the national units, particularly in Guinea, (see Figure S5), a collection of four location models was considered. The six schema illustrated within Fig. 1 all consider a single node to represent each of Liberia and Sierra Leone, while Guinea is represented by two distinct regions, Guinea 1 constituting the eastern (predominantly, rural) and Guinea 2, western (more urban) counties, respectively. The nature of mobility between the four locations varies between the six model structures, depending on the number and type of connections between the four nodes. Schemes 1 and 1i represent the fully connected graph, containing all five edges informed by the shared borders between the locations, while Schemes 3 and 3i represent a minimally connected graph which links all four locations with only three edges. The type of edge connecting the two Guinea nodes is altered in Schemes 1i, 2i, and 3i, where a sparking event, implemented via the temporary edge, $\rho_{34} I_T(t)$, is explored.

While these are not the only spatial connectivity schemes possible in the model space, here we limit the scope of the models explored by using intuition gained through analysis of the geographic distribution of cases and the timing of the initial wave of infection (Figure S5) as well as a review of the literature (D'Silva and Eisenberg, 2017; Rivers et al., 2014; Diaz et al., 2018). The constraints on the model scope stem from the initial progression of the disease across the four regions. The epidemic started in the eastern and rural Guinea 1, and next was observed in the western, urban Guinea 2. However, we conjecture that after this initial transfer, the western epidemic spreads primarily independent of the outbreak in the eastern region that spread to Liberia and Sierra Leone. Indeed this concurrent spread is observed in other modeling efforts (Kramer et al., 2016). There is effectively a physical separation between the two regions aligned with the least populous areas of the country which could serve to severely dampen population mixing between the two regions. Hence, this study serves as a demonstration of the application of the model class and the Bayesian inference methods for model selection and parameter estimation, rather than as a comprehensive exploration of the entire model space.

Among the explored spatial structures there are aspects of the model and Bayesian inference process that remain consistent due to the particular application and computational implementation. Primarily, as an infectious Ebola individual is extremely unlikely to be mobile due to the debilitating nature of the symptoms, our implementation of the model does not consider a mobile infectious class. Thus, for any value of $i = 1, \ldots, k$ the infectious population at time $t + \Delta t$ will be defined as $I_i(t + \Delta t) = I_i(t) + C_i(t) - D_i(t)$ with each population defined as in (2). Though the data is observed weekly, the model will consider a constant time step as $\Delta t = 1$ day. The general model structure accommodates a temporally-dependent and/or spatially-dependent β by redefining it as a function of t and/or i, respectively. For this application it is appropriate to consider a spatiotemporally-dependent infection rate

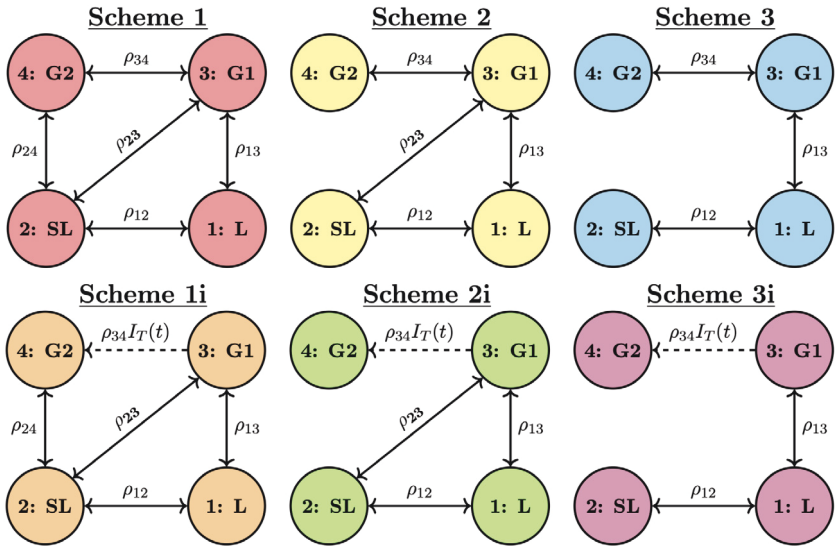

Fig. 1. Four Node Models.
Four node model schema with differing connectivity structures informed by geographical boundaries and timing of epidemic spread. Schemes 1i, 2i, and 3i include a time indicator function from Guinea 1 to Guinea 2 informed by geographical boundaries and timing of epidemic spread.

Table 1
Standard priors of model parameters.

Parameter	Prior
γ	$\gamma = 1/10$
α	$\alpha \sim \Gamma(4, 20)$
ρ_{ij}	$\rho_{ij} \sim \beta(2,2) - .5$
p_{ij}	If $\rho_{ij} > 0$ $p_{ji} \sim \mathrm{U}(0, \frac{(1-\rho_{ij})}{k})$, $p_{ij} = p_{ji} + \rho_{ij}$ If $\rho_{ij} < 0$ $p_{ij} \sim \mathrm{U}(0, \frac{(1-\lvert\rho_{ij}\rvert)}{k})$, $p_{ji} = p_{ij} + \lvert\rho_{ij}\rvert$
$b^i_{0,1,2}$	$(b^i_0, b^i_1, b^i_2) \sim N(\hat{\mu}, 3I\hat{\sigma}^2)$
T	$T = [t^* - \tau, t^* + \tau]$ $t^* = 70$ days and $\tau = 7$ days

The standard priors used for each parameter in the model are detailed. Additional details are included to explain non-intuitive aspects of the priors used.

$\beta_i(t)$ to approximate the effects of interventions that reduce the overall infection strength, as well as the inherent spatial heterogeneity of the spread. To simplify the parameter estimation, rather than estimating each daily infection rate we will estimate a particular functional form of the infection rate, namely

$$\beta_i(t) = \exp\left(b^i_0 + b^i_1 t + b^i_2 t^2\right).$$

With this model definition and the associated simplifying assumptions, the parameters that require estimation are a single latent period parameter α, the $3k$ infection rate functional coefficients b^i_0, b^i_1, b^i_2, and each ρ_{ij} that is defined for a given spatial connectivity structure shown in Fig. 1. As a given individual is infectious for 7–14 days (6–16 days if untreated), this parameter is estimated in the model by the average infectious period $\frac{1}{\gamma}$. This value is well known and simpler to measure because Ebola symptoms are clear and violent. For these reasons, we will treat it as known in our modeling efforts, and it will be fixed by $\gamma = \frac{1}{10}$. The mean latent period $1/\alpha$ has been estimated between 2 and 21 days, but a 4–10 day time span is typical. That being said, there is some evidence that the latent period for this particular epidemic was significantly longer, ranging up to 30 or 40 days. One of the reasons for this discrepancy is that the recommended quarantine period for individuals exposed to Ebola was 21 days (Eichner et al., 2011; Haas, 2014; Velásquez et al., 2015), but there were cases in which an individual who had been originally deemed uninfected later exhibited symptoms after being quarantined. To address this while still reflecting the higher likelihood of the standard latency periods, a gamma distribution reflecting a mean latency period of 5 days with a suitably large standard deviation was selected to serve as a prior for α. Finally, the spatial connectivity schemes possess consistent initial conditions defined by initializing the susceptible population in all three locations to be the total population in each country with the exception of two initial infectious individuals in Guinea. The priors are summarized in Table 1 for each location and connecting edge, which is indicated by each of the graph structures shown in Fig. 1. A relatively vague prior centered at zero and ranging from $[-0.5, 0.5]$ was chosen for ρ_{ij} in order to prevent extremely large differences between the movement probabilities p_{ij} and p_{ji}. This introduces a constraint, $|\rho_{ij}| \leq 0.5$, which prevents invalid values of the probabilities p_{ij} and p_{ji}. Following the parameterization detailed in Section 3.1 a proposed ρ_{ij} is drawn from its prior and then, depending on its sign, either p_{ij} or p_{ji} is proposed and the remaining parameter is calculated using the definition $\rho_{ij} = p_{ij} - p_{ji}$.

Following the ABC-SMC algorithm, a decreasing sequence of acceptance thresholds

$$\vec{\epsilon} = (0.75, 0.65, 0.55, 0.50, 0.45, 0.40, 0.35, 0.30, 0.25, 0.20)$$

was implemented using the norms given by (4). All norms were required to be satisfied in all spatial locations simultaneously to generate an acceptance. For the purposes of this data set, $T_{\max} = 123$ weeks is the final reporting date. Each stage of ABC-SMC was simulated until $N \geq 1000$ acceptances were observed. Detailed information regarding the implementation of the ABC-SMC algorithm is available in Supplemental Information. All six proposed models generated a sufficient number of acceptances at the final threshold, though clear differences in the suitability of the spatial structures become clear. Table 2 summarizes the results for these structures with the final stage of ABC-SMC at $\epsilon = 0.20$.

Model selection via Bayes Factors was applied upon reaching the final threshold, as at less stringent thresholds it is not clear which model(s) can generate acceptances at more restrictive thresholds. This is abundantly clear when considering the evolution of the Bayes factors over the ϵ thresholds for the six model structures shown in Fig. 2. The stark contrast of the evolution of the Bayes factors for Schemes 1 and 2i

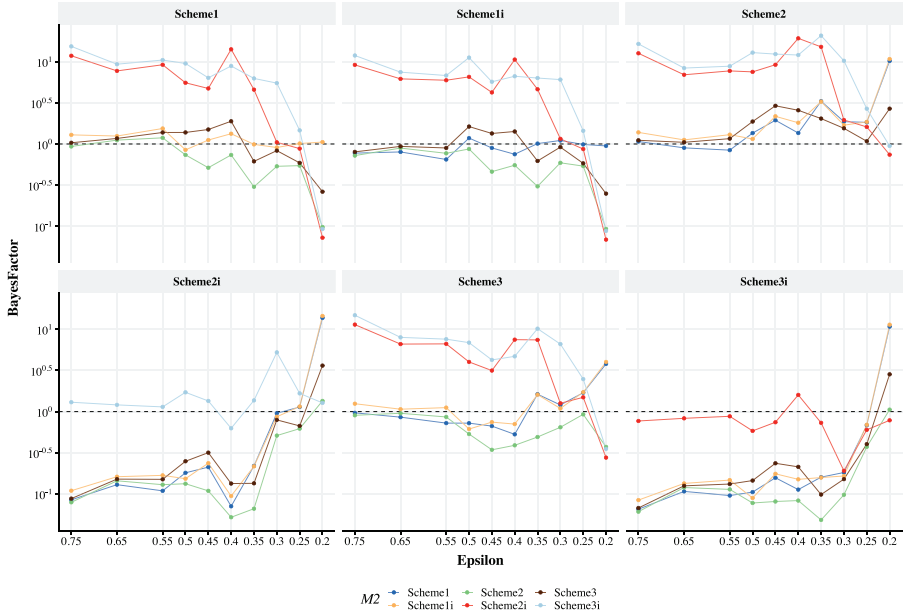

Fig. 2. Bayes Factors.
Each facet shows the Bayes Factors calculated for a given model by dividing the acceptance rate of the titular model (Model A) by the modeled indicated by the color scheme (Model B) for each ϵ threshold used for ABC-SMC. The Bayes Factors are shown on a log scale and values greater than one indicate positive evidence for Model A while values less than one indicate evidence in favor of Model B.

Table 2
ABC-SMC results for six schemes.

Scheme	Epsilon	Simulations	Acceptances	Acceptance rate
Scheme2i	0.20	9.35×10^9	1011	1.08×10^{-7}
Scheme2	0.20	5.65×10^9	448	7.94×10^{-8}
Scheme3i	0.20	9.78×10^9	739	7.55×10^{-8}
Scheme3	0.20	2.74×10^9	82	2.99×10^{-8}
Scheme1	0.20	3.17×10^9	22	6.94×10^{-9}
Scheme1i	0.20	2.09×10^9	12	5.74×10^{-9}

The total simulations, acceptances and acceptance rate are reported for each of the six spatial mobility schemes when ABC-SMC uses a threshold of $\epsilon = 0.20$

are demonstrative of this uncertainty early in the ABC-SMC Algorithm. At the final threshold, $\epsilon = 0.20$, the Bayes factor matrix **B** is generated by dividing the acceptance rate for the model indicated by the column by the acceptance rate for the model indicated by the row, and is given by

$$\mathbf{B} = \begin{array}{c} \\ 1 \\ 1i \\ 2 \\ 2i \\ 3 \\ 3i \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 1i & 2 & 2i & 3 & 3i \end{array} \\ \begin{pmatrix} - & 0.83 & 10.71 & 14.50 & 3.77 & 10.45 \\ 1.21 & - & 12.97 & 17.55 & 4.57 & 12.65 \\ 0.09 & 0.08 & - & 1.35 & 0.35 & 0.98 \\ 0.07 & 0.06 & 0.74 & - & 0.26 & 0.72 \\ 0.27 & 0.22 & 2.84 & 3.84 & - & 2.77 \\ 0.10 & 0.08 & 1.03 & 1.39 & 0.36 & - \end{pmatrix} \end{array}. \quad (6)$$

After evaluating all pairwise Bayes Factor comparisons, Scheme 2i is the model which is favored among the candidates. This evidence is substantial in the case of Scheme 2i compared to Scheme 3, strong when compared to either Scheme 1 or 1i, but only slight when compared to Scheme 2 and Scheme 3i. In general, Schemes 2, 2i, and 3i stand apart from the other three models as displaying evidence in the data to support them over the others. Additionally, all spatial structures that included the sparking event between Guinea 1 and Guinea 2, in the form of the indicator function all have Bayes Factors greater than 1 when compared to their non-indicator parallel model.

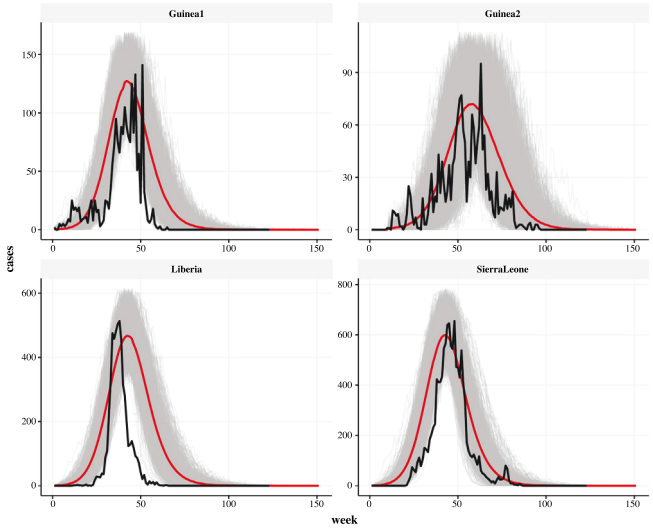

Fig. 3. Accepted Epidemics for Scheme 2i.
The accepted epidemics for the Scheme 2i four node model at the SMC threshold of $\epsilon = 0.20$ are plotted (gray) with the observed data (black) and the mean simulation (red).

4.3. Analysis of optimal structure (Scheme 2i)

For the purposes of demonstration of the results we will limit further analysis to only those of Scheme 2i. Fig. 3 displays the evolution of accepted epidemics $\epsilon = 0.20$. The division of Guinea into two regions and the reduction of edges connecting the western region of Guinea (Guinea 2) to the remaining graph structure aids in modeling the early weeks of the epidemic within that country. In fact, the three highest-ranked model structures (Schemes 2, 2i, and 3i) demonstrate this capability, further solidifying the hypothesis of disconnected concurrent spread in Guinea, as well as the occurrence of a sparking event as the mode of spread to the urban region of Guinea. Even so, the nature of the super exponential growth applied at the entire country level

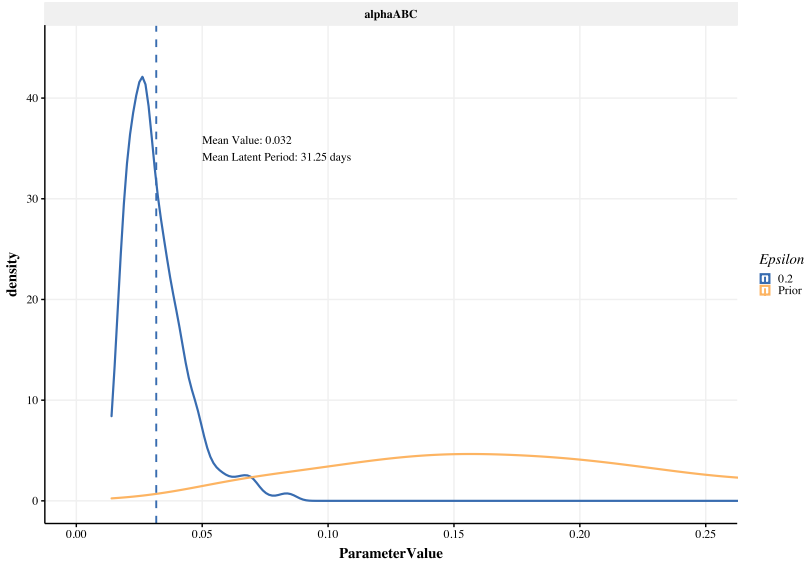

Fig. 4. Posterior Distribution of the Latent Period Parameter α.
The posterior distribution (blue) of the inverse mean latent period α with $\epsilon = 0.20$ is compared to the prior (yellow). The mean value of the posterior is indicated by the dashed vertical line.

of Liberia is challenging to represent via the standard SEIR structure, even with spatiotemporal infection and contact rates $\beta_i(t)$. While further spatial divisions and other model structures may address this issue, it is likely that the data itself has not captured the true situation on the ground. This is particularly suspect in Liberia, where the super-exponential rise of the data is indicative of reporting errors (The World Health Organization, 2020a). Such errors arise from a variety of factors, including the overwhelmed state of the healthcare system during the epidemic or issues within the governmental agencies that provide the data, as unlike the other two countries, there is no unifying federal system in Liberia that directs the public health response, particularly the collection of data from each of its subregions. Both causes are likely to introduce intrinsic errors in the data. Thus, as the dynamics represent the disease progression in the other three regions so well, the model may actually reveal a more accurate description of the epidemic in Liberia.

4.3.1. Epidemiological parameter estimation

The ABC-SMC algorithm produces approximate posterior distributions for epidemiological parameters in the model, namely α and the coefficients $b^i_{0,1,2}$ that determine the infection rate $\beta_i(t) = \exp\left(b^i_0 + b^i_1 t + b^i_2 t^2\right)$. Fig. 4 shows the approximated posterior distribution for α as compared to its prior. A numerical summary for each of the epidemiological parameters can be found in Supplemental Table S2.

It is particularly useful to examine the numerical values of α as this quantity represents the time that an infected individual spends within the exposed compartment, and the mobility of exposed individuals facilitates the spatial spread of infection. The accepted values of α at $\epsilon = 0.20$ generate a distribution with mean $\bar{\alpha} = 0.0317$, median $\tilde{\alpha} = 0.0288$, and equal-tailed credible interval $[0.0175, 0.0545]$. This indicates a mean latent period of approximately 31 days ranging from a minimum of 11 days to a maximum of 71 days. Though the bulk of the distribution is centered at values that are greater than the expected latency period for Ebola (ranging from 2–21 days), this dynamic is unsurprising due to the simplified nature of the underlying compartmental model and the mode of infection transfer between multiple locations. To accommodate the absence of other infection pathways, such as hospital infections or infections caused by deceased individuals, the latency period becomes inflated to reproduce the data and generate the observed spatial spread. Particularly, as the only mode in this model framework of infection transfer between two spatial locations is that due to the movement of an exposed individual, there is an additional bias towards longer latency periods. Besides the model structure, as noted previously, post-mortem analysis has shown that the standard quarantine of 21 days was insufficient as there were a non-negligible number of individuals who experienced longer latency periods, a dynamic that was newly observed during this particular outbreak of EVD.

Table 3
Flux parameter means, medians, and credible intervals.

Flux(ρ_{ij})	Mean	Median	ETI CI
ρ_{12}	−0.0579	−0.0610	[−0.1578, 0.0529]
ρ_{13}	0.2683	0.2736	[0.0672, 0.4578]
ρ_{23}	0.2686	0.2754	[0.0579, 0.4686]

The indices $i = 1, 2, 3, 4$ refer to Liberia, Sierra Leone, Guinea 1, and Guinea 2, respectively, and the CI is a 89% equal-tailed credible interval. These values are calculated for the distributions generated by the threshold $\epsilon = 0.20$. Recall that every flux is a dimensionless difference between two probabilities.

In order to examine the functional form of the spatiotemporal infection rate for each set of accepted coefficients, the pointwise, temporal quantiles of $\beta_i(t)$ were calculated and plotted in Fig. 5, where the color ranges from the darkest point representing the greatest density of $\beta_i(t)$ to the lightest representing the least density. For reference, inset on the top right panel of Fig. 5 is the same plot generated by a random sample of the coefficients from the priors. It is clear that in all four locations there is a convergence to an estimate of the functional form for $\beta_i(t)$ for all $t > 0$. This is not due to a convergence of the coefficients themselves as Supplemental Figure S11 demonstrates widely varying behavior of each individual distribution.

4.3.2. Mobility parameters

The fluxes ρ_{ij} between two spatial locations and their underlying directional movement probabilities p_{ij}, p_{ji} are critical to the ultimate spread of the disease due to the design of the model, as infection can only spread across a border between two spatial locations via the movement of individuals. While the mobility of susceptible individuals impacts the overall number of potential contacts that could spread infection, the mobility of the exposed class is indistinguishable from their susceptible counterparts and is the primary mode of transmission across space. Thus, the model can uncover the potential pathways of disease spread. A positive flux ρ_{ij} indicates an overall flow towards

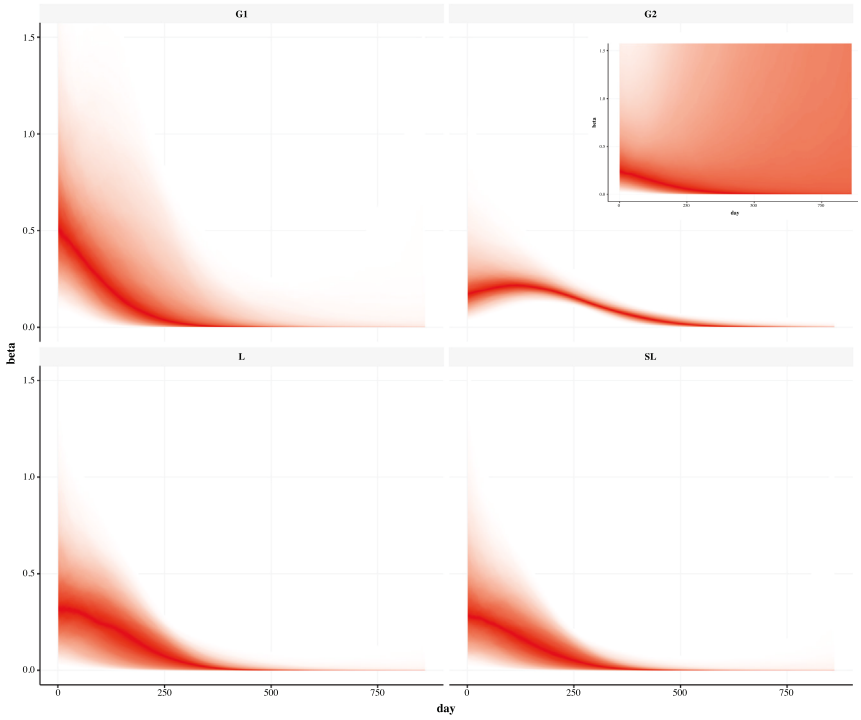

Fig. 5. Functional Distribution of $\beta_i(t)$.
The distribution of the accepted functional values of $\beta_i(t)$ are plotted for $\epsilon = 0.20$ with the greatest density at each time point denoted by the darkest shade of red. Inset on the top right panel, the distribution of the $\beta_i(t)$ prior for all $i = 1, 2, 3, 4$ is plotted as a reference, with the greatest density at each time point denoted by the darkest shade of red.

spatial location i from j. Alternatively, a negative flux favors movement to location j from i. Table 3 catalogues the mean, median, and 89% equal-tailed credible interval for each of the three constant fluxes in this model, while Fig. 6 demonstrates the final posterior for each compared to the prior, as well as joint posterior densities for pairs of fluxes. The joint posterior densities aid in our understanding of the individual fluxes because though the fluxes are independent random variables, they do impact one another within the model. These results signal a few notable dynamics. First, the flux between locations 1 (Liberia) and 2 (Sierra Leone) is near zero, with the weight of the distribution slightly favoring movement towards Sierra Leone. With a near zero flux, it is worth examining the individual probabilities that result from said flux. If the probable mobility between each location is high in both directions, this could be compelling evidence for the merger of the two locations due to the high level of mixing which results. Fig. 7 demonstrates in the panels for Liberia and Sierra Leone that the dominating probability for each location is the probability that an individual, exposed or susceptible, does not change location. The probabilities p_{11} and p_{22} are greater than 0.50 and have means close to 0.75 while the probabilities p_{12} and p_{21} are less than 0.25 and means close to 0.10. Clearly, the system considers these two spatial locations as distinct, with only small levels of movement being necessary and sufficient for the spread of the infection between the two locations.

In contrast to the dynamics between Sierra Leone and Liberia, the mobility of individuals in Guinea 1 is drastically different. Both ρ_{13} and ρ_{23} indicate an overall flow of individuals away from Guinea 1. The mean for both is 0.26 meaning that the probability of leaving Guinea 1 for one of the other locations is on average 0.26 higher than the probability of entering Guinea 1. In fact credible intervals indicate that 89% of all realizations of both fluxes are positive, meaning that a different mobility dynamic favoring entering Guinea 1 is very rare. The epidemic is started on Day 0 in Guinea 1, and thus there must be movement from Guinea 1 to at least one of the other spatial locations to perpetuate the spatial spread. Additionally, as the data suggests, the case counts in Guinea 1 remain relatively low throughout the epidemic, while the bulk of the infection occurs in Sierra Leone and Liberia. There is evidence that the case counts maintain low levels in Guinea 1 for longer periods of time due to rare re-seeding of epidemics from Liberia and Sierra Leone (D'Silva and Eisenberg, 2017). The borders at the nexus of these three regions are extremely porous, and though the region is technically divided between three countries for administrative and political purposes, it was difficult to implement closed borders to slow the spread of the virus from Guinea 1 (Kiskowski, 2014; Cohen, 2016). Despite the clarity of the overall flux away from Guinea 1, the distribution of these positive fluxes have wider spreads, indicating a larger range of acceptable movement dynamics within that limitation. This is especially clear when observing the individual probabilities for Guinea 1 as seen in the final panel of Fig. 7. The three distributions of the probabilities p_{33}, p_{31}, p_{32} have wide ranges of high density regions ranging anywhere from 0 to nearly 0.75 for all three probabilities. Clearly, the three probabilities must sum to one, so all three cannot be simultaneously either high or low, instead they must balance under the constraints of the fluxes. Analyzing various joint probability density functions of combinations of the fluxes can aid in elucidating the shared dynamics. In Fig. 6 the joint densities of pairs of the fluxes are visualized using contour and color density plots. The regions of lightest blue are the regions with the highest joint density. From this it can be deduced that the average behavior of the fluxes overall is that there is a near zero, though slightly negative flux between Liberia and Sierra Leone, the fluxes involving Guinea 1 are likely to be large and equally positive, favoring movement away from Guinea 1. Additionally, as ρ_{12} decreases, ρ_{23} decreases slightly and ρ_{13} increases slightly to compensate. Similarly the relationship between ρ_{23} and ρ_{13} is definitively negative with ρ_{13} increasing to compensate for a reduction in ρ_{23}.

Furthermore, the three-dimensional joint kernel density estimates are computed to identify the three flux values that occur simultaneously with greatest density, detailed in Table 4. Across the board the modal values align with high density regions of the individual distributions, but the interactions between the three different variables may shift the

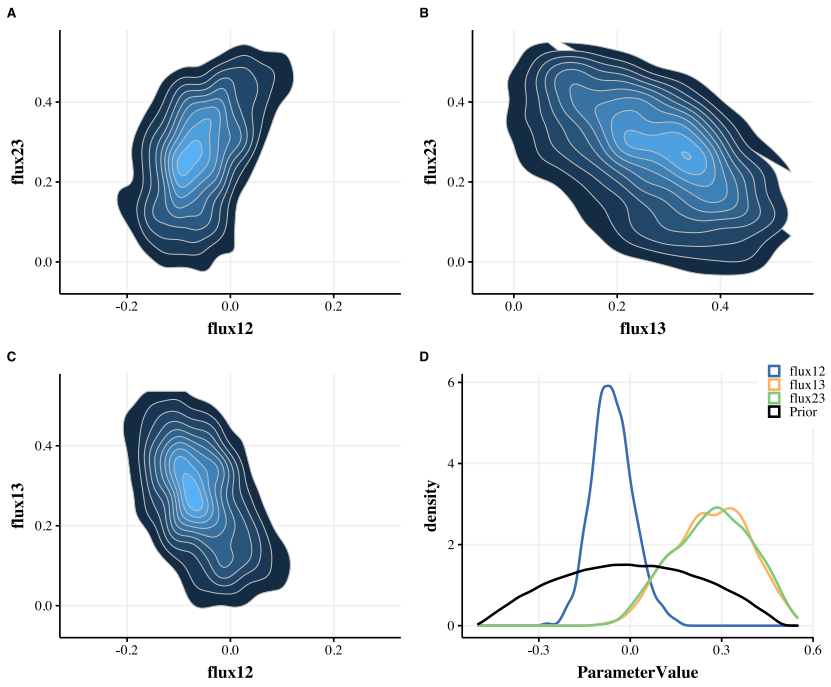

Fig. 6. Marginal and Joint Distributions of Flux Parameters.
Panels (A)–(C) are two dimensional joint density plots of pairs of the three fluxes $\rho_{12}, \rho_{13}, \rho_{23}$ with the lightest blue denoting the highest density region and the gray contours indicated equal density values in the plane. Panel (D) is plots the marginal distributions for each of the fluxes, ρ_{12} (blue), ρ_{13} (yellow), ρ_{23} (green) compared to their prior (black). All distributions are generated with $\epsilon = 0.20$.

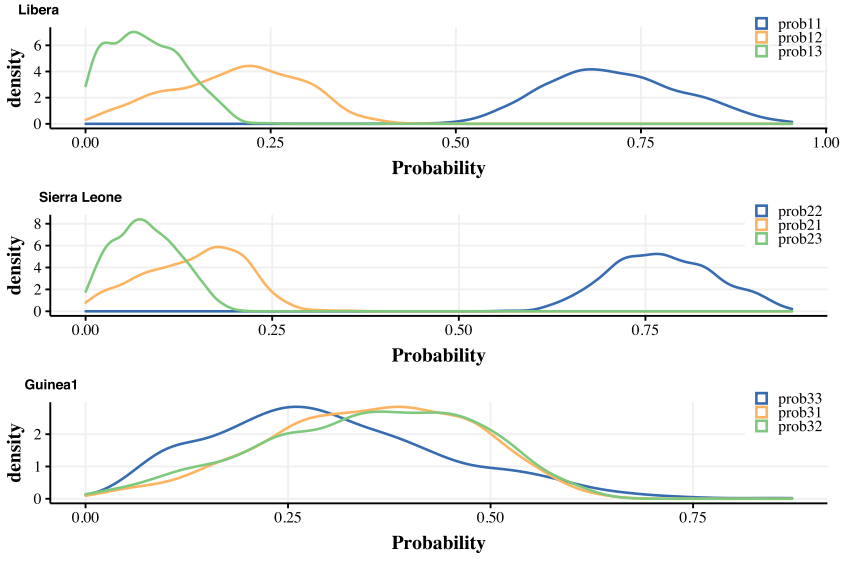

Fig. 7. Distributions of Movement Probabilities.
The distribution of the three probabilities generated by the fluxes are plotted for each spatial location besides Guinea 4. Guinea 4 does not have mobility in this scheme. All distributions are generated with $\epsilon = 0.20$.

values slightly from the distinct means or medians. For example the means and medians of ρ_{13} and ρ_{23} are nearly identical, but the joint modal values indicate slightly more movement from Guinea 1 to Sierra Leone than to Liberia. This dynamic is only revealed by exploring the joint posteriors.

4.3.3. Basic reproduction number

As detailed in the Methods section, there are a variety of tools that modelers use to estimate the epidemiological definition of the basic reproduction number. Here, this important epidemiological value is estimated via an empirically adjusted reproductive number given by

Table 4
Joint modal flux values.

	Fluxes ρ_{ij}		
Parameter	ρ_{12}	ρ_{13}	ρ_{23}
Joint modal value	−0.060	0.244	0.291

The mode of joint distribution of the three fluxes allows for the calculation of the joint modal values of the three fluxes for $\epsilon = 0.20$.

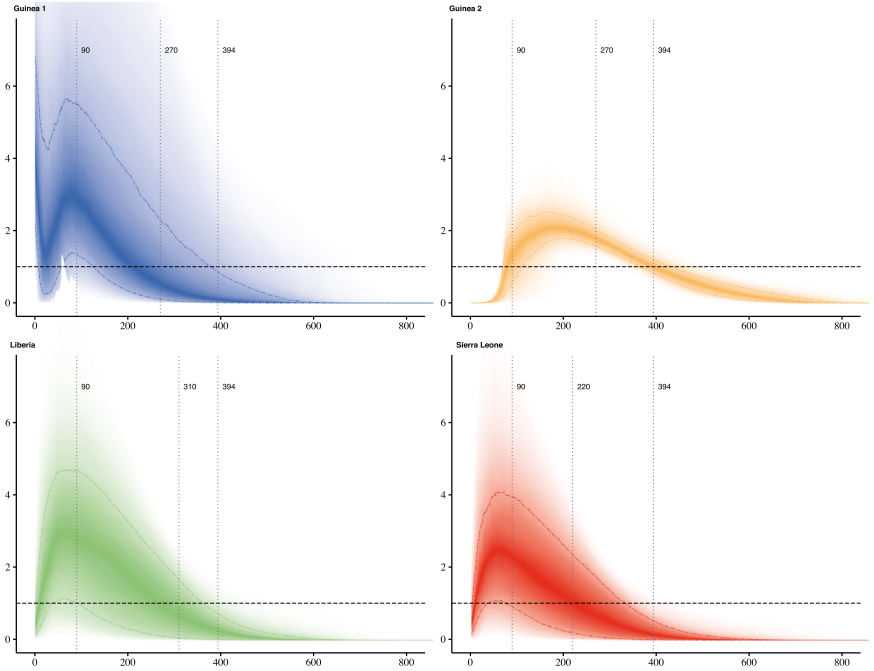

Fig. 8. Functional Distributions of $R_i^{(ea)}(t)$.
The distribution of the accepted functional values with $\epsilon = 0.20$ of $R_i^{(ea)}(t)$ are plotted, with the highest density values at each time point denoted by the darkest color. The dot/dashed lines represent the 50% quantile. The horizontal dashed line on all plots highlights the threshold of one. Additionally, dotted vertical gray lines annotate each spatial location, indicating epidemic days that correspond to interventions or other notable events in each region.

(5) rather than that of the underlying deterministic, continuous-time compartmental model. A comparison of the two estimations of the reproductive number can be found in the Supplemental Information, wherein a spatiotemporally-varying, but unadjusted reproductive number $R_0^i(t)$ clearly does not exemplify the dynamics of the observed or simulated epidemics.

In contrast, the average dynamics of $R_i^{(ea)}(t)$ appear to match the general behavior of both the observed data and the accepted simulated epidemics. Fig. 8 contains more detailed plots with the quantiles of $R_i^{(ea)}(t)$ calculated from the accepted SMC sample represented by the range of color shading, with the darkest region representing the median value of $R_i^{(ea)}(t)$ at each time point. While the accepted sample of parameters and epidemics do produce variability in the empirically adjusted reproduction number, there is a clear trend that the time series for each spatial location follows. The peak and the decay occur in the same places, with only the magnitude presenting the uncertainty, the level of which depends on the spatial location. Clearly $R_4^{(ea)}(t)$, describing Guinea 2, displays low levels of uncertainty at all times, indicating the dynamics of this location are distinctive and must follow certain trends to produce acceptable epidemics. The increasing uncertainty in the other three regions is intuitive due to the connections between the spatial locations and the reality that a secondary infection may not induce subsequent infections within the location in which it was generated.

Finally, though explicit intervention parameters were not included in this model formulation, the calculated reproduction number may uncover links between interventions on the ground and the average number of secondary infections. Additionally, it can be useful in formulating hypotheses concerning the efficacy of different intervention strategies or the accuracy of real time analysis. That being said, it is not possible to fully disentangle the effect of a public health response with a single model parameter. For example, this model employs a time-varying $\beta_i(t)$ to capture factors like intervention and behavior or mobility changes. Alternatively, the mobility changes could have been parameterized by more complex time-varying mobility parameters and fluxes.

The intervention timeline during this epidemic was particularly complex. A more complete description and illustration is given in the Supplemental Information (Figure S7). Fig. 8 is also annotated with vertical dotted lines highlighting the timing of potentially important epidemiological interventions or significant events in the location (West African Ebola virus epidemic timeline, 2020). The timestamp on Day 90 aligns with the peak value of the reproduction number in Guinea 1, Sierra Leone and Liberia, signaling that once the epidemic garnered sufficient attention, in the form of WHO awareness and tracking, the number of secondary cases began to decay in these regions. This does not mean the epidemic ceased to propagate, as the reproduction values remain above one; instead, it signals that the rate of spread was not accelerating. An attempt at closing the borders was made in Sierra Leone on Day 170 and in Liberia on Day 216, a change that the model does not impose as the fluxes are constant. It is unclear how effective this intervention was because while $R_i^{(ea)}(t)$ continues to decrease from its peak, it remained well above one after Day 170/216, and only drops below one on average after Sierra Leone enlisted the help of the military to enforce quarantines on Day 220 and after the implementation of improved mortuary practices in Liberia that were reported to have been effective on Day 310, but occurred prior. Epidemiologists have noted that while travel restrictions around the world aided in preventing significant spread beyond West Africa, the porous borders in the region greatly contributed to the spread of the disease within West Africa (Kiskowski, 2014; Cohen, 2016), supporting the hypothesis that closing the borders was not entirely successful at halting the spatial spread. In Guinea, distrust of international aid was high, which mitigated the impact of interventions to slow the spread. When aid workers and journalists were killed in a prefecture in Guinea 1 on Day 270, the corresponding reproductive number estimate was also high, $R_i^{(ea)}(t) \approx 2$ in both Guinea 1 and Guinea 2. After subsequent efforts by the Guinean government, like cancellation of festivities on Day 283, the reproduction number dropped across all three regions. Finally, on Day 394 all three countries reported their lowest transmission rates since August 2014, which matches the dynamics of the reproduction number this model suggests as this is the first time in which all four

locations have a mean $R_i^{(ea)}(t) \leq 1$ signaling the eventual extinction of the epidemic.

5. Conclusions

The current study introduces a spatially-heterogeneous extension of an embedded stochastic SEIR model that describes the spread of an infectious disease across a graph structure with k different spatial locations and mobility along the edges of the graph by the various disease compartment populations. As demonstrated by the application of the model to the Ebola outbreak in West Africa between 2014 and 2016, the proposed class of models can be adapted to fit a variety of spatial structures and mobility schemes to capture infectious disease spread, as well as rare behaviors such as sparking events. Additionally, the SEIR compartmental structure allows for estimation of important epidemiological quantities, model parameters, and analysis of the basic reproduction number, which can be compared and validated against other methods and models. The design of this class of models further allows for model comparison using Bayes Factors. The use of ABC methods, particularly ABC-SMC, avoids some of the computational challenges observed in this setting with MCMC techniques. The range of complexity available within the class of models can be scaled up or down, both through the mobility and compartmental structure depending on the application, but ABC methods require norms that gauge the relative proximity of proposed simulations, which must be tailored to the proposed model. The summary norms detailed herein provide flexibility and a greater rate of acceptances even though the number of norms increases with the number of spatial locations. The norms themselves are also subject to additional scrutiny, as various adjustments to the existing norm structure may be appropriate. Overall, the flexibility of the model and associated analytic methods present both advantages to exploit and challenges that must be addressed depending on the application.

An ongoing challenge for future applications is to refine the present procedure for determining the spatial mobility structure. The number of spatial locations and the mobility structures in this study are determined *ad hoc*, but in theory the collection of potential spatial mobility structures is arbitrarily large. This study limited the exploration space by specifying small numbers of spatial locations and subsequently imposing assumptions on the mobility that are based on the observed data. However, it is possible to relax these constraints and build a system by which the model space can be explored more holistically. For example, the four location model is a step in the direction of further systematization, as a similar method could be applied to generate a number of mobility schemes for a system with k spatial locations for any desired $k \in \mathbb{N}$. Additional complexity in the application of the operator that introduces "sparking events", by varying the timing, duration, and location of such events, could also be considered when constructing a desired model space to explore. An algorithmic method for determining spatial structures would be particularly valuable in the case of an ongoing crisis, as the course of the epidemic and the dynamics of its future spatial spread would necessarily be unknown, and thus well suited to the adaptable nature of the class of models proposed herein.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The first author was supported by the US National Science Foundation via an NSF Graduate Research Fellowship and under award DMS-1551229. The third author was supported by the US National Science Foundation under awards DMS-1614586 and DMS-1911145.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.sste.2022.100505.

References

Agusto, F.B., Teboh-Ewungkem, M.I., Gumel, A.B., 2015. Mathematical assessment of the effect of traditional beliefs and customs on the transmission dynamics of the 2014 ebola outbreaks. BMC Med. 13 (1), 96. http://dx.doi.org/10.1186/s12916-015-0318-3.

Allen, L.J.S., 2008. An introduction to stochastic epidemic models. In: Brauer, F., van den Driessche, P., Wu, J. (Eds.), Mathematical Epidemiology. In: Lecture Notes in Mathematics, Springer, Berlin, Heidelberg, pp. 81–130. http://dx.doi.org/10.1007/978-3-540-78911-6_3.

Baylis, J., Grimaldi, R.P., Kalmanson, K., 2003. Discrete and Combinatorial Mathematics, an Applied Introduction, Vol. 71. (455), Publication Title, The Mathematical Gazette. URL , Pearson, p. 86, https://www.academia.edu/40379028/Discrete_and_Combinatorial_Mathematics_An_Applied_Introduction_Fifth_Edition_by_Ralph_P_Grimaldi.

Beaumont, M.A., 2010. Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41 (1), 379–406. http://dx.doi.org/10.1146/annurev-ecolsys-102209-144621, URL http://www.annualreviews.org/doi/10.1146/annurev-ecolsys-102209-144621.

Beaumont, M., Cornuet, J.-M., Marin, J.-M., Robert, C., 2009. Adaptive approximate Bayesian computation. Biometrika 96 (4), 983–990, URL http://www.jstor.org/stable/27798882.

Bellman, R., Åström, K.J., 1970. On structural identifiability. Math. Biosci. 7 (3), 329–339. http://dx.doi.org/10.1016/0025-5564(70)90132-X, URL http://www.sciencedirect.com/science/article/pii/002555647090132X.

Brauer, F., 2017. Mathematical epidemiology: Past, present, and future. Infect. Dis. Model. 2 (2), 113–127. http://dx.doi.org/10.1016/j.idm.2017.02.001, URL http://www.sciencedirect.com/science/article/pii/S2468042716300367.

Brauer, F., Castillo-Chávez, C., 2001. Mathematical models in population biology and epidemiology. In: Texts in Applied Mathematics. Springer, URL http://books.google.com/books?id=ahhhwB72JUgC.

Brown, G.D., Oleson, J.J., Porter, A.T., 2016. An empirically adjusted approach to reproductive number estimation for stochastic compartmental models. Biometrics 72 (2), 335–343, URL http://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12432.

Center for Disease Control, 2014. Ebola outbreak in west africa - reported cases graphs (sep. 2015). URL http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/cumulative-cases-graphs.html.

Center for Disease Control, 2015. Ebola (ebola virus disease) transmission (jan. 2015). URL https://www.cdc.gov/vhf/ebola/transmission/index.html.

Cohen, N.J., 2016. Travel and border health measures to prevent the international spread of Ebola. MMWR Suppl. 65, http://dx.doi.org/10.15585/mmwr.su6503a9, URL https://www.cdc.gov/mmwr/volumes/65/su/su6503a9.htm.

Diaz, P., Constantine, P., Kalmbach, K., Jones, E., Pankavich, S., 2018. A modified SEIR model for the spread of Ebola in Western Africa and metrics for resource allocation. Appl. Math. Comput. 324, 141–155. http://dx.doi.org/10.1016/j.amc.2017.11.039, URL https://linkinghub.elsevier.com/retrieve/pii/S0096300317308214.

Dilorenzo, B.D.a.S., 2014a. Ebola virus is surging in places where it was beaten back: experts, library catalog: www.ctvnews.ca section: Health (sep. 2014). URL https://www.ctvnews.ca/health/ebola-virus-is-surging-in-places-where-it-was-beaten-back-experts-1.1996179.

Dilorenzo, S., 2014b. West africa's ebola outbreak is 'totally out of control', library catalog: www.ctvnews.ca section: Health (jun. 2014). URL https://www.ctvnews.ca/health/west-africa-s-ebola-outbreak-is-totally-out-of-control-1.1878133.

D'Silva, J.P., Eisenberg, M.C., 2017. Modeling spatial invasion of Ebola in West Africa. J. Theoret. Biol. 428, 65–75. http://dx.doi.org/10.1016/j.jtbi.2017.05.034, URL http://www.sciencedirect.com/science/article/pii/S0022519317302576.

Eichner, M., Dowell, S.F., Firese, N., 2011. Incubation period of Ebola hemorrhagic virus subtype Zaire. Osong Public Health Res. Perspect. 2 (1), 3–7. http://dx.doi.org/10.1016/j.phrp.2011.04.001, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3766904/.

Fast, S.M., Mekaru, S., Brownstein, J.S., Postlethwaite, T.A., Markuzon, N., 2015. The role of social mobilization in controlling ebola virus in lofa county, liberia. PLOS Curr. Outbreaks http://dx.doi.org/10.1371/currents.outbreaks.c3576278c66b22ab54a25e122fcdbec1, URL https://pubmed.ncbi.nlm.nih.gov/26075140/.

Gilks, W.R., Richardson, S., Spiegelhalter, D., 1995. Markov Chain Monte Carlo in Practice. TRXrMWY_i2IC, CRC Press, google-Books-ID.

Haas, C.N., 2014. On the quarantine period for ebola virus. PLoS Curr. 6, http://dx.doi.org/10.1371/currents.outbreaks.2ab4b76ba7263ff0f084766e43abbd89, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4205154/.

Hunter, E., Mac Namee, B., Kelleher, J.D., 2017. A taxonomy for agent-based models in human infectious disease epidemiology. J. Artif. Soc. Soc. Simul. 20 (3), 2.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Statist. Assoc. 90 (430), 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572, URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572.

Kiskowski, M.A., 2014. A three-scale network model for the early growth dynamics of 2014 West Africa Ebola epidemic. PLoS Curr. Outbreaks http://dx.doi.org/10.1371/currents.outbreaks.c6efe8274dc55274f05cbcb62bbe6070, URL https://pubmed.ncbi.nlm.nih.gov/25685614/.

Kramer, A.M., Pulliam, J.T., Alexander, L.W., Park, A.W., Rohani, P., Drake, J.M., 2016. Spatial spread of the West Africa Ebola epidemic. Royal Soc. Open Sci. 3 (8), 160294. http://dx.doi.org/10.1098/rsos.160294, URL https://royalsocietypublishing.org/doi/full/10.1098/rsos.160294.

Ladner, J.T., Wiley, M.R., Mate, S., Dudas, G., Prieto, K., Lovett, S., Nagle, E.R., Beitzel, B., Gilbert, M.L., Fakoli, L., Diclaro, J.W., Schoepp, R.J., Fair, J., Kuhn, J.H., Hensley, L.E., Park, D.J., Sabeti, P.C., Rambaut, A., Sanchez-Lockhart, M., Bolay, F.K., Kugelman, J.R., Palacios, G., 2015. Evolution and spread of bola virus in Liberia, 2014–2015. Cell. Host Microbe. 18 (6), 659–669. http://dx.doi.org/10.1016/j.chom.2015.11.008, URL https://www.cell.com/cell-host-microbe/abstract/S1931-3128(15)00462-X.

Lahodny, G.E., Allen, L.J.S., 2013. Probability of a disease] outbreak in stochastic multipatch epidemic models. Bull. Math. Biol. 75 (7), 1157–1180. http://dx.doi.org/10.1007/s11538-013-9848-z.

Lekone, P.E., Finkenstädt, B.F., 2006. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. Biometrics 62 (4), 1170–1177. http://dx.doi.org/10.1111/j.1541-0420.2006.00609.x, URL http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2006.00609.x.

Mamo, D.K., Koya, P.R., 2015. Mathematical modeling and simulation study of SEIR disease and data fitting of Ebola epidemic spreading in West Africa. J. Multidiscip. Eng. Sci. Technol. 2 (1), 9.

McCormack, R.K., Allen, L.J.S., 2007. Multi-patch deterministic and stochastic models for wildlife diseases. J. Biol. Dyn. 1 (1), 63–85. http://dx.doi.org/10.1080/17513750601032711.

Merler, S., Ajelli, M., Fumanelli, L., Gomes, M.F.C., y. Piontti, A.P., Rossi, L., Chao, D.L., Longini, I.M., Halloran, M.E., Vespignani, A., 2015. Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. Lancet Infect. Dis. 15 (2), 204–211. http://dx.doi.org/10.1016/S1473-3099(14)71074-6, URL https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(14)71074-6/abstract.

Modica, G., Poggiolini, L., 2012. Discrete time Markov chains. In: A First Course in Probability and Markov Chains. John Wiley & Sons, Ltd, pp. 168–240. http://dx.doi.org/10.1002/9781118477793.ch5, section 5, URL http://onlinelibrary.wiley.com/doi/abs/10.1002/9781118477793.ch5.

Pell, B., Baez, J., Phan, T., Gao, D., Chowell, G., Kuang, Y., 2016a. Patch models of EVD transmission dynamics. In: Chowell, G., Hyman, J.M. (Eds.), Mathematical and Statistical Modeling for Emerging and Re-Emerging Infectious Diseases. Springer International Publishing, Cham, pp. 147–167. http://dx.doi.org/10.1007/978-{3}-319-40413-4_10.

Pell, B., Baez, J., Phan, T., Gao, D., Chowell, G., Kuang, Y., 2016b. Patch models of EVD transmission dynamics. In: Chowell, G., Hyman, J.M. (Eds.), Mathematical and Statistical Modeling for Emerging and Re-Emerging Infectious Diseases. Springer International Publishing, Cham, pp. 147–167. http://dx.doi.org/10.1007/978-3-319-40413-4_10, URL http://link.springer.com/10.1007/978-3-319-40413-4_10.

Perez, L., Dragicevic, S., 2009. An agent-based approach for modeling dynamics of contagious disease spread. Int. J. Health Geogr. 8 (1), 50. http://dx.doi.org/10.1186/1476-072X-8-50.

Porter, A.T., Oleson, J.J., 2014. A multivariate CAR model for mismatched lattices. Spatial Spatio-Temporal Epidemiol. 11, 79–88. http://dx.doi.org/10.1016/j.sste.2014.08.001, URL http://www.sciencedirect.com/science/article/pii/S1877584514000604.

Rivers, C.M., Lofgren, E.T., Marathe, M., Eubank, S., Lewis, B.L., 2014. Modeling the impact of interventions on an epidemic of ebola in sierra leone and liberia. PLoS Curr. 6, http://dx.doi.org/10.1371/currents.outbreaks.fd38dd85078565450b0be3fcd78f5ccf, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4399521/.

Santermans, E., Robesyn, E., Ganyani, T., Sudre, B., Faes, C., Quinten, C., Van Bortel, W., Haber, T., Kovac, T., Van Reeth, F., Testa, M., Hens, N., Plachouras, D., 2016. Spatiotemporal evolution of Ebola virus disease at sub-national level during the 2014 West Africa epidemic: Model scrutiny and data meagreness. PLoS One 11 (1), http://dx.doi.org/10.1371/journal.pone.0147172, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4714854/.

The World Health Organization, 2016a. Ebola virus disease fact sheet (jan. 2016). URL http://www.who.int/mediacentre/factsheets/fs103/en/.

The World Health Organization, 2016b. Frequently asked questions on ebola virus disease (jan. 2016). URL http://www.who.int/csr/disease/ebola/faq-ebola/en/.

The World Health Organization, 2020a. Liberia: A Country – and Its Capital – are Overwhelmed with Ebola Cases. World Health Organization, URL http://www.who.int/entity/csr/disease/ebola/one-year-report/liberia/en/. (Accessed August 2020).

The World Health Organization, 2020b. Ebola data and statistics, library catalog: apps.who.int. URL https://apps.who.int/gho/data/node.ebola-sitrep.quick-downloads?lang=en. (Accessed August 2020).

Valdez, L.D., Aragão Rêgo, H.H., Stanley, H.E., Braunstein, L.A., 2015. Predicting the extinction of ebola spreading in liberia due to mitigation strategies. Sci. Rep. 5, http://dx.doi.org/10.1038/srep12172, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4507172/.

Van den Driessche, P., Watmough, J., 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. 180 (1), 29–48.

Velásquez, G.E., Aibana, O., Ling, E.J., Diakite, I., Mooring, E.Q., Murray, M.B., 2015. Time from infection to disease and infectiousness for Ebola virus disease, a systematic review. Clin. Infect. Dis. 61 (7), 1135–1140. http://dx.doi.org/10.1093/cid/civ531, URL https://academic.oup.com/cid/article/61/7/1135/290096.

West African Ebola virus epidemic timeline, 2020. Page version ID: 965450034 (jul. 2020). URL https://en.wikipedia.org/w/index.php?title=West_African_Ebola_virus_epidemic_timeline&oldid=965450034.

WHO Ebola Response Team, 2014. Ebola virus disease in west africa the first 9 months of the epidemic and forward projections. N. Engl. J. Med. 2014 (371), 1481–1495.

Wong, Z.S.Y., Bui, C.M., Chughtai, A.A., Macintyre, C.R., 2017. A systematic review of early modelling studies of Ebola virus disease in West Africa. Epidemiol. Infect. 145 (6), 1069–1094. http://dx.doi.org/10.1017/S0950268817000164, URL https://www.cambridge.org/core/journals/epidemiology-and-infection/article/systematic-review-of-early-modelling-studies-of-ebola-virus-disease-in-west-africa/154353B9A815326FE3656046AD6390B6.