

# Spatiotemporal Epidemic Modeling with libSpatialSEIR Specification, Fitting, Selection, and Prediction

G. D. Brown, J. J. Oleson

University of Iowa, 145 N. Riverside Drive, Iowa City, IA, USA  
Telephone: (319) 384-1595  
Email: grant-brown@uiowa.edu  
Email: jacob-oleson@uiowa.edu

## 1. Introduction

The ability to effectively model the spread of infectious diseases over space and time is an important tool in an increasingly connected world. Epidemic modeling allows analysts to estimate the size of ongoing outbreaks, quantify pathogen reproductive behavior, evaluate public health interventions, and predict the degree of future spread. Despite these attractive features, epidemic models can be difficult to specify and computationally impractical to fit.

We propose a general class of epidemic models which allows the straightforward specification of a wide range of spatial structures, and have developed the open source libSpatialSEIR epidemic modeling software to implement it. Set in the stochastic compartmental modeling framework, these techniques track the transition of populations through the Susceptible, Exposed, Infectious, and Removed disease states, and can address a wide array of spatial and non-spatial hypotheses (Porter and Oleson 2013; Lekone and Finkenstädt 2006; Chowell et. al. 2004). Our software aims to expand both the class of models which may be feasibly fit on modest hardware, in addition to the number of researchers with the ability to use them. It includes both a simplified, high level, API for standard analyses and a set of tools for specifying the model components individually to allow for maximum flexibility.

In this work, we introduce the conceptual framework behind these models, discuss several important computational tools, and finally illustrate the model fitting, selection, and prediction process in the context of the ongoing Ebola epidemic in West Africa.

## 2. Stochastic Compartmental Models

Compartmental techniques have a long history in the epidemic modeling literature, and were originally introduced in the context of deterministic systems of differential equations (Kermack and McKendrick, 1927; Hethcote, 2000). Stochastic formulations of these techniques introduce probabilistic transitions between disease states, allowing for full estimation of the uncertainty of important model parameters. Our work concerns the stochastic spatial SEIR model class, which incorporates measurements of disease processes which occur over discrete time and space. This often corresponds directly to the spatiotemporal scale on which epidemic data is available: periodic counts of disease incidence in administrative regions. The temporal process employed by these models is given in Equations 1 and 2, for time points  $\{t_i : i = 1, \dots, T\}$  and spatial locations  $\{s_j : j = 1, \dots, n\}$ .

$$\begin{aligned}
S_{ij} &= S_{i-1,j} + S_{ij}^* - E_{ij}^* & S_{ij}^* &\sim \text{binom}(R_{ij}, \pi_j^{(RS)}) \\
E_{ij} &= E_{i-1,j} + E_{ij}^* - I_{ij}^* & E_{ij}^* &\sim \text{binom}(S_{ij}, \pi_{ij}^{(SE)}) \quad (2) \\
I_{ij} &= I_{i-1,j} + I_{ij}^* - R_{ij}^* & I_{ij}^* &\sim \text{binom}(E_{ij}, \pi^{(EI)}) \\
R_{ij} &= R_{i-1,j} + R_{ij}^* - S_{ij}^* & R_{ij}^* &\sim \text{binom}(I_{ij}, \pi^{(IR)})
\end{aligned} \quad (1),$$

We introduce spatial and temporal heterogeneity into the prior structure for the exposure probability,  $\pi^{(SE)}$ , by including a linear predictor term to capture shared and location specific explanatory variables, and a set of distance matrices,  $\{D_z : z = 1, \dots, Z\}$ , to model contact between spatial locations. Each distance matrix is associated with a single spatial parameter,  $\rho_z$ , which determines the contribution of the corresponding distance metric to the epidemic mixing process. With the introduction of several minor distributional assumptions, this exposure probability can be shown to have the parametric form given in Equation 3 (Brown et. al. 2015). Here,  $\theta_{ij}$  denotes the value of the linear predictor corresponding to the time point  $t_i$  and spatial location  $s_j$ . An example of practical use of this linear predictor term is given in Section 4.

$$\begin{aligned}
\pi_{ij}^{(SE)} &= 1 - \exp \left( \left\{ -\eta_{i.} - \sum_{z=1}^Z \rho_z (\mathbf{D}_z \eta_{i.}) \right\}_j \right) \\
\eta_{i.} &= \left( \frac{I_{i1}}{N_1} e^{\theta_{i1}}, \dots, \frac{I_{in}}{N_n} e^{\theta_{in}} \right) \quad (3)
\end{aligned}$$

A simple exponential form is employed for the E to I and I to R transitions, and is illustrated in Equation 4. In this parameterization,  $1/\gamma_{(EI)}$  and  $1/\gamma_{(IR)}$  correspond to the average latent and infectious periods in the chosen temporal unit, respectively.

$$\begin{aligned}
\pi^{(EI)} &= 1 - \exp(-\gamma_{(EI)}) \\
\pi^{(IR)} &= 1 - \exp(-\gamma_{(IR)}) \quad (4)
\end{aligned}$$

### 3. Software

The libSpatialSEIR modeling software, freely available online (Brown, 2014), comprises a C++ library and R-package interface, and aims to provide an easy to use and flexible implementation of the spatial SEIR and SEIRS model classes. The software was designed to allow researchers to quickly begin modeling epidemics, while allowing enough flexibility to explore numerous hypotheses. An example of how simple the R code required to perform such an analysis can be is given in Example 1. The code presented here reads in incidence data for the 1995 Ebola outbreak in the Democratic Republic of the Congo, and fits a model to evaluate the effectiveness of the intervention efforts. A complete analysis of these data using libSpatialSEIR may be found in our manuscript on the Empirically Adjusted Reproductive Number (Brown et. al. 2015).

### Example 1: High Level API

```
library(spatialSEIR)
cases = read.csv("http://bit.ly/1w3m3Xr")

results = fit.qSEIR(Count ~ daysSinceIntervention,
  p_ei=1-exp(-1/5),
  p_ir=1-exp(-1/7),
  data = cases,
  N = 5.36e6,
  transition_ess=1000,
  seed=12345,
  n.cores=3,
  return.cluster=TRUE)
```

## 4. Analysis

In order to explore the range of decisions which inform epidemic data analysis in the spatial SEIR setting, we demonstrate a complete analysis of the 2014-2015 Ebola epidemic in West Africa. Data were originally collected from World Health Organization situation reports, and were pre-processed and smoothed for clarity. In Example 2, we begin by loading the requisite R packages and reading in the processed data. As we have found no complete and detailed account of all public health intervention activities, we employ a set of basis splines to capture changes in population behavior over the course of the epidemic. Using the model selection techniques discussed in Brown et. al. (2015), we selected a three degree of freedom basis, created with the *ns* function from the *splines* package. This temporal basis is then combined with a separate intercept for each nation to form the intensity process design matrix, *Z*. Finally, we define a distance matrix for each national border. This spatial structure is defined by a set of indicator functions, equal to one when the spatial locations with corresponding row and column indices share a border and zero otherwise. While such a spatial structure is discrete, one could instead include informative weights motivated by, for example, a gravity model. Our development may be seen explicitly in Example 2.

Specification of hierarchical models is always complex. In *libSpatialSEIR*, we define an array of model objects which correspond to each level of the model. This includes a data model to relate the observations to the underlying epidemic parameters, process models to capture exposure, any reinfection activity, latent and infectious durations, and spatial heterogeneity, starting values for unknown population counts, and finally instructions for initial configuration of the MCMC samplers. These components are created in Example 3, and combined into a functional model by the *buildSEIRModel* function.

## Example 2: Ebola Model Preparation

```
library(spatialSEIR)
library(splines)

# Read in Data
processedData = read.csv("http://bit.ly/1DYcLQx")

# Build temporal basis
basis = ns(cumsum(processedData$offset), df = 3)
Z = cbind(diag(3)[rep(1:3, each = nrow(processedData)),],
          basis[rep(1:nrow(basis), 3),])

# Declare neighborhood matrices.
DM1 = matrix(c(0,1,0,
              1,0,0,
              0,0,0), nrow=3, byrow=TRUE)
DM2 = matrix(c(0,0,1,
              0,0,0,
              1,0,0), nrow=3, byrow=TRUE)
DM3 = matrix(c(0,0,0,
              0,0,1,
              0,1,0), nrow=3, byrow=TRUE)

# Define the data set
cases = cbind(processedData$Guinea, processedData$Liberia,
              processedData$SierraLeone)

# Declare the population matrix and initial infectious counts
N = matrix(c(1.005e7, 4.1285e6, 6.1902e6), nrow = nrow(cases),
           ncol = 3, byrow=TRUE)
I0 = c(86,0,0)
```

## Example 3: Low Level API

```
DataModel = buildDataModel(cases, type = "overdispersion", phi = 1)
ExposureModel = buildExposureModel(Z, nTpt=nrow(processedData),
                                   nLoc=3, offset=processedData$offset)
ReinfectionModel = buildReinfectionModel("SEIR")
SamplingControl = buildSamplingControl(iterationStride=500)
InitContainer = buildInitialValueContainer(data=cases, N=N,
                                           S0=N[1,]-3*I0,
                                           E0 = I0,
                                           I0 = I0)
DistanceModel = buildDistanceModel(list(DM1, DM2, DM3),
                                   priorAlpha = 1,
                                   priorBeta = 10)
TransitionPriors = buildTransitionPriorsFromProbabilities(
  p_ei = 1-exp(-1/5),
  p_ir = 1-exp(-1/7),
  p_ei_ess = 1000,
  p_ir_ess = 1000)
SEIRModel = buildSEIRModel("samples.csv", DataModel,
                           ExposureModel, ReinfectionModel,
                           DistanceModel, TransitionPriors,
                           InitContainer, SamplingControl)
```

While the high level *qSEIR* and *qSpatialSEIR* functions automatically run several MCMC chains until convergence, the general model specification tools used here require the user to configure and execute samples. This process is illustrated in Example 4, which also produces the example infectious count summary shown in Figure 1.

#### Example 4: Sampling and Inference

```
# Keep track of compartment counts, not just basis parameters
sapply(0:2, function(i) {SEIRModel$setTrace(i)})

# Configure samplers; these are reasonable values for many analyses.
SEIRModel$compartmentSamplingMode=17
SEIRModel$useDecorrelation=10
SEIRModel$performHybridStep=11

# Run the model
SEIRModel$simulate(100000)

# Read in the MCMC samples
mcmc.samples = read.csv("samples.csv")

# Clean up C++ objects
rm(SEIRModel, DataModel, ExposureModel, ReinfectionModel,
    DistanceModel, TransitionPriors, InitContainer,
    SamplingControl)

# Example summary measure, shown in figure 1
hist(mcmc.samples$I_2_46[100:nrow(mcmc.samples)],
     main = paste("Estimated Remaining Infectious Individuals",
                  "\n Sierra Leone - 3/29/2015"),
     freq=FALSE, breaks = 20, xlab = "Currently Infectious")
```

Example R code to generate predictions using MCMC samples is available in the supplemental companion package to Brown et. al. (2015). Predictions made using the latest available data in January 2015 showed a decrease in infection size in all three nations with an epidemic extinction time in late April. Updated models at the end of March validate the observed decrease in Sierra Leone and Guinea in particular, but indicate that different dynamics have taken over as the epidemic has shrunk (Figure 2). Such behavior is likely due to unmodeled spatial heterogeneity; the epidemic spreads within and between villages and counties, so nationally aggregated data provides a relatively coarse view of the underlying disease dynamics

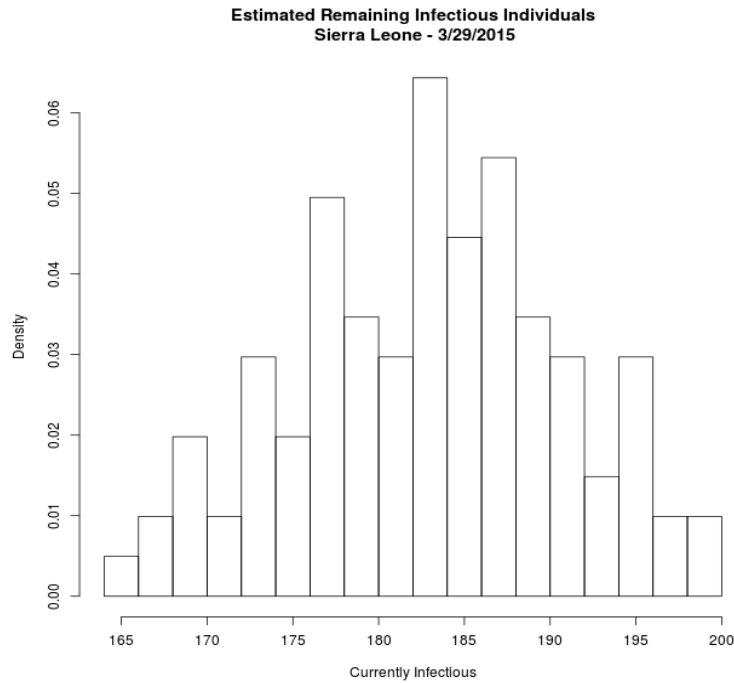


Figure 1: Example Output

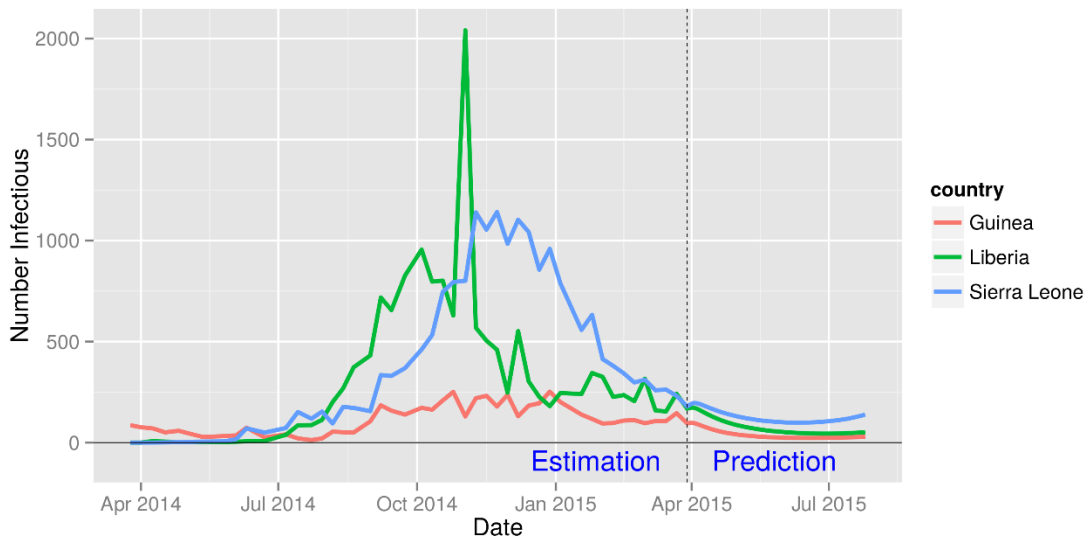


Figure 2: Current Predictions

## 5. Impact

These powerful analytical methods, while intuitive, have long lacked high level computational tools. The ability to quickly assess the behavior of emerging pathogens, characterize the effectiveness of interventions, and evaluate the drivers of geographic spread is of great importance in the management of epidemics. Our software addresses these analytical needs in three ways. First, the development of empirically driven

reproductive number estimates provides a tool to estimate changes in epidemic spread over time (Brown et. al. 2015). Second, the general parameterization of the exposure process allows the inclusion of intervention effects, climactic and weather related data, and numerous other quantities which vary over both space and time. Finally, the flexible and intuitive spatial structure of these models allows the incorporation of diverse geospatial data, and leverages existing spatial computation and data expertise. We believe that our software has the potential to make such analyses feasible for a wider community of researchers. Moreover, the development emphasis on computational efficiency allows the use of larger data sets than could be practically analyzed in this setting in the past, and work is ongoing to better utilize the heterogeneous computing architectures increasingly available even on consumer devices.

## 6. References

- Brown, G. D. 2014, spatialSEIR: a framework for fitting Bayesian spatial SEIR epidemic models. R package, available at <https://github.com/grantbrown/libSpatialSEIR>
- Brown, G. D., Oleson, J. J., and Porter, A. T. 2015, An empirically adjusted approach to reproductive number estimation for stochastic compartmental models: A case study of two Ebola outbreaks. *Submitted Manuscript*. <https://github.com/grantbrown/EARNMC>
- Chowell, G. et. al. 2004, The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda. *Journal of Theoretical Biology*
- Hethcote, H. W. 2000, The mathematics of infectious diseases. *SIAM Review*
- Kermack, W. and McKendrick, A. 1927, A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society, London*
- Lekone, P. E. and Finkenstadt, B. F. 2006, Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*
- Porter, A. T. and Oleson, J. J. 2013, A path-specific SEIR model for sue with general latent and infectious time distributions. *Biometrics*