2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Temporal spatial modelling of Ebola

Supervisor :
Prof. dr. Marc AERTS
Prof.dr. Christel FAES

Rachid Muleia

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Temporal spatial modelling of Ebola

Supervisor :
Prof. dr. Marc AERTS
Prof.dr. Christel FAES

Rachid Muleia
Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

Acknowledgments

First and foremost, I would to like to thank God, the Almighty, for granting me the opportunity to pursue my masters studies. I also thank him for giving me strength and courage to do this work.

I am very grateful to Desafio Program and the Flemish Inter-university (VLIR UOS) for the scholarship, your financial support throughout my masters has made it possible for me to completely focus on my studies. I will always feel fortunate and honored to have had this opportunity.

I would like to express my sincere gratitude to Prof. dr. Christel Faes and Prof. dr. Marc Aerts who gave me the topic and all the necessary material needed throughout the whole thesis. A word of gratitude also goes to all the Professors who taught me throughout my masters studies, without them I would not have made this far. Last but not the least, I would like to thank my family and classmates.

Contents

List of Figures

List of Tables

1. Introduction

The current Ebola outbreak in West Africa that started in May 2014, is the largest since Ebola was first identified in Sudan in 1976. Currently the disease has mostly affected Liberia, Guinea and Sierra Leone, and as of May 2014, it has been reported that 27,443 people contracted the disease and about 11,207 died (World Health Organization; 2015b). On 9 May 2015, Liberia was declared Ebola free after 42 days without any new reported case, however, on June 29 a new case was detected (Center for Disease Control; 2015a). In Guinea the epidemic is still ongoing and 3 new confirmed cases were reported in the week to 16 August, while in Sierra Leone no new case was reported (World Health Organization; 2015a). Among all countries, Sierra Leone is the worst affected country, with 13,534 confirmed, probable, and suspected cases, as of 23^{rd} May 2014 (Sierra Leone Ministry of Health and Sanitation; 2015).

To better understand the transmission dynamics of infectious diseases, mathematical modeling is indispensable (Siettos and Russo; 2013). Over the years a wide range of mathematical models has been developed. In the field of modeling infectious diseases, deterministic SEIR (Susceptible-Exposed-Infectious-removed) has been extensively used. And currently, Althaus (2014) has used a deterministic SEIR model that can be described by a set of ordinary differential equations (ODE) to estimate the basic and the effective reproductive number for the current Ebola outbreak. However, the deterministic SEIR model described by a set of ODE has some limitation. In first place, it assumes that the observed process is in a continuous time, while in practice most of the times, the data come in a time-series. Secondly it assumes that all states (susceptible, exposed, infectious and removed) are observable while in reality only reported cases and deaths are observed (Yin; 2008; Carlin and Louis; 2008). Furthermore, King et al. (2015) and Erickson (2015) recommend that, whenever possible, deterministic models should be avoided, because such models tend to overestimate the predictions on the course of the epidemic and substantially underestimate the uncertainty around the estimates and the predictions. Therefore, stochastic models, which have room for randomness and which can precisely address uncertainty, should be used instead (King et al.; 2015).

In this report we use a stochastic discrete-time SEIR model proposed by Lekone and Finkenstädt (2006) to estimate the reproductive number and the incubation period in the districts of Sierra Leone. This model is appropriate to model the current Ebola outbreak since it accounts for the fact that, the reported cases are time series, that is they are observed in a discrete-time, and also has room for randomness by allowing the movements of individuals from one state to another to be governed by probability density. Because the model has many unobservables states, Bayesian approach using Markov Chain Monte Carlo methods is used, since it allows to numerically integrating over the latent variables (Lekone and Finkenstädt; 2006). This report is not limited to only estimate the reproductive number and incubation period, but also intends to study the spatial structure of the transmission rate. Therefore, for this purpose Conditional

Autoregressive (CAR) models are also considered.

The objective of this report was to apply a stochastic discrete-time SEIR model to study the dynamic transmission of the current Ebola outbreak in the all affected districts of Seirra Leone. In addition to that see how the transmission rate varies across the country.

The reminder of this report is organized as follows: Section 2.1 provides a description of the data. In section 2.2, a discussion on compartmental epidemiological model and deterministic SEIR model is given. Section 2.3 introduces the stochastic discrete-time SEIR model together with the estimation procedure, and section 2.4 presents the Conditional autoregressive model for spatial analysis. Next , results of the study are presented in section 3. Discussion and main conclusions are described in Section 4. Finally, in section 5 we close by describing the limitations of the reports and giving avenues for further research.

2. Methodology

2.1. Data Description

Ebola, also known as Ebola hemorrhagic fever, is a rare and deadly disease caused by an infection of one of the Ebola virus strains. Ebola can cause diseases in humans and nonhuman primates (monkeys, gorillas, and chimpanzees) (World Health Organization; 2015c). It is mainly transmitted through direct contact with infected bodily fluids and contaminated materials. At the early stage, the disease is characterized by initial-flu symptoms , high fever, severe headache followed by pharyngitis and abdominal pain, whereas the late stage the disease is marked by vomiting, diarrhea, rash, and internal and external bleeding (Yamin et al.; 2015). After being exposed to the Ebola virus disease (EVD) the symptoms can show up anywhere from 2 to 21 days, but on average the incubation period is around 8 to 10 days (Center for Disease Control; 2015b). Breman et al. (1978) reported that after contracting the infection, the person stays infectious for a mean period between 3.5 and 10.7 days. Currently the disease has been affecting Liberia, Sierra Leone and Guinea, with a case fatality rate of 30 to 90%, depending on the virus species (Baize et al.; 2014). There are four identified strains of Ebola virus, three of which are the Zaire, Ivory Coast and Sudan strains, and they have shown to cause disease in both humans and nonhuman primates, with the Zaire strain exhibiting the highest fatality rate (Feldmann et al.; 1994; Sanchez et al.; 1996).

In this report we study the data of the current Ebola outbreak in 14 districts of Sierra Leone, namely Bo, Bombali, Bonthe, Kailahun, Kambia, Koinadagu, Kenema, Kono, Moyamba, Port Loko, Pejehu, Tonkolili, Western Area Rural, and Western Area Urban. The Ebola outbreak in Sierra Leone had its first case on the 24^{th} of May 2014, in the hospital of Kenema. Nevertheless, in this report we analyze the cumulative cases counts reported by the Ministry of Health of Sierra Leone (Sanitation Moha; 2015) from July 2014 to June 2015. From the mentioned period, in the entire country 13,088 cases of Ebola were reported, with Western Area Urban district being the most affected with 3,459 reported cases, and Bonthe being the least affected with 89 reported cases. Figure 2 (b) well depicts the distribution of Ebola cases in all affected districts of Sierra Leone.

2.2. Data Transformation

The incidence counts were reported in irregular time interval, so to avoid problems associated with considering time steps that are too large, daily incidence curves of EVD were estimated. Additionally, when using difference equations model to model a disease, which is the case in this report, Vynnycky and White (2010) suggest using a time step that is less than the shortest average duration that individuals spend in a given compartment. To estimate the daily incidence smoothing splines were fitted to the cumulative cases of EVD in each district. Then, the daily difference of the cumulative cases were taken to obtain the daily incidence time series. The

smoothing splines were fitted in R Core Team (2015) using the function `smooth.spline`. For more details see Nishiura and Chowell (2014), and Chandler and Scott (2011). Note that in the subsequent analyses, the models used did not account for the uncertainty introduced by the fitted smoothing splines.

2.3. **Compartmental Epidemiological Model**

The use of compartmental epidemic models dates back to 1927 in a paper by Kermack and McKendrick (1927). They developed a nonlinear system of ordinary differential equation that opened doors for the use of many mathematical modes in modeling infectious disease, such as, SIR (Susceptible, Infectious, Recovered) and SEIR (Susceptible, Exposed, Infectious, Recovered) models. The modeling of infectious diseases can either be done by deterministic or stochastic model. The deterministic models in general describe on average what happens in the population, while stochastic models, apart from describing the average course of the disease, they also allow the number of individuals who move from one compartment to another to vary through chance. Therefore, providing a range in which an outcome is more likely to occur and making them more appealing for decision making (Vynnycky and White; 2010). The well known stochastic model in the context of compartmental epidemic models is the chain binomial model introduced by Greenwood (1931). This model is an extension of the Susceptible-Infective-Recovered (SIR) model, where individual transition probabilities from susceptible to diseased states are given in terms of the binomial distribution (Tuckwell and Williams; 2007).

The Susceptible-Infective-Recovered (SIR) model is one of the basic compartmental model in infectious disease epidemiology and is broadly used and well suited to model many infectious diseases (Hens et al.; 2012). It assumes that hosts within a population are categorized as susceptible (if previously unexposed to the pathogen), infective (if currently colonized by the pathogen), and recovered (if they have successfully cleared the infection) (Keeling and Rohani; 2008). However, in many infectious disease, infected individuals are exposed before becoming infectious. Therefore, the hosts in the population can no longer be categorized as susceptible, infectious or recovered. This leads the extension of SIR model to SEIR model, in which E represents the exposed compartment In practice, modeling of infectious diseases is usually done by means of a deterministic SEIR model, using a set of ordinary differential equation (Althaus; 2014; Lekone and Finkenstädt; 2006). A graphical representation of SEIR model can be seen in Figure 1 below

The dynamics of the SEIR model in a continuous time can be described using the following set of ordinary differential equations (ODEs):

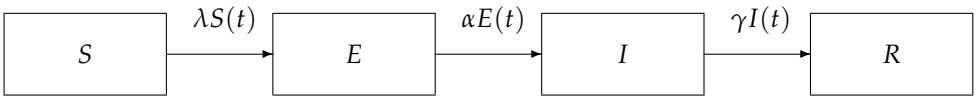


Figure 1: Graphical illustration of SEIR model showing the flows among disease state

$$\frac{dS(t)}{dt} = -\lambda S(t) \tag{1}$$
$$\frac{dE(t)}{dt} = \lambda S(t) - \alpha E(t) \tag{2}$$
$$\frac{dI(t)}{dt} = \alpha E(t) - \gamma I(t) \tag{3}$$
$$\frac{dR(t)}{dt} = \gamma I(t) \tag{4}$$

A distinct feature of Ebola virus disease is that individuals exposed to the virus who become infectious do so after a mean incubation period (Bashar et al.; n.d.). Hence, to reflect this feature a SEIR model is deemed appropriate and will be considered throughout this report. In this model, Susceptible (S) individual in contact with the virus enter the exposed (E) state at a rate of $\beta I/N = \lambda$ (Known as force of infection), where β is the transmission rate per person per day. The exposed experiences an average incubation period of $1/\alpha$ days before passing to the infectious (I) state. The exposed state is assumed to be asymptomatic as well as noninfectious. Infective (I) individuals move to the R state, either recovered or dead, at a rate of γ.

Moreover, in the SEIR model, most often only some of the transition among compartment are observed. This makes the development and the implementation of valid estimation methods such as likelihood-based approaches difficult (Carlin and Louis; 2008). Therefore, proper Bayesian analysis implemented by Markov Chain Monte Carlo (MCMC) can be used to handle the missing information by integrating over the distribution of the unobserved variables (Lekone and Finkenstädt; 2006). In addition to that, data are often available at discrete time points while the true underlying process is continuous in time, but the deterministic SEIR model, which is described by set of ordinary differential equation, ignores that. In this report we use a model proposed Lekone and Finkenstädt (2006), which is a stochastic discrete-time model, to study the transmission dynamics of the EVD in the 14 districts of Sierra Leone, in which a Bayesian analysis is performed.

2.4. Stochastic compartmental model

2.4.1 Model Specification

The analysis presented in this report is based on a stochastic discrete-time susceptible-exposed-infectious-recovered (SEIR) model wherein at time t, each individual is in one of the four states, namely susceptible ($S(t)$), exposed ($E(t)$), Infectious ($I(t)$) and recovered ($R(t)$), and without loss of generality, t will represent time in days. The discrete-time stochastic SEIR model can be specified by the following set of difference of equations.

$$S(t+h) = S(t) - B(t) \tag{5}$$
$$E(t+h) = E(t) + B(t) - C(t) \tag{6}$$
$$I(t+h) = I(t) + C(t) - D(t) \tag{7}$$
$$S(t) + E(t) + I(t) + R(t) = N \tag{8}$$

where h represents the interval between the time points at which the measurements are taken, here $h = 1$ day. $B(t)$ represents number of susceptible individuals who become infected, $C(t)$ the number of cases by date of symptom onset, and $D(t)$ the number of cases who are removed (die or recover) from the infectious class during the time interval $(t, t+h]$.

In the above stochastic discrete-time model, transition from one state to another, in a day, follows a probability distribution where $B(t)$, $C(t)$ and $D(t)$ are random variables with binomial distribution ($Bin(n, p)$) , that is

$$B(t) \sim Bin(S(t), P(t)), \quad C(t) \sim Bin(E(t), p_C), \quad \text{and} \quad D(t) \sim Bin(I(t), p_R) \tag{9}$$

with

$$P(t) = 1 - \exp\Big[-\frac{\beta(t)}{N} I(t) h\Big], \quad p_C = 1 - \exp(-\alpha h), \quad \text{and} \quad p_R = 1 - \exp(-\gamma h) \tag{10}$$

In the equation (9), usually the $S(t) \approx N$ (the population size) and as the epidemic progresses $E(t)$ and $I(t)$ may get larger. Therefore, knowing that the Poisson distribution is the limiting case of a Binomial distribution when the number of trials n gets very large and the probability p gets small, the transition among states can me modeled through Poisson distribution instead (Yu et al.; 2010). So the equation (9) can be rewritten as follow:

$$B(t) \sim Poisson\Big(\frac{\beta(t) I(t) S(t)}{N}\Big), \quad C(t) \sim Poisson(\alpha E(t)) \quad \text{and} \quad D(t) \sim Poisson(\gamma I(t)) \tag{11}$$

where the parameter $\beta(t)$, α, and γ represent the time varying transmission rate, the incubation rate and the infectious rate, respectively.

2.4.2 The reproductive number

The basic reproductive number (R_0) can be defined as the average number of secondary infections that occur when one infective is introduced into a completely susceptible host population (Hethcote; 2000). This quantity is of major importance for epidemiologist and public health authority, because it can be used to characterize an epidemic, wherein $R_0 > 1$ indicates that a large epidemic outbreak may occur, and if $R_0 < 1$ then the epidemic will die out (Clancy et al.; 2008). For the model above proposed by Lekone and Finkenstädt (2006), the basic reproductive number is the ratio between the constant transmission rate (β) and the infectious rate (γ), that is $R_0 = \frac{\beta}{\gamma}$. However, the basic reproductive number only applies to the beginning of the epidemic, therefore, a time varying reproductive number, which is called effective reproductive number ($R(t)$), has been useful in the field of infectious disease. This is because it reports the expected number of new infections that an individual who becomes infectious at day t will induce (Carlin and Louis; 2008). Furthermore, it helps account for external influences on transmission such as variation in human behavior and introduction of intervention measures. Chowell et al. (2004) define the effective reproductive number as follow:

$$R(t) = \frac{\beta(t)S(t)}{\gamma N} \tag{12}$$

because $S(t) \approx N$ the effective reproductive number can be approximated by $R(t) \approx \beta(t)/\gamma$.

2.4.3 Inference and Model Building

The inference on the basic and effective reproductive number depends on the estimation of the parameters of the model.However, because the daily counts in all states are not observable, maximum-likelihood estimation is not straightforward. In this study, the only observable quantities are the population size N and the number of cases by date symptom onset $C(t)$. Therefore, if one is willing to apply maximum-likelihood based estimation, a likelihood function should have to be constructed for all the missing components in the model, and implemented through an EM algorithm (Yin; 2008; Carlin and Louis; 2008). Nevertheless, a Bayesian approach implemented through MCMC approach simplifies matters by allowing us to numerically integrate over the probability distribution of the unobserved processes (Lekone and Finkenstädt; 2006; Carlin and Louis; 2008). Moreover, it has a net benefit of producing posterior estimates for all latent information in the model, and it also allows the specification of prior information through historical study and knowledge of the experts (Stokes et al.; 2014). As the series $B(t)$, $C(t)$ and $D(t)$ are conditionally independent, then according to Lekone and Finkenstädt (2006) the likelihood function can be approximated by

$$L(\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}|\Theta) = \prod_{t=0}^{\tau^*} f_1(B(t)|\cdot) f_2(C(t)|\cdot) f_3(D(t)|\cdot) \tag{13}$$

where f_1, f_2 and f_3 are the Poisson transition density given in 11 conditioned on Θ and on all the information up to time t. In this report, to simplify matters, we assume that infectious period

$(1/\gamma)$ is known and is equal to 7 days (Lekone and Finkenstädt; 2006; Chowell et al.; 2004). Thus, on average, $\gamma I(t)$ leave the infectious compartment on day t. This implies the following

$$I(t+h) = I(t) + C(t) - \gamma I(t) \tag{14}$$

and consequently the likelihood in (13) can be written as follows

$$L(\boldsymbol{B}, \boldsymbol{C}|\Theta) = \prod_{t=0}^{\tau^*} f_1(B(t)|\cdot) f_2(C(t)|\cdot) \tag{15}$$

The parameter estimates can be obtained by maximizing (15), though the time series $B(t)$ is latent. Thereby, Bayesian MCMC method can be used to sample from the probability distribution of latent process. In this scope, multiplying the likelihood function (15) by the prior distribution $p(\Theta)$ we have, up to a constant proportionality, the posterior distribution we are going to sample from, i.e

$$p(\Theta, \boldsymbol{B}|\boldsymbol{C}) \propto L(\boldsymbol{B}, \boldsymbol{C}|\Theta) p(\Theta) \tag{16}$$

To estimate the effective reproductive number, and account for possible intervention measures, two time varying transmission models were considered:

- The transmission rate is assumed to be constant before the intervention measures, and to decay exponentially with intervention measures (Lekone and Finkenstädt; 2006) , that is

$$\beta(t) = \begin{cases} \beta, & t < t^* \\ \beta e^{-k(t-\tau)}, & t \geq t^* \end{cases} \tag{17}$$

  where t^* is the time by which intervention measures are introduced. In this report we assume that control measures were taken at the beginning of October 2014, as by the end of the year a decrease in the number of new cases of EVD is observed.

- We also model the transmission rate $\beta(t)$ as constant piece-wise function (Dureau et al.; 2013), assuming that for each month the transmission rate is constant.

For the parameters β, α and k we specified the same independent informative Gamma prior as used by Lekone and Finkenstädt (2006), i.e

$$\beta \sim Gamma(20, 100), \quad \alpha \sim Gamma(20, 100) \quad \text{and} \quad k \sim Gamma(2, 10) \tag{18}$$

The parameters β, α and k represent the transmission rate, incubation rate; and the rate at which the transmission rate decays, respectively. The specified priors in (18) have mean 0.2. In

the sampling process, the gamma priors for β and α were constrained such that $0 < \beta < 1$ and $1 < 1/\alpha < 21$, respectively.

Additionally, the estimation of the parameters depends on initialization of equations (6)-(8), so at time zero we assume that the number of exposed individuals is equal to $C(1)/\gamma$ and $I(0) = 1$. Thus, the number of susceptible individuals is $S(0) = N - E(0) - I(0)$ (Carlin and Louis; 2008).

After identifying the set of candidate models, the next step is to select the most appropriate model, that is the model that provides a better fit to the data. Bayesian statistics offers a wide range of selection criteria that can help select the best model among the candidates. The most popular and widely used is the Deviance Information Criteria (DIC) which is a measure of model "support" that intends to balance the fit (deviance) and complexity (effective number of parameters) of a model (Spiegelhalter et al.; 2002). The DIC can be seen as a counterpart of Akaike Information Criteria (AIC) in Bayesian context. The rule of thumb for using DIC is roughly the same as for AIC, in which a difference in DIC of more that 10 rules out the model with higher DIC (Lesaffre and Lawson; 2012). In this report DIC and predictive values of new cases of EVD were used to assess the fit of the model.

2.4.4 Sensitivity analysis

The sensitivity analysis is done in order to see how the conclusion change when one changes some aspects in the original model and also to check how plausible assumptions of the model are. In this report it was assumed that the number of exposed individuals at time 0 ($E(0)$) is fixed and known. A sensitivity analysis was done allowing the $E(0)$ to be estimated from data. Three different discrete uniform distribution were used to estimate $E(0)$, namely, $U(1, 50)$, $U(1, 100)$ and $U(1, 1000)$. Additionally, instead of using the gamma priors as used by Lekone and Finkenstädt (2006), uniform priors were used to estimated the transmission rate as well incubation period, that is $\beta \sim U(0, 1)$ and $\alpha \sim U(0.05, 1)$. The choice of the uniform prior in this range is motivated by the constrain made by Chowell et al. (2004) where they say that $0 < \beta < 1$ and $1 < 1/\alpha < 21$.

2.5. Spatial Analysis

Spatial analysis aims to identify and describe a spatial pattern in the data, and also helps identify spatial dependence, where observations from units close together are expected to be more similar than those relating to units further apart. In the literature there is a wide range of approaches that can be used to model the spatial dependence of the data. In this report we use a Conditional Autoregressive (CAR) model introduced by Clayton and Kaldor (1987) in an empirical Bayes setting and developed by Besag et al. (1991), to study the spatial structure of the transmission rate across the affected districts. The fitted mode is as follows:

$$Y_k|\mu_k \sim f(y_k|\mu_k, \sigma^2), \qquad g(\mu_k) = \theta + u_k + v_k \tag{19}$$

where Y_k stands for transmission rate observed in the k-th region, $g(\cdot)$ is a link function. For this report we assume that the transmission rate in different districts can be modeled by a normal distribution, thus, we use an identity link. The parameter θ will be the overall transmission rate, v_k is the uncorrelated heterogeneity which is assumed to be normally distributed, i.e. $v_k \sim N(0, \sigma_v^2)$, and u_k is the correlated heterogeneity and $u_k \sim CAR$ prior, i.e

$$u_k | u_{-k}, \boldsymbol{W}, \tau^2 \sim N\Big(\frac{\sum_{i=1}^{K} w_{ki} u_i}{\sum_{i=1}^{K} w_{ki}}, \frac{\sigma_u^2}{\sum_{i=1}^{K} w_{ki}}\Big) \tag{20}$$

where $\boldsymbol{W}$ is a $K \times K$ weigth matrix made of zeros and ones, entries with one indicate that two locations are neighbors and entries with zero indicate that they are not. Moreover, the variance parameter are given uninformative inverse gamma prior

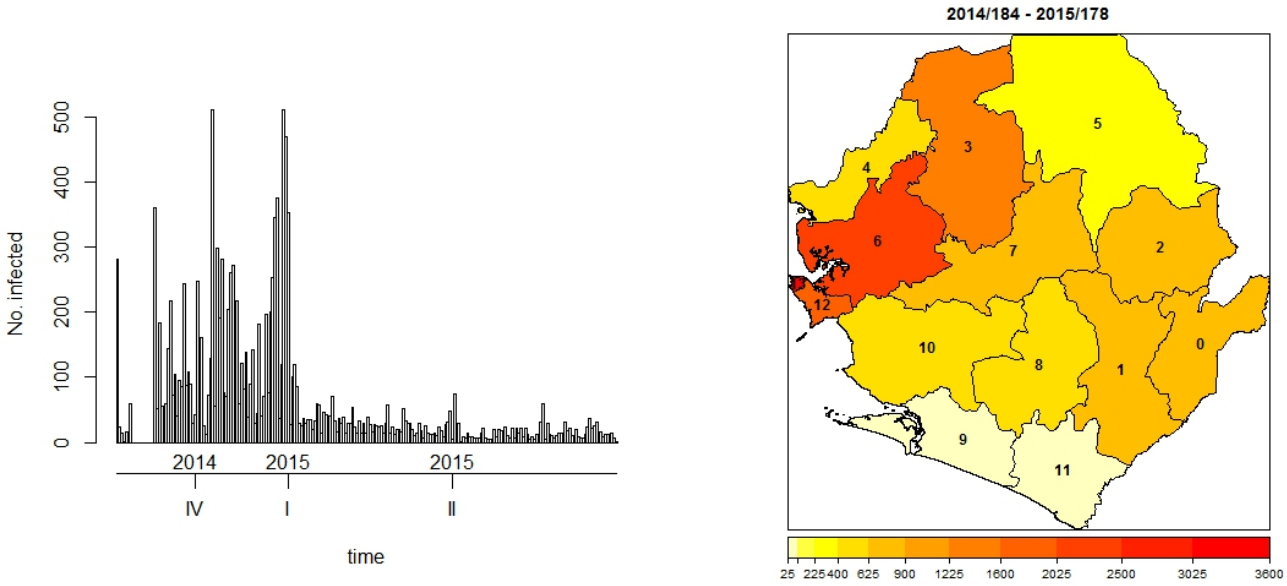
$$\sigma^2 \sim IG(a, b), \quad \sigma_u^2 \sim IG(a, b) \tag{21}$$

with $a = 0.0001$ and $b = 0.001$. In this model, σ_u^2 and σ_v^2 controls the variability of u and v, respectively. The considered CAR model was fitted to district-specific transmission rate for the first 30 days, next 30 days of the epidemic and last for the month of April 2015.

3. Results

3.1. Exploratory Data Analysis

The total number of cases of EVD overtime and per district are well depicted in Figure 2 (a) and 2 (b), respectively. From Figure 2 (a) it is observed that the epidemic had its highest peak by the end of 2014 and in the first quarter of 2015. Furthermore, after the first quarter of 2015 a decrease in the number of cases is observed, however small peaks can still be observed. This suggests that as the epidemic dies out, resurgence of new cases are possible. Turning to Figure 2 (b), it is apparent that Western Area Urban (13), Western Area Rural (12), and Port Loko (6) were the worst affected districts. While Bonthe (9) and Pujehun (11) were the least affected districts. In addition to that, one can see that districts from southeast have similar number of cases of EVD as well as the districts from northwest. This suggests a pattern in the transmission of the disease.



(a) (b)

Figure 2: Number of cases Ebola over time (a) and mapping of the number of cases of Ebola (b) per district in Sierra Leone during 07/2014-07/2015.

In Figure 3 plots of individual time series of new cases of EVD are presented. These plots support the findings suggested by Figure 2 (a) and (b). As can be seen in almost all districts, the peak of the epidemic is observed in the last months of 2014. By the end of 2015, one can see that few cases of EVD are observed, except for Western Area Urban (13) and Port Loko (6) where we still observe a substantial number of cases of EVD. This might be an indication that in Western Area Urban (13) and Port Loko (6), the epidemic is not yet under control.

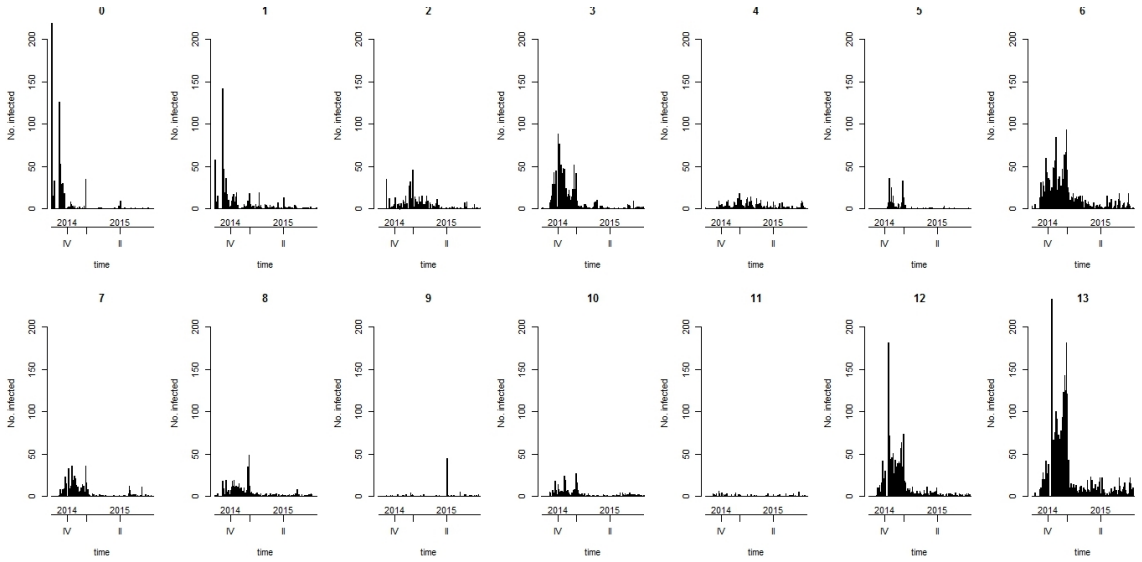


Figure 3: Individual time series of the number of cases of Ebola in the 14 affected district

Figure 4 shows the cumulative number of new cases of EVD for Bo (left panel) and Tonkolili (right panel) together with the smoothing spline fits.Since the number of cases of EVD was reported in irregular time interval, smoothing spline was fit to the cumulative cases of EVD, then the daily difference were taken to obtain the daily incidence, which were later used in subsequent analysis. From the plots in Figure 4 one can clearly see that the fitted smoothing spline provides a good fit to the the cumulative cases of EVD, suggesting that the obtained daily incidence are reasonable. The cumulative number of cases together with the smoothing spline fits for other districts can be found in Appendix 1.

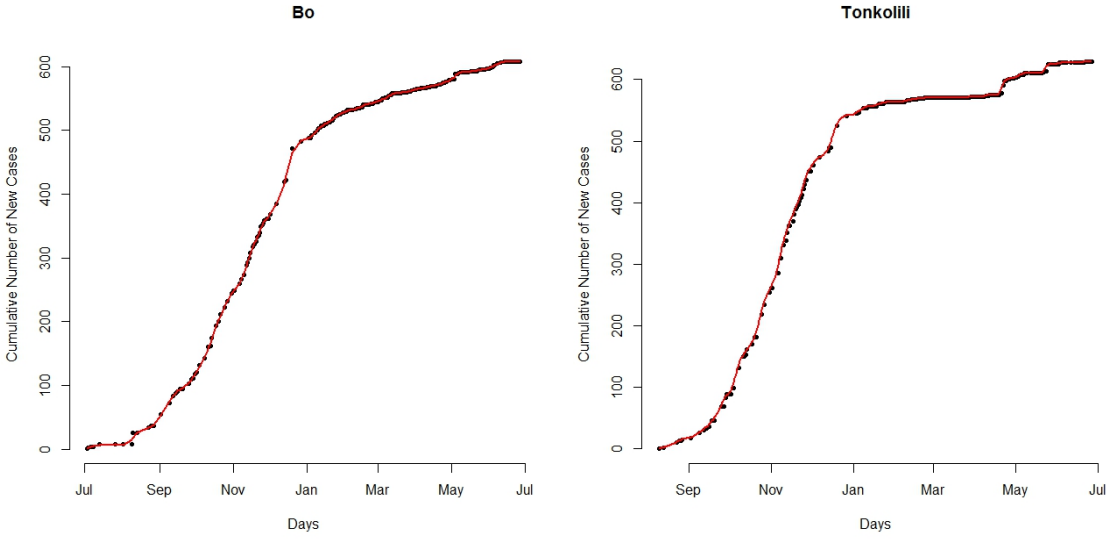


Figure 4: Black dots are the cumulative number of EVD. The solid red line represents the smoothing spline fitted to the cumulative number of new cases of EVD.

3.2. Statistical Analysis

3.2.1 Model Selection

Here we consider three candidate models. First, we consider a model which assumes a constant transmission rate (model 1). Secondly, we consider a model where the transmission rate is assumed to decay exponentially with control intervention (model 2). Finally, we assume a constant piece-wise transmission rate (model 3). The DICs together with the parameter estimates of the three mentioned models are shown in Table 1 below. Comparing the DICs of the different models, one can clearly see that model 3 shows a better fit in 10 districts(Bombali Bonthe, Kailahum, Kambia, Kenema, Koinadugu and Kono). For the incidence data from Pujehu, we can see that among the fitted model, there is no clear winner as comparing their DICs the difference is not less than 5. In addition to that, it possible to see that their parameter estimates are pretty similar, this suggests that the three models can be used to describe the epidemic in Pujehu. In Moyamba we observe that DIC is lower under model 2, and in Bo and Port Loko the DIC is lower under model 1. However, as model 1 is only applicable for the beginning of the epidemic ,and in addition to that, it does not account for external influences in the transmission rate, we therefore rule it out.

Table 1: DIC and parameter estimates under three different models: Model 1 assumes that the transmission rate is constant, model 2 assumes that it decays exponentially with control intervention and model 3 assumes a piece-wise constant transmission rate

	Model 1				Model 2				Model 3			
Districts	β	α	R	DIC	β	α	R	DIC	β	α	R	DIC
Bo	0.142	0.217	0.995	1264.2	0.165	0.18	1.154	1439.3	0.231	0.153	1.615	1794
Bombali	0.144	0.238	1.006	1707.5	0.642	0.019	4.491	1843.6	0.169	0.199	1.181	1613.1
Bonthe-	0.151	0.139	1.059	629.2	0.209	0.151	1.461	671.2	0.185	0.088	1.296	571
Kailahum	0.075	0.014	0.525	11475.2	0.241	0.01	1.684	10609.8	0.158	0.098	1.107	1050.1
Kambia	0.146	0.138	1.025	972.6	0.209	0.103	1.465	960.1	0.18	0.102	1.261	935.5
Kenema	0.127	0.198	0.888	1845	0.219	0.016	1.536	2314.2	0.05	0.244	0.353	1829.9
Koinadugu	0.141	0.107	0.989	622.7	0.32	0.094	2.237	624.9	0.236	0.103	1.65	620.1
Kono	0.125	0.041	0.874	1667	0.219	0.029	1.536	1668.7	0.101	0.189	0.705	1280.4
Moyamba	0.143	0.166	1.003	1083.5	0.161	0.157	1.124	1043.7	0.172	0.103	1.206	1075.8
Port Loko	0.143	0.374	0.999	2400	0.53	0.015	3.708	2767.8	0.134	0.133	0.937	2731.2
Pujehu	0.14	0.079	0.983	432.5	0.158	0.078	1.107	435	0.16	0.075	1.117	436
Tonkolili	0.142	0.181	0.996	1401.1	0.553	0.021	3.871	1244.2	0.232	0.148	1.624	1071.8
Western Area Rural	0.143	0.394	1.004	1341.1	0.636	0.019	4.451	1567.6	0.204	0.214	1.426	1680.7
Western Area Urban	0.144	0.156	1.011	2207	0.804	0.013	5.631	2266.3	0.164	0.068	1.151	2036

In addition to the DIC, we also assess the goodness of fit of model 3 based on the predictive values of new cases (simulated values) of EVD. Figure 5 belows we present the plots of the observed and the predictive number of new cases of EVD together with 95% credible interval. From these plots it can be seen that model 3 does not have a good fit for all incidence data, for instance, although model 3 had a lower DIC for the incidence data from Kambia, one can clear see that the predictive values do not fit well to the data. Besides Kambia, a poor fit is also observed in the incidence data from Pujehun. This suggests that the inference based on this model for the mentioned districts should be made with caution.

In table 1, based on DIC, it was observed that model 3 does not fit well the incidence data from Western Area Rural, and as the model with constant transmission rate was ruled out, model 2 was found to be the best. However, looking at the plots of observed and predictive number of new cases of EVD in Figure A1 under model 2 and 3, it is apparent that the curve of the predictive values under model 2 (left panel) is much smoother than the curve of the predictive values under model 3 (right panel). This clearly indicates that model 2 does not describe the course of the epidemic in Western Area Rural very well. Additionally, we observed that the posterior mean incubation rate under model 2 was very small (0.019) and so was the credible interval (0.017-0.020; not shown in the table). This indicates that the posterior distribution was not well explored, and therefore model 2 cannot be used to infer about the course of the epidemic in this region.

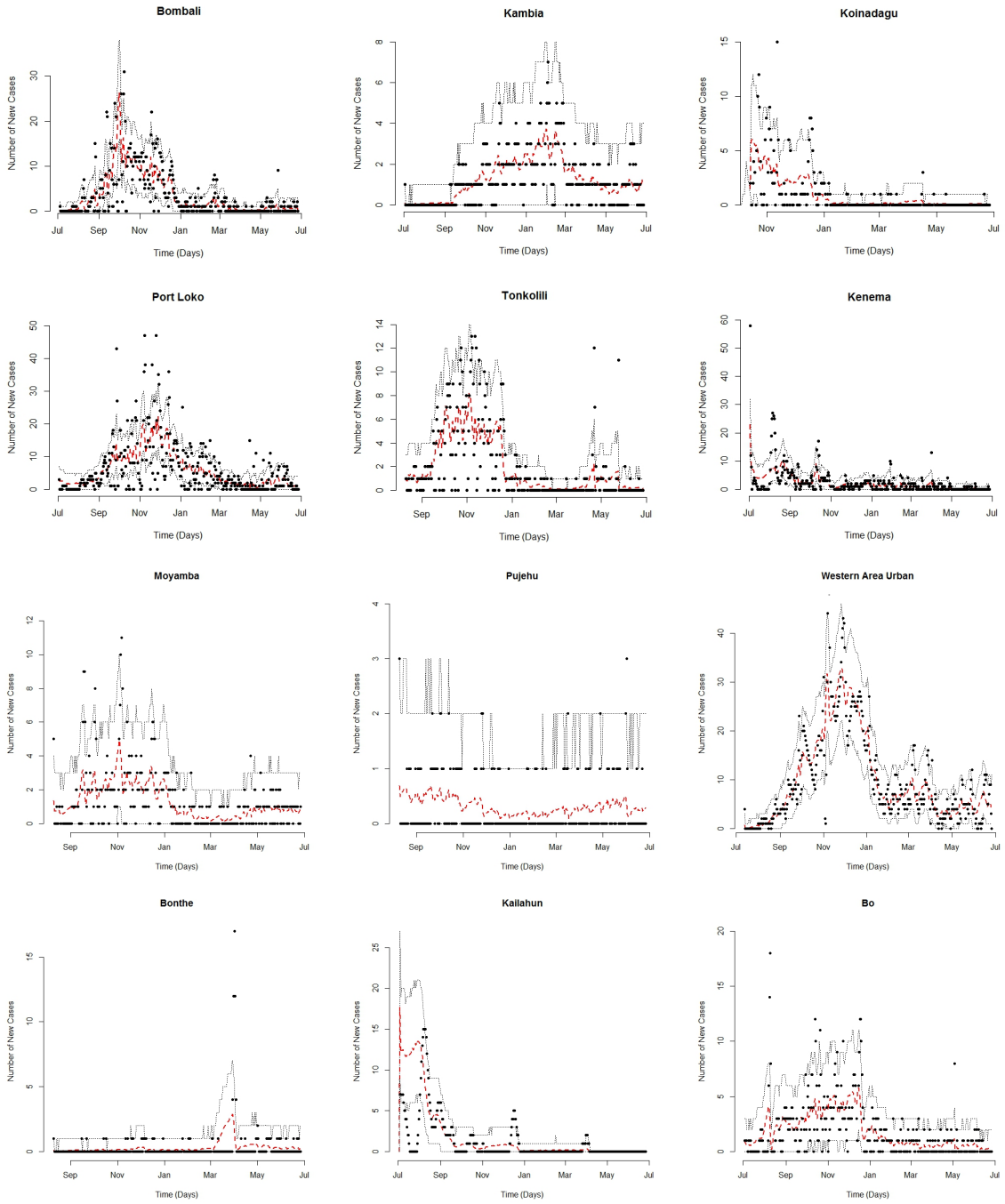


Figure 5: Observed (black dots) and predictive (red dashed line) number of new cases. Dashed lines are the 95% credible interval

Similarly in Moyamba, although the DIC was found to be smaller under model 2, looking at the predictive number of new cases of EVD in Figure A1, we also observed that both models can reasonably describe the course of the epidemic.

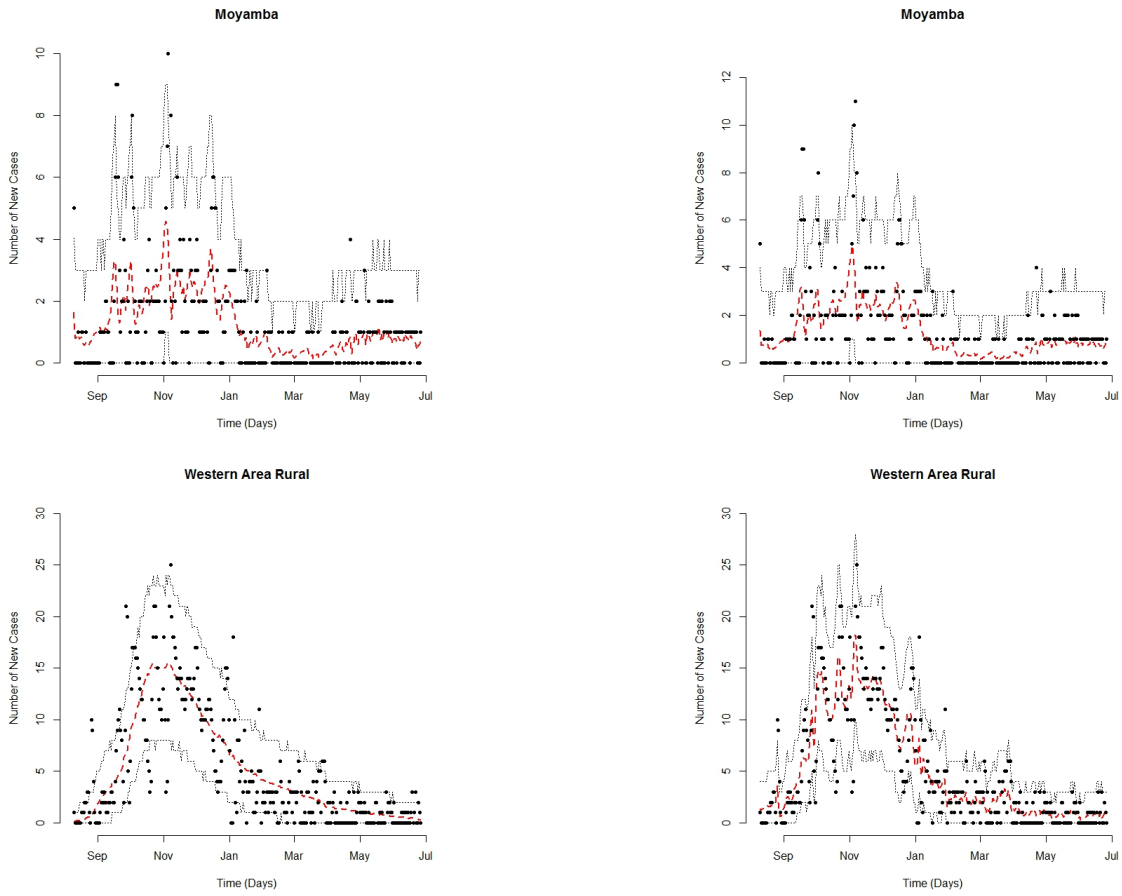


Figure 6: Observed (black dots) and predictive (red dashed line) number of new cases. Dashed lines are the 95% credible interval. Comparison of model 2 (left panel) and 3 (right panel) under Western Area Rural, Moyamba and Bo

All the analysis done in this section points out that model 3 can well describe the course of the epidemic. Therefore, model 3 is considered throughout the report. However, it is worth emphasizing that there is a substantial variability around the predictive number of new cases of EVD. This is not so surprising, because by nature, a stochastic model introduces uncertainty (Lekone and Finkenstädt; 2006; King et al.; 2015). Furthermore, note that the number of cases of EVD used to fit the model are not the true observed cases, but the number of cases that were obtained based on the fitted smoothing splines to the cumulative number of new cases of EVD. Therefore, this can also introduce variability in the data.

3.2.2 Estimating the reproductive number per district

Here, we present the curves of the effective reproductive number per district. In addition to that, we also present the posterior mean of the reproductive number in the month of January and the posterior mean of the incubation period for each district.

Table 2 shows the posterior mean of the reproductive number for the month January as well as the incubation period for each district. It can be seen that in all districts, except Kambia, the reproductive number is less than one. This is in agreement with the exploratory data analysis, where as of January, we observed a decrease in the number of new cases of EVD (see Figure 2 (a)). The estimated average incubation period ranges from 4.11 to 14.73 days, with kenema having the least incubation period and Western Area Urban with the highest incubation period. A clear distribution of the incubation period can be see in 3.2.4.

Table 2: Posterior mean for the reproductive number and incubation period together with 95% credible interval for the month of January 2015

	Reproductive number R		Incubation period $1/\alpha$	
Districts	Mean	CI	Mean	CI
Bo	0.804	0.57-1.063	7.41	5.413-10.432
Bombali	0.724	0.525-0.955	6.40	5.157-7.782
Bonthe	1.112	0.723-1.61	11.82	7.751-16.682
Kailahum	0.919	0.571-1.381	10.3	9.303-12.207
Kambia	1.311	1.017-1.62	10.28	6.638-14.948
Kenema	1.007	0.772-1.292	4.11	3.721-4.571
Koinadugu	0.408	0.255-0.578	9.79	8.307-11.710
Kono	0.968	0.777-1.176	5.32	4.507-6.094
Moyamba	1.073	0.826-1.33	7.40	5.13-10.394
Port Loko	0.724	0.597-0.859	7.60	5.922-9.520
Pujehu	1.161	0.709-1.713	13.86	9.214-20.186
Tonkolili	0.699	0.545-0.864	6.90	5.706-9.525
Western Area Rural	0.832	0.729-0.948	4.97	3.559-6.695
Western Area Urban	0.337	0.243- 0.442	14.73	12.322-17.338

Figure 7 shows the curves of the estimated effective reproductive number. It can be seen that the estimated effective reproductive number ranges between values less than 1 and around 3. Around January to May it is possible to see that almost all district have $R_e(t)$ less than 1. Moreover, for the last observed period some districts show estimate of $R_e(t)$ greater than 1. However, from May to the last period of the epidemic we can observe that Western Area Rural/Urban, Port Loko, Kambia, Bombali and Bo show an oscillation of $R_e(t)$ aroud the threshold value 1, suggesting that reappearance of new cases cannot be ruled out in these districts. It is also important to stress that, although the $R_e(t)$ for Pujehu is larger than 1, the credible interval covers the threshold (R=1) suggesting that the epidemic in this region is under control , but little can be said due to the fact that there is a substantially large variability around the it. It also is worth emphasizing that for all districts the estimated $R_e(t)$ have a quite large variability.

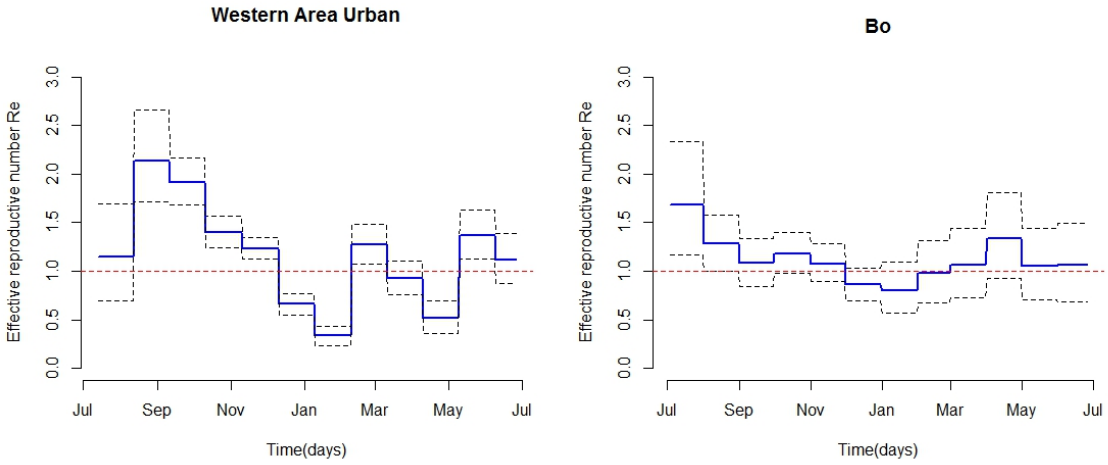


Figure 7: Effective reproductive number and 95% credible interval for the 14 district of Sierra Leone. The red dotted line represents $R = 1$, which is the threshold for control. Continues in Figure 8

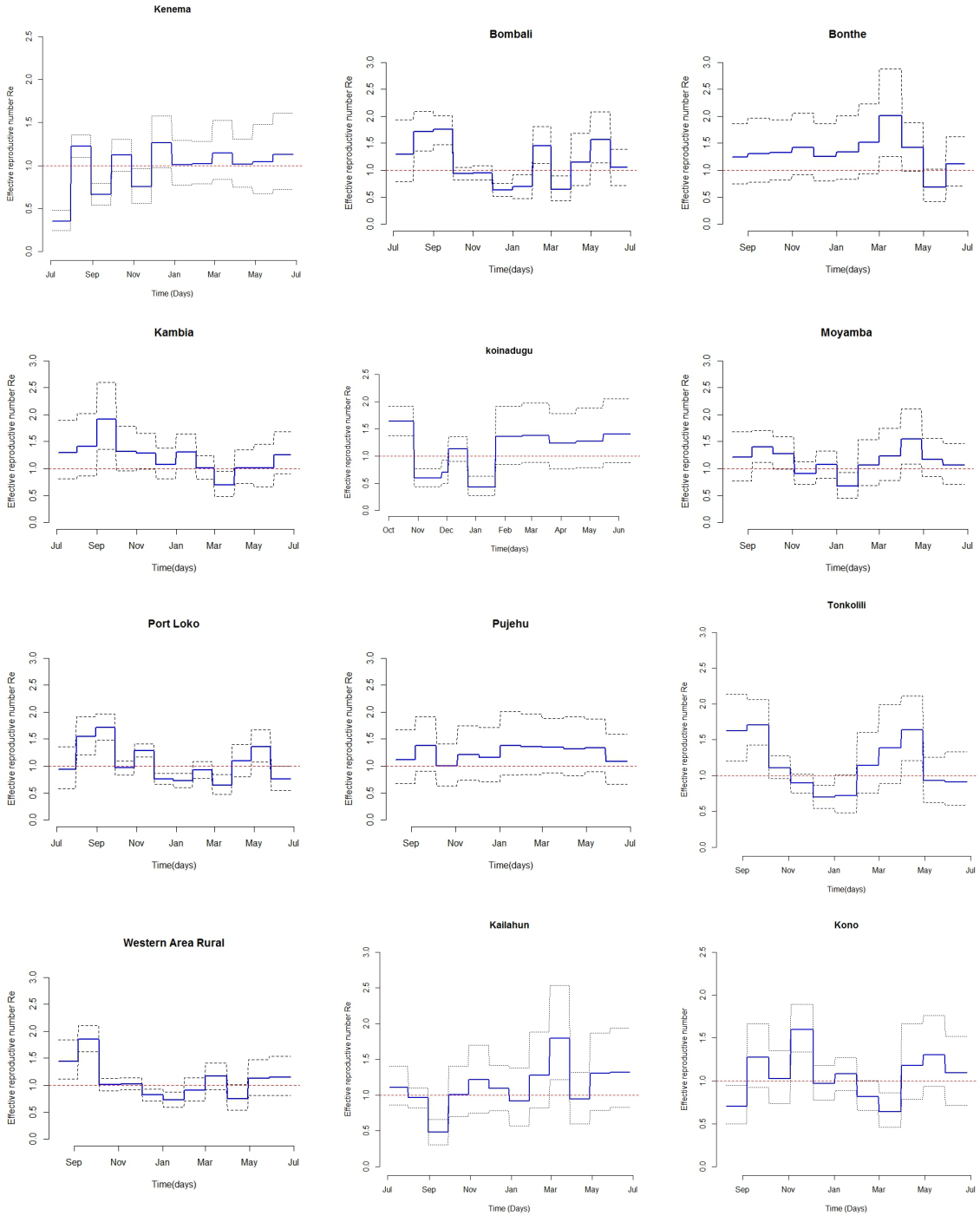


Figure 8: Effective reproductive number and 95% credible interval for the 14 district of Sierra Leone. The red dotted line represents $R = 1$, which is the threshold for control.

3.2.3 Sensitivity analysis

Table 3 shows parameter estimate for the model assuming a constant piece-wise transmission rate under different prior information. It can bee seen that estimates under vague uniform prior has much narrower credible intervals than the estimates under informative gamma prior, and they are substantially different. Additionally, it is worth noting that effective reproductive number for the second month ($R(2)$) in both scenario is significantly larger than 1. Though, the DIC of the model considering vague uniform prior is smaller, the estimate of the incubation rate under uniform prior is quite close to the boundary of the uniform prior distribution. This may be an indication that the posterior distribution of the incubation rate under vague uniform prior is not well explored. Thus, inferences based on vague uniform prior, should be made with caution.

Table 3: Parameter estimates together with 95% credible interval for model with a constant piece-wise transmission rate considering a vague uniform prior and and an informative gamma prior. Western Area Urban

	Scenario 1		Scenario 2	
Prior	$\beta \sim U(0,1)$, $\alpha \sim U(0.05,1)$		$\{\beta,\alpha\} \sim Gamma(20,100)$	
Parameters	Mean	CI	Mean	CI
$\beta(1)$	0.069	0.003-0.233	0.164	0.103-0.238
$\beta(2)$	0.433	0.345-0.53	0.303	0.241-0.369
α	0.051	0.05-0.055	0.068	0.057-0.08
$R(1)$	0.486	0.018-1.631	1.151	0.721-1.669
$R(2)$	3.032	2.414-3.711	2.123	1.685-2.585
DIC	1987.9		2036	

In table 4, we present a sensitivity analysis. In the first scenario, we estimate the parameters with the assumption that the number of exposed is fixed. In the other scenario, we estimate the parameters from the data by placing uniform prior on the number of exposed of individuals at time zero. Three discrete uniform prior in the range $(1,50)$, $(1,100)$ and $(1,1000)$ were considered. According to Lekone and Finkenstädt (2006) the incubation period can be approximated by a geometric distribution with mean $1/p_c$. Previous studies have reported an incubation period of 7 days (Center for Disease Control; 2015b; Chowell et al.; 2004). This implies that $p_C = 0.143$, thus, we use an estimated number of exposed at time zero, which is compatible with p_C, i.e. $E(0) = 7C(1)$. As at the beginning the number of cases in Western Area Urban is 4, then $E(0)$ is set to 28. From this table, we observe that fixing $E(0)$ to 28 leads to a model with a lower

DIC. We also noticed that the incubation period for model 4 is quite large (14.73 days) compared to the incubation period estimated from the scenario where $E(0)$ is estimated from the data. Additionally, it is also worth noting that, although the DIC of scenario 1, 2 and 3 are quite different, the estimates and the credible interval are pretty similar. Furthermore, the credible intervals of the aforementioned scenarios are wider than credible interval of scenario 4. This suggests that estimating $E(0)$ from data introduces extra variability in the parameter estimates.

Table 4: Comparison of the parameter estimates and 95% credible interval considering that the number of exposed at time 0 is estimated from data using a uniform prior and fixing ()$E(0) = 28$).

	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
Prior	$E(0) = 28$		$E(0) \sim U(1, 50)$		$E(0) \sim U(1, 100)$		$E(0) \sim U(1, 1000)$	
Parameter	Mean	CI	Mean	CI	Mean	CI	Mean	CI
$\beta(1)$	0.164	0.103-0.238	0.22	0.143-0.303	0.217	0.152-0.295	0.218	0.147-0.299
$\beta(2)$	0.303	0.241-0.369	0.246	0.194-0.303	0.239	0.19-0.297	0.241	0.193-0.295
$1/\alpha$	14.73	12.322-17.338	6.128	3.998-9.524	5.577	3.568 -8.064	5.712	4.087-8.018
$R(1)$	1.151	0.721-1.669	1.543	1.001-2.123	1.519	1.065-2.063	1.526	1.028-2.091
$R(2)$	2.123	1.685-2.585	1.72	1.361-2.118	1.676	1.331-2.08	1.684	1.348-1.348
$E(0)$		4.423	3-7	4.261	3-6	4.254	3-6	
DIC	2036		2660		2887.6		2493.8	

3.3. **Spatial Analysis**

In this section, we study the spatial distribution of the transmission rate across the country of Sierra Leone. First, we begin by doing an exploratory data analysis by displaying the maps of estimated incubation period, basic reproductive number and the estimated transmission rates. Then, using a conditional autoregressive model we explore the structured and the unstructured heterogeneity of the estimated transmission rate across the country. The transmission rate is analyzed at three time points, namely, first 30 days of the epidemic, next 30 days and for the month of April 2015.

Figure 9 shows the estimated incubation period (left panel) and basic reproductive number (right panel) for each district across the country. From the map of the estimated incubation period one can clearly see that for most of the districts the estimated incubation period rages between 5 and 7.5 days. In addition to that, one can also see that the districts with the estimated incubation period in this range are located in the center of the country. Two districts had the estimated incubation period greater than 12 days, namely: Western Area Urban in the northwest region of the country and Pujehun in the south region of the country. Looking at the map of the estimated basic reproductive number, it is clearly seen than most districts have the estimated basic reproductive number greater than one.

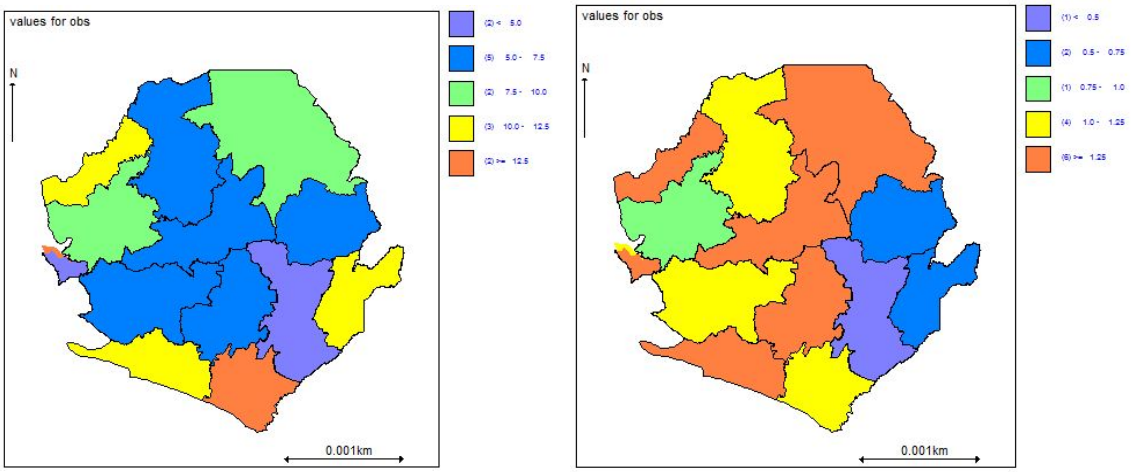


Figure 9: Illustration of the estimated incubation period (left panel) and the basic reproductive (right panel) number across the districts of Sierra Leone

In Figure 10, we give an illustration of the transmission rates across the districts of Sierra Leone in order to see how the transmission rate varies across the different districts. The maps were plotted at three different selected time points. Firstly, for the first 30 days, secondly for the following 30 days and the last case for the month of April. From the maps, one can see that for the first 30 days (top map on the left), there is no substantial difference in the transmission rate across the districts. Moreover, the map suggests some clustering as we can see districts with similar transmission rates close together. Looking at the map for the second selected time point, it can be seen that it is much smoother than the previous one, suggesting more similarity in the transmission rate among the districts. On the third map (bottom on the left), although we see a different pattern, it can be seen that compared to the latter period the transmission rate in many districts has decreased.

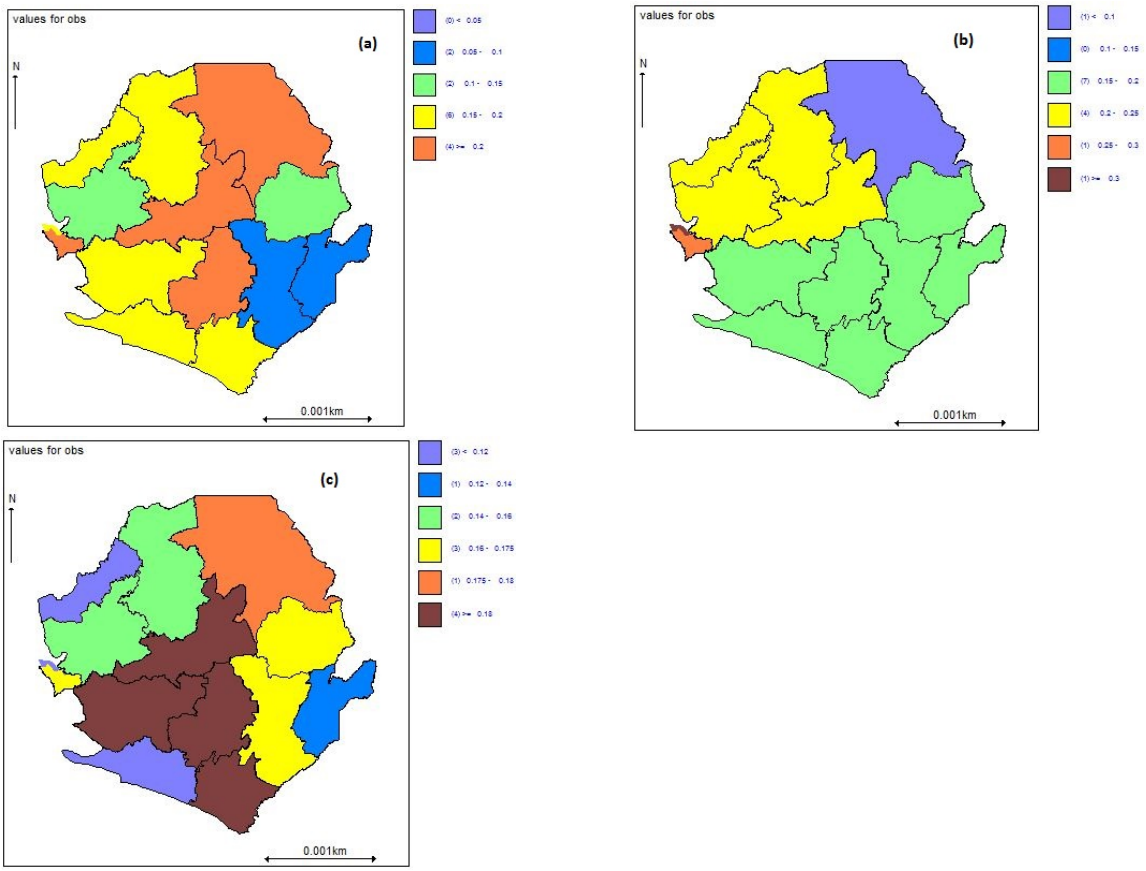


Figure 10: Illustration of the estimated transmission rate across the region at three different selected time points : Map (a) transmission rate for the first month days of the epidemic; Map (b) the transmission rate for the second month of the epidemic; Map (c) transmission rate for the April 2015

Now that we have estimated the district-specific transmission rate, we fit a CAR model in order to study the spatial structure across the affected region . Table 5 shows the parameter estimates from the CAR model. In all the fitted models the there is spatially structured variation as we can see from the table. Furthermore, it is worth noting that in all the fitted models the spatial structured variation is quite small, indicating the presence for the presence of spatial correlation between neighboring district . Moreover, from plot (b) Figure 4, one can see the presence of two outstanding regions with region from the southeast (green) having a lower transmission rate as opposed to its counterpart from the northwest region (yellow) of clustering of district-specific transmission rate.

Table 5: Parameter estimates and 95% credible interval for the CAR model: Period 1 considers the district-specific transmission rate for the 1^{st} 30 days of the epidemic; Period 2 considers district-specific transmission rate for the next 30 days of the epidemic; and period 3 consider district-specific transmission rate for the month of April 2015.

	Period 1		Period 2		Period 3	
Parameter	Mean	CI	Mean	CI	Mean	CI
β	0.1675	0.1343-0.2006	0.2054	0.1769-0.2342	0.1573	0.1311-0.1839
σ^2	0.002565	0.0004703-0.007457	0.001864	0.0003793-0.005493	0.001585	0.00037-0.004451
σ_u^2	0.002253	0.0001848-0.01017	0.002306	0.0001993-0.009327	0.001427	0.000163-0.006031
σ_v^2	0.001316	0.0001738-0.004673	0.001011	0.0001571-0.003547	8.86E-04	0.0001559-0.00288

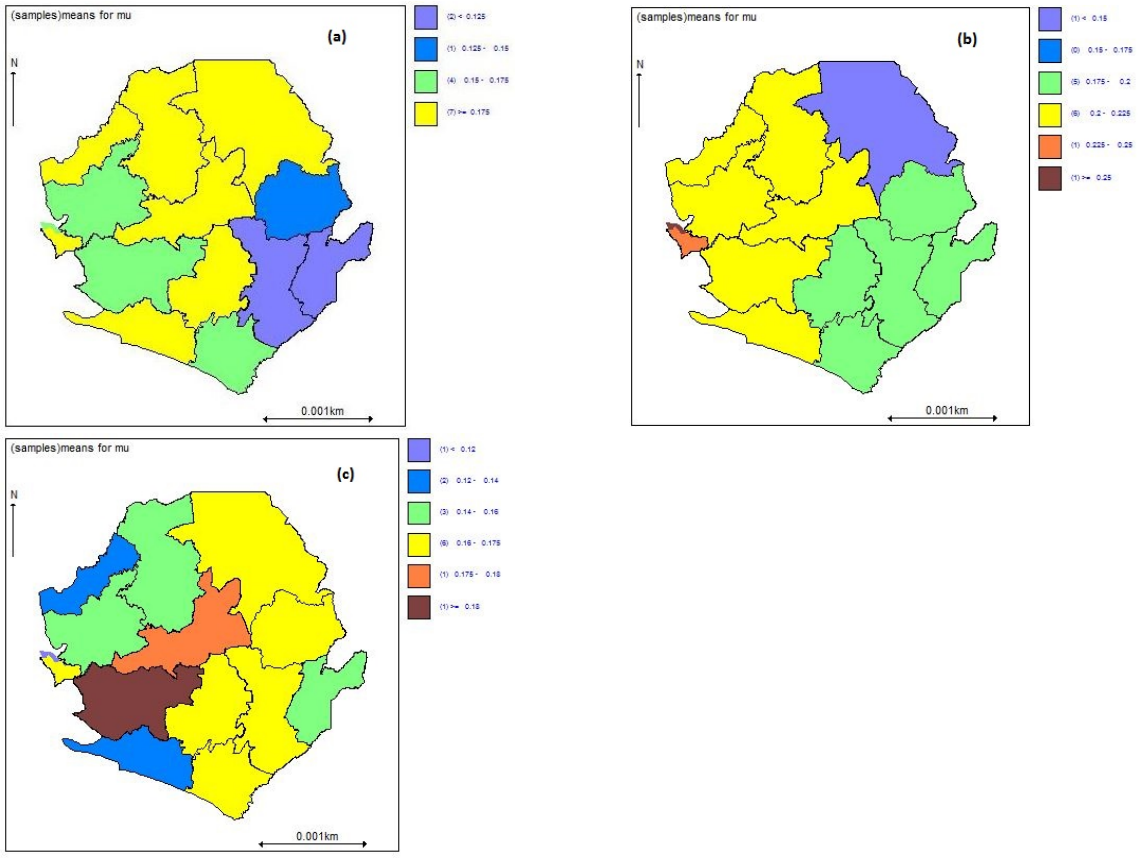


Figure 11: Illustration of the estimated transmission rate across the region at three different selected time points after smoothing : Map (a) estimated transmission rate for the first 30 days of the epidemic; Map (b) estimated transmission rate for the second month of the epidemic; Map (c) estimated transmission rate for the April 2015.

4. Conclusion and Discussion

On the 8^{th} of August 2014, the World Health Organization (WHO) declared the Ebola outbreak as an international public health problem of extreme emergency after it had started in Guinea and then spread to the neighboring countries of Liberia and Sierra Leone (World Health Organization; 2014b,a). Since the Ebola outbreak started, it has killed more than 11,000 people, and a number of control measures have been applied in order to control the spread of the disease. In addition to that, epidemiologist and mathematicians have been studying the dynamic transmission and predicting the number of new cases using a variety of deterministic SEIR model. Nevertheless, the deterministic SEIR model has been extensively criticized in the literature, because they do not account for stochasticity, and they usually lead to overestimated predictions and underestimated uncertainty. Furthermore, the models are fitted using a cumulative number of new cases by means of ordinary least square, which assumes that the measurement errors are independent and normally distributed with constant variance. However, when the cases are cumulated, this assumption does not hold (King et al.; 2015). To overcome issues related to deterministic SEIR model, in this report, we used a stochastic discrete-time SEIR model developed by Lekone and Finkenstädt (2006) to estimate the reproduction number across the districts, to study how it evolves overtime, and also to estimate the average incubation period. The stochastic discrete-time SEIR model was fitted to daily incidence data of each district of Sierra Leone.

Turning to the findings, the stochastic discrete-time SEIR model was first fitted to all districts considering three scenario. In the first cases, we considered a constant transmission rate, while in the second case we assumed it to decay exponentially with control intervention, and finally, we assumed a piece-wise constant transmission rate. Results showed that the model with piece-wise constant transmission rate provides a better fit to the data. Furthermore, the model is capable of predicting if there will be resurgences of new cases of Ebola or not. In general, the model had good a fit to all incidence data, except for the incidence data from Kambia, Tonkolili, and Pujehu. Despite the good fit, it was observed that the predictive values together with the parameter estimates had substantially wide credible intervals.However, this is not surprising since according to Lekone and Finkenstädt (2006), the analysis based on stochastic discrete-time SEIR model is fully probabilistic, and in addition to that, many events and variables are unobserved. Therefore, this induces variability which is reflected in the posterior estimates.

Between July and August 2014, the R_t in Kenema, Kono, and Kalilahum was less than one, indicating that in this period the epidemic seemed to be under control. These results are in agreement with Team et al. (2014) who analyzed the first 9 months of the epidemic in West African affected countries and they report a substantial decrease of R_t between June and August since the case incidence was stabilized. In Koinadugu, R_t was found to be less than one between November 2014 and January 2015, and from February to June a reproduction number larger than one was

observed. However, the credible intervals suggested that the disease is under control as it included the threshold value ($R = 1$). Similarly, Pejehun and Bonthe, though the R_t is greater than one, the credible intervals showed evidence that the disease is in endemic state. These results line up with what is currently being seen in the field; as 9 weeks have passed since the last case was observed (World Health Organization; 2015a). The reproduction number in Port Loko, Western Area Urban and Tonkolili decreased to less than one from June and January, though as of February, R_t has been oscillating around the control value, suggesting that reappearance new cases of EVD is likely in these districts.

Recently, Satermans et al. (2015) studied the spatiotemporal evolution of Ebola virus disease at sub-national level. They found that between January and May 2015, the reproductive number in Western Area Urban was less than 1. More still , their results suggested a decrease in the reproductive number between June and May 2015. These findings are in agreement with what we observed for Western Area Urban. Our results are also consistent with the findings by Camacho et al. (2015), who studied the temporal changes of Ebola in Sierra Leone. For the month of January 2015, they reported a reproductive number of 0.32 for Western Area, which is in line with what we found.

In our analyses, for some districts, we observed that the basic reproduction number was less than 1, but the epidemic still persisted. This is not surprising as Li et al. (2011) state that, when stochastic effects are included in the SEIR model, the threshold at $R_0 = 1$ may be disturbed as well as variations in individuals parameters. Therefore, the reported results should be interpreted with caution.

Furthermore, we performed a sensitivity analysis considering only the incidence data of Western Area Urban. The choice of different priors did not affect the posterior mean noticeably, except for the incubation period. Under uniform prior, for instance, the incubation rate was much smaller than under the gamma prior. This indicates that, an individual will stay an average period of 19 days before they become infectious. However, this average incubation period may be considered unrealistic as the average incubation period is around 8 to 10 days. Additionally, when we estimate the number of exposed individuals at time zero, the posterior mean did not vary considerably as compared to when we assume that it is known. Nevertheless, it was observed that when we place a uniform prior on the number of exposed individuals at time zero the credible intervals were wider . This suggests that allowing the number of exposed individuals to be estimated from the data introduces variability in the posterior mean.

The model used in this report is subject to some limitation. It assumes that the disease can only be transmitted through contact with infected people, while it is known that the disease can also be transmitted through contact with dead bodies. We also assume that the infectious rate is similar to the previous outbreak, however, Meltzer et al. (2014) warn that this approach may lead inaccurate results. Moreover, the stochastic discrete-time SEIR model proposed by Lekone and Finkenstädt (2006), assumes perfect reporting of the new cases of EVD, while in reality this is not the case . In this report we did not account for underreporting. Ignoring this, may lead to bias when making inference for the model parameters (Gamado et al.; 2014). We therefore, recommend that all aforementioned drawbacks of the model should be taken into account for further research.

Finally, after having estimated the transmission rate for all districts, we investigated the spatial structure of transmission rate across the country. Results did not indicate much differences in the transmission rate across the regions. However, it we noticed that districts from northwest had a higher transmission rate than districts from the south were found have similar and lower transmission rate. In conclusion, our analysis suggests a decline in the transmission rate during February to June 2105. Moreover, it was observed that around the end of June and beginning of July 2015 the R_t was around the control value indicating that the disease is under control

Bibliography

Althaus, C. L. (2014). Estimating the reproduction number of ebola virus (ebov) during the 2014 outbreak in west africa, *PLoS currents* **6**.

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., Soropogui, B., Sow, M. S., Keïta, S., De Clerck, H. et al. (2014). Emergence of zaire ebola virus disease in guinea, *New England Journal of Medicine* **371**(15): 1418–1425.

Bashar, S., Percy, M. and Singhai, R. (n.d.). Predicting the 2014 ebola outbreak in west africa using network analysis.

Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics, *Annals of the institute of statistical mathematics* **43**(1): 1–20.

Breman, J., Piot, P., Johnson, K., White, M., Mbuyi, M., Sureau, P., Heymann, D., Van Nieuwenhove, S., McCormick, J., Ruppol, J. et al. (1978). The epidemiology of ebola hemorrhagic fever in zaire, 1976, *Ebola virus haemorrhagic fever* pp. 103–124.

Camacho, A., Kucharski, A., Aki-Sawyerr, Y., White, M. A., Flasche, S., Baguelin, M., Pollington, T., Carney, J. R., Glover, R., Smout, E. et al. (2015). Temporal changes in ebola transmission in sierra leone and implications for control requirements: a real-time modelling study, *PLoS currents* **7**.

Carlin, B. P. and Louis, T. A. (2008). *Bayesian methods for data analysis*, CRC Press.

Center for Disease Control (2015a). 2014 ebola outbreak in west africa - case counts, `http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/case-counts.html`. [Online; accessed 22-August-2015].

Center for Disease Control (2015b). Ebola virus disease, `www.cdc.gov/vhf/ebola/pdf/ebola-factsheet.pdf`. [Online; accessed 16-August-2015].

Chandler, R. and Scott, M. (2011). *Statistical methods for trend detection and analysis in the environmental sciences*, John Wiley & Sons.

Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W. and Hyman, J. (2004). The basic reproductive number of ebola and the effects of public health measures: the cases of congo and uganda, *Journal of Theoretical Biology* **229**(1): 119–126.

Clancy, D., O'Neill, P. D. et al. (2008). Bayesian estimation of the basic reproduction number in stochastic epidemic models, *Bayesian Analysis* **3**(4): 737–757.

Clayton, D. and Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics* pp. 671–681.

Dureau, J., Kalogeropoulos, K. and Baguelin, M. (2013). Capturing the time-varying drivers of an epidemic using stochastic dynamical systems, *Biostatistics* **14**(3): 541–555.

Erickson, J. (2015). Faulty modeling studies led to overstated predictions of ebola outbreak, `http://ns.umich.edu/new/releases/22783-faulty-modeling-studies-led-to-overstated-predictions-of-ebola-outbreak`. [Online; accessed 22-August-2015].

Feldmann, H., Nichol, S. T., Klenk, H.-D., Peters, C. J. and Sanchez, A. (1994). Characterization of filoviruses based on differences in structure and antigenicity of the virion glycoprotein, *Virology* **199**(2): 469–473.

Gamado, K. M., Streftaris, G. and Zachary, S. (2014). Modelling under-reporting in epidemics, *Journal of mathematical biology* **69**(3): 737–765.

Greenwood, M. (1931). On the statistical measure of infectiousness, *The Journal of hygiene* **31**(3): 336.

Hens, N., Shkedy, Z., Aerts, M., Faes, C., Van Damme, P. and Beutels, P. (2012). *Modeling infectious disease parameters based on serological and social contact data: a modern statistical perspective*, Vol. 63, Springer Science & Business Media.

Hethcote, H. W. (2000). The mathematics of infectious diseases, *SIAM review* **42**(4): 599–653.

Keeling, M. J. and Rohani, P. (2008). *Modeling infectious diseases in humans and animals*, Princeton University Press.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 115, The Royal Society, pp. 700–721.

King, A. A., de Cellès, M. D., Magpantay, F. M. and Rohani, P. (2015). Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to ebola, *Proceedings of the Royal Society of London B: Biological Sciences* **282**(1806): 20150347.

Lekone, P. E. and 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.

Lesaffre, E. and Lawson, A. B. (2012). *Bayesian biostatistics*, John Wiley & Sons.

Li, J., Blakeley, D. and Simith, R. J. (2011). The failure of r0. computational and mathematical methods in medicine, (527610): 17.

Meltzer, M. I., Atkins, C. Y., Santibanez, S., Knust, B., Petersen, B. W., Ervin, E. D., Nichol, S. T., Damon, I. K. and Washington, M. L. (2014). Estimating the future number of cases in the ebola epidemic—liberia and sierra leone, 2014–2015, *MMWR Surveill Summ* **63**(suppl 3): 1–14.

Nishiura, H. and Chowell, G. (2014). Early transmission dynamics of ebola virus disease (evd), west africa, march to august 2014, *Euro Surveill* **19**(36): 20894.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Sanchez, A., Trappier, S. G., Mahy, B., Peters, C. J. and Nichol, S. T. (1996). The virion glycoproteins of ebola viruses are encoded in two reading frames and are expressed through transcriptional editing, *Proceedings of the National Academy of Sciences* **93**(8): 3602–3607.

Sanitation Moha (2015). Ebola situation, `http://health.gov.sl/?page_id=583`. [Online; accessed on 1-July-2015].

Satermans, E., Robesyn, E., Ganyani, T., Sudre, B., Faes, C., Quinte, C., Van Bortel, W., Haber, T., Kovac, T., Van Reeth, F., Testa, M., Hens, N. and Plachouras, D. (2015). Spatiotemporal evolution of ebola virus disease at sub-national level during the 2014 west africa epidemic: model scrutiny and data meagreness, *PLOS ONE* .

Sierra Leone Ministry of Health and Sanitation (2015). Ebola virus disease - situation report–21 august, 2015, `http://health.gov.sl/?page_id=583`. [Online; accessed 22-August-2015].

Siettos, C. I. and Russo, L. (2013). Mathematical modeling of infectious disease dynamics, *Virulence* **4**(4): 295–306.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4): 583–639.

Stokes, M., Chen, F. and Gunes, F. (2014). An introduction to bayesian analysis with sas/stat® software, *Proceedings of the SAS Global Forum 2014 Conference, SAS Institute Inc, Cary, NC*.

Team, W. E. R. et al. (2014). Ebola virus disease in west africa—the first 9 months of the epidemic and forward projections, *N Engl J Med* **371**(16): 1481–95.

Tuckwell, H. C. and Williams, R. J. (2007). Some properties of a simple stochastic epidemic model of sir type, *Mathematical Biosciences* **208**(1): 76–97.

Vynnycky, E. and White, R. (2010). *An Introduction to Infectious Disease Modelling*, OUP Oxford.

World Health Organization (2014a). Ebola response roadmap situation report 15 october 2014, `http://apps.who.int/iris/bitstream/10665/136508/1/roadmapsitrep15Oct2014.pdf?ua=1`. [Online; accessed 25-August-2015].

World Health Organization (2014b). Statement on the 1st meeting of the ihr emergency committee on the 2014 ebola outbreak in west africa, http://www.who.int/mediacentre/news/statements/2014/ebola-20140808/en/. [Online; accessed 25-August-2015].

World Health Organization (2015a). Ebola situation report - 19 august 2015, http://apps.who.int/ebola/current-situation/ebola-situation-report-19-august-2015. [Online; accessed 22-August-2015].

World Health Organization (2015b). Ebola situation report - 24 june 2015, http://apps.who.int/ebola/current-situation/ebola-situation-report-24-june-2015. [Online; accessed 22-August-2015].

World Health Organization (2015c). Ebola virus disease, http://www.who.int/mediacentre/factsheets/fs103/en/. [Online; accessed 15-August-2015].

Yamin, D., Gertler, S., Ndeffo-Mbah, M. L., Skrip, L. A., Fallah, M., Nyenswah, T. G., Altice, F. L. and Galvani, A. P. (2015). Effect of ebola progression on transmission and control in liberia, *Annals of internal medicine* **162**(1): 11–17.

Yin, Y. (2008). *Bayesian analysis of infectious disease time series data and optimal constrained Bayesian updating*, THE JOHNS HOPKINS UNIVERSITY.

Yu, B., Wang, J., McGowan, M., Vaidyanathan, G. and Younger, K. (2010). Gryphon: a hybrid agent-based modeling and simulation platform for infectious diseases, *Advances in Social Computing*, Springer, pp. 199–207.

Appendix

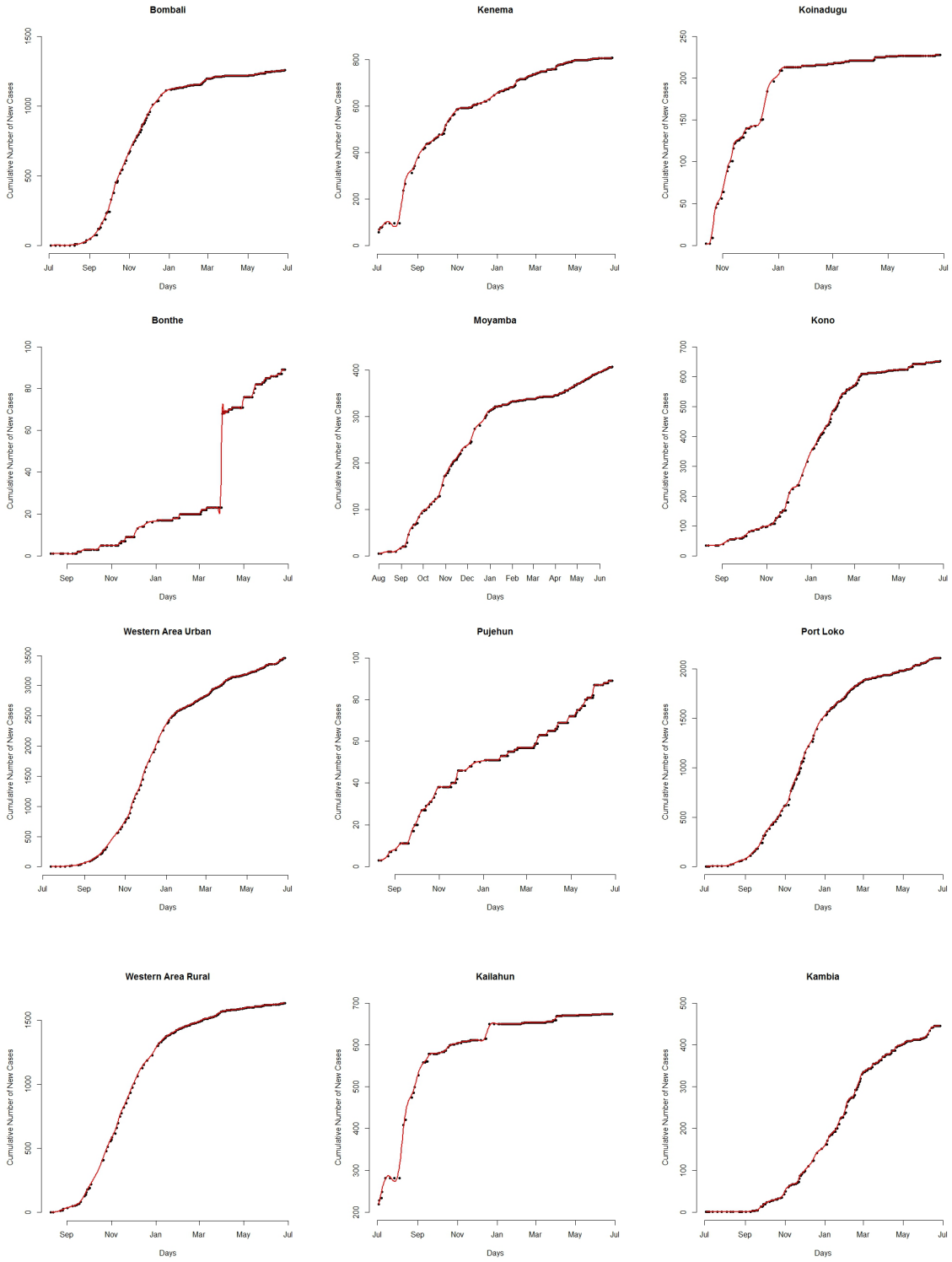


Figure A1: Black dots are the cumulative number of EVD. The solid red line represents the smoothing spline fitted to the cumulative number of new cases of EVD.

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Temporal spatial modelling of Ebola

Richting: **Master of Statistics-Biostatistics**
Jaar: **2015**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Muleia, Rachid Joel Guidion

Datum: **1/09/2015**