Patch Models of EVD Transmission Dynamics

Bruce Pell, Javier Baez, Tin Phan, Daozhou Gao, Gerardo Chowell and Yang Kuang

Abstract Mathematical models have the potential to be useful to forecast the course of epidemics. In this chapter, a family of logistic patch models are preliminarily evaluated for use in disease modeling and forecasting. Here we also derive the logistic equation in an infectious disease transmission context based on population behavior and used it for forecasting the trajectories of the 2013–2015 Ebola epidemic in West Africa. The logistic model is then extended to include spatial population heterogeneity by using multi-patch models that incorporate migration between patches and logistic growth within each patch. Each model's ability to forecast epidemic data was assessed by comparing model forecasting error, parameter distributions and parameter confidence intervals as functions of the number of data points used to calibrate the models. The patch models show an improvement over the logistic model in short-term forecasting, but naturally require the estimation of more parameters from limited data.

1 Introduction

The 2013–2015 Ebola epidemic in West Africa has become the most severe Ebola virus disease (EVD0) outbreak in history, with a case fatality rate of 70–71 % and

B. Pell · J. Baez · T. Phan · Y. Kuang (✉)
Arizona State, Tempe, USA
e-mail: kuang@asu.edu

B. Pell
e-mail: bepell@asu.edu

D. Gao
Mathematics and Science College, Shanghai Normal University,
Shanghai 200234, China

G. Chowell
School of Public Health, Georgia State University,
Atlanta, GA 30302, USA

a hospitalized fatality rate of 57–59 % [7, 23, 24]. This epidemic is significantly different in both size and duration compared to previously reported EVD epidemics. As of August 30, 2015, over 28,000 cases have been reported, of which over 11,000 patients have succumbed to the disease, making it the deadliest Ebola epidemic in history [24]. This latest outbreak far surpasses the number of reported cases and deaths from ten major previous ebola outbreaks combined with an estimated 1,531 cases and 1,002 deaths [6].

Although, EVD was first discovered in 1976, the virus had not triggered a major regional epidemic until Dec. 2013. Standard practices to prevent the outbreak in these countries were not as effective partly due to their poor health infrastructure, including the lack of public health surveillance systems to rapidly detect emerging outbreaks [11]. In addition, no licensed vaccine against EVD was available during the 2013–2015 epidemic [1, 26]. Instead, quarantine, isolation and education programs were used to mitigate the spread of the disease.

Measuring the effect that control interventions have on epidemics can be achieved by measuring shifts in R_0 and $R_e(t)$, the basic and effective reproduction numbers, respectively. R_0 is defined as the average number of secondary infections generated by one infectious agent in a completely susceptible population. Nevertheless, R_0 assumes the epidemic first occurs in a fully susceptible population and thus does not account for time-dependent variations. $R_e(t)$ is defined as the actual average number of new infections by one infectious agent in a population with both infected and uninfected individuals at time t. $R_e(t)$ shows time-dependent variation due to the implementation of control strategies and the decline in susceptible individuals.

Several studies have used mathematical models to quantify the effect that control interventions and behavior changes have on managing the epidemic. In [3], Althaus, employs an SEIR (susceptible-exposed-infectious-removed, [4]) model and the estimated effective reproduction number to gain insights into the real-time intervention effects for the 2013–2015 EVD epidemic. They suggest that the effective reproduction numbers in Guinea and Sierra Leone decreased to around unity by the end of May and July 2014 due to sufficient control measures. However, that was not the case in Liberia where efforts needed to be improved. In a similar spirit, Chowell et al. [8], employed the logistic model to capture early signs of intervention and behavior changes in the population. Furthermore, they showed that phenomenological models are useful for understanding early epidemic dynamics, specifically because of the small number of parameters that need to be estimated. With more complexity, Agusto et al. [2], used a mathematical model to explore the effects of traditional belief systems and customs on the transmission process, concluding that the 2014 outbreaks may be controllable by using a moderately-effective basic public health intervention plan.

Other studies have used mathematical models to investigate the affects of spatial structure on disease dynamics. For instance, Valdez et al. [22], embeds a compartmental model into a 15-patch spatial framework (representing 15 counties of Liberia) and shows that reducing mobility only delays the overall control of the epidemic. Their findings suggest that safe burials and hospitalizations are key to controlling EVD. In particular, if safe burials and hospitalizations were established in mid-July 2014, their model predicts that the epidemic would have been three months shorter

and infected individuals would have been 80 % less than if the controls were implemented in mid-August. Gomes et al., employs the Global Epidemic and Mobility Model that incorporates mobility and demographic data at a worldwide scale coupled to a stochastic epidemic model [13]. They concluded that the probability of the disease spreading outside of Africa was highly unlikely. Merler et al., employs a spatial agent-based model to examine the effectiveness of safe burials, household protection kits and to estimate Ebola virus transmission parameters [18]. They suggest that the majority of infections occur within hospitals and households. Their findings indicate that the decline in disease incidence is due in part by the increased number of Ebola treatment units, safe burials and household protection kits. Using a discrete, stochastic SEIR model that is embedded within a three-scale community network model, Kiskowski, shows that effects from community mixing along with stochasticity can explain the different growth rates of reported cases observed in Sierra Leone, Liberia and Guinea [15].

Multiple studies have used mathematical models for forecasting the potential number of future cases and estimating transmission parameters for the 2013–2015 Ebola epidemic. Meltzer et al., constructs the EbolaResponse modeling tool that tracks patients through multiple stages of infection and categorizes patient infectiousness depending on whether they are in a hospital, a low-risk community setting or at home with no isolation [17]. The EbolaResponse model was used to estimate how control and prevention measures could stop the epidemic and to forecast future cases. Meltzer et al., suggest that policy makers rapidly increase the number of Ebola treatment units. In another study by Shaman et al., a stochastic compartmental model is coupled with the Ensemble Adjusted Kalman Filter (EAKF) to forecast state variables and parameters six weeks into the future [21]. The EAKF adjusts the parameters and ensemble state variables as more data becomes available. Parameter estimations provided some evidence that the epidemic growth was slowing down in Liberia.

We present a simple approach that phenomenologically connects the effects of behavior changes to mitigate transmission rates and population spatial structure. Our method derives the logistic equation from an assumption about the effect of population behavior and introduces spatial heterogeneity via logistic patch models. In particular, we contribute the following:

- The logistic model is derived from a susceptible-infected compartmental model in Sect. 2.1, justifying its use in [8].
- Formulas for the basic and effective reproduction numbers are presented in Sect. 2.2.
- We build upon the work done in [8], by incorporating spatial heterogeneity via logistic patch models.
- Models are validated by comparing their fits to total reported case data in Sect. 4.1.
- As seen in Fig. 4, we show that these models improve upon the short term forecasting error in Sect. 4.2. Furthermore we perform Kruskal–Wallis tests to analyze the variation across the different models.
- Further model validation and comparison is presented in Sect. 4.3, via parameter estimations and confidence intervals. This section shows that patch models are not well constrained due to limited data.

- We provide estimates and 95 % confidence intervals of R_0 for Liberia, Sierra Leone and Guinea respectively in Sect. 4.4.

2 Modeling Methods

2.1 Logistic Equation as an Ebola Cumulative Infections Case Model

From a basic SI compartmental model and an assumption about population behavior we can derive the logistic equation. Assuming there are no births, natural deaths or immigration of susceptible individuals and that infected individuals do not return to the susceptible class, the classical Kermack and McKendrick infectious disease model can be adapted to obtain the following:

$$
\begin{aligned}
S(t)' &= -\frac{\beta S(t) I(t)}{S(t) + I(t)}, \\
I(t)' &= \frac{\beta S(t) I(t)}{S(t) + I(t)} - \mu I(t),
\end{aligned}
\tag{1}
$$

where β is the infection rate and μ is the disease induced death rate. From system (1) the cumulative number of infections at time t, denoted by $x(t)$, has derivative $x'(t) = \beta \frac{SI}{S+I} \approx \beta I$, (assuming $\frac{S}{S+I} \approx 1$). Below we assume that $x'(t) = \beta I$.

As an increasing number of cases are reported during an outbreak, the behavior of the individuals in the affected region may change due to disease education programs, an increase in care or quarantine facilities and help from health care workers.

As an example, dead bodies infected with Ebola virus remain infectious, causing participants to unknowingly contract the infection during funeral burials. In the beginning stage of the outbreak, unsuspecting mourners would carry the infection back to other parts of the community and would infect more individuals. By having specific handling guidelines of human remains, communities were able to decrease exposure to the Ebola virus [27]. In general, this is the notion of a positive behavioral change in the community. Based on these observations we make what we call the behavior assumption:

- **(Behavior assumption)**: During an epidemic, a change in behavior in the community that mitigates the transmission rates is expected as an epidemic unfolds. This response is modeled by a function of the total reported cases and has a decreasing effect on per-capita infection rate. That is,

$$
\frac{I'(t)}{I(t)} = f(x(t))
\tag{2}
$$

is a decreasing function of the total number of reported cases $x(t)$.

In the following, we assume that $f(x(t)) = r(1 - ax(t))$ for some positive constants $r := \beta - \mu$ and a. Hence
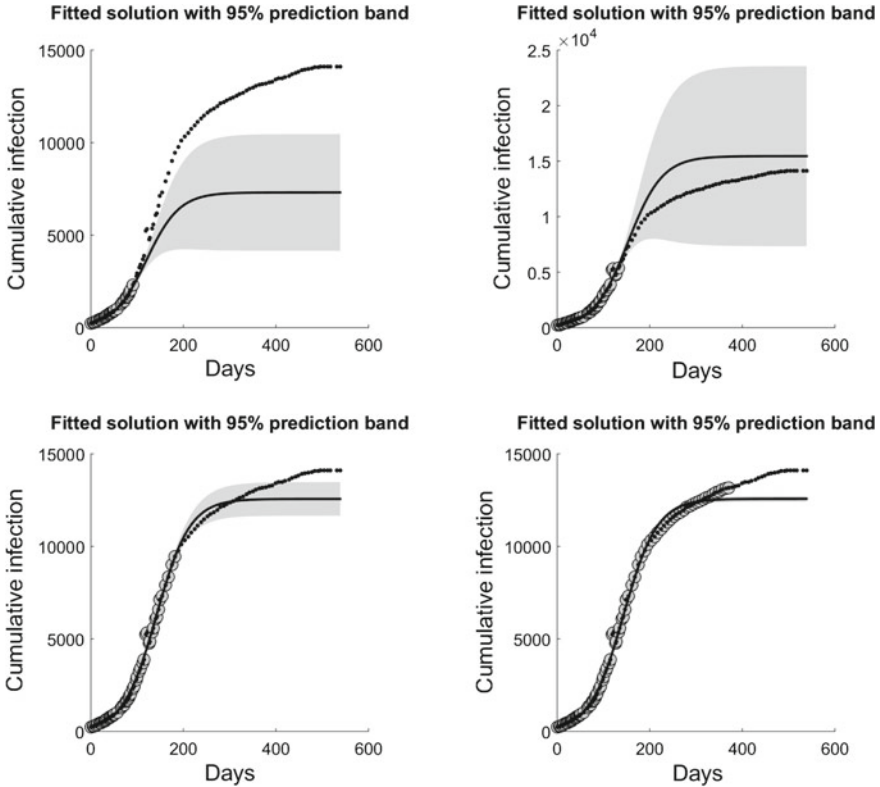
$$I'(t) = rI(t)(1 - ax(t)) = \frac{r}{\beta}x'(t)(1 - ax(t)).$$

Therefore,

$$I(t) - I(0) = \frac{r}{\beta}\left(x(t) - \frac{a}{2}[x(t)]^2\right) - \frac{r}{\beta}\left(x(0) - \frac{a}{2}[x(0)]^2\right).$$

Since $I(0) = x(0) \approx 0$, we see that $I(t)$ can be approximated by $\frac{r}{\beta}\left(x(t) - \frac{a}{2}[x(t)]^2\right)$. Therefore

$$x'(t) = \beta I(t) = r\left(x(t) - \frac{a}{2}[x(t)]^2\right) = rx(t)\left(1 - \frac{x(t)}{K}\right), \qquad (3)$$

Fig. 1 Predictions of the cumulative number of Ebola cases in Sierra Leone by the logistic growth Eq. (3). Data points start on June 2, 2014 and end December 23, 2015. 95 % prediction bands are superimposed. *Gray disks* are data points for model calibration, while *black dots* are forecasting data points

where $K = 2/a$. Here we interpret r as the *intrinsic infection rate*, a is a proportionality constant that corresponds to strength and effectiveness of disease interventions and preventive strategies and K is the *final epidemic size*.

In [8], the saturation effect of the logistic equation was used to implicitly account for the behavior change in the population. The above derivation provides a rigorous framework of this modeling effort and emphasizes the role behavior plays in the saturation effect. Figure 1 shows the change in the 95 % prediction band using the delta method as more data points are incorporated when fitting the logistic model to epidemic data [5].

2.2 Derivation of R_0 and R_e

During an outbreak, there may not be enough data to calibrate mechanistic models of the exact transmission processes, thus the logistic model can provide useful insights into the early outbreak dynamics. To derive R_0 and R_e first observe that,

$$I(t + T) = R_e(t) I(t), \tag{4}$$

where T is the mean generation interval and is defined as the time between infection in an index case patient and infection in a patient infected by that index case patient [23]. From Eq. (2), we have that $I'(t) = f(x(t)) I(t)$, integrating both sides from t to $t + T$ yields

$$\ln(I(t + T)) - \ln(I(t)) = \int_t^{t+T} f(x(s)) \, ds.$$

Solving for $I(t + T)$ and dividing by $I(t)$ yields $\frac{I(t+T)}{I(t)} = e^{\int_t^{t+T} f(x(s))ds}$, which from Eq. (4) yields

$$R_e(t) = e^{\int_t^{t+T} f(x(s))ds}. \tag{5}$$

Lastly, define $R_0 := e^{rT}$ which is approximately equal to the usual definition of the basic reproduction number, $\frac{\beta}{\mu}$, of model 1 when $\frac{\beta}{\mu}$ is close to 1.

2.3 Incorporating Population Heterogeneity: Multi-patch Models

District geography, topology, health care centers and quarantined regions can influence population movement. This motivates the need for incorporating spatial structure in transmission models. We do this by partitioning a district into a network of

two or more sub-districts (patches). In each sub-district, cumulative infections obey logistic growth individually.

Let x_i be the **cumulative infections** in patch i and let m_{ij} be the rate of cumulative infections that travel from patch i to patch j, where $i, j = 1, 2, i \neq j$.

The equations for the two-patch model are:

$$x_1' = r_1 x_1 \left(1 - \frac{x_1}{K_1} \right) - m_{12} x_1 + m_{21} x_2,$$

$$x_2' = r_2 x_2 \left(1 - \frac{x_2}{K_2} \right) - m_{21} x_2 + m_{12} x_1.$$

Similarly, the three-patch model is given by:

$$x_1' = r_1 x_1 \left(1 - \frac{x_1}{K_1} \right) - (m_{12} + m_{13}) x_1 + m_{21} x_2 + m_{31} x_3,$$

$$x_2' = r_2 x_2 \left(1 - \frac{x_2}{K_2} \right) - (m_{21} + m_{23}) x_2 + m_{12} x_1 + m_{32} x_3,$$

$$x_3' = r_3 x_3 \left(1 - \frac{x_3}{K_3} \right) - (m_{31} + m_{32}) x_3 + m_{13} x_1 + m_{23} x_2.$$

In addition, we will consider two special cases of each model: symmetric migration (S) with $m_{ij} = m_{ji}$ and homogeneous migration (H) with, $m_{ij} = m$ for all i, j and $i \neq j$.

Assume that r_i and K_i are positive in the above models. It is easy to see that these patch models are cooperative in nature which generate a strictly monotone semiflow. It is shown that the positive solutions of the above models tend to a unique positive steady state (see Lemma 3.1 in [12]).

Let $x = \sum_{i=1}^{N} x_i$. As with the derivation of R_e and R_0 for the logistic model above, define the basic reproduction number for an N-patch model as

$$R_e(t) = \exp \left(\hat{r} \int_t^{t+T} 1 - \frac{2}{\hat{K}} x(s) \, ds \right),$$

where $\hat{r} = \frac{\sum_{i=1}^{N} r_i K_i}{\hat{K}}$, $\hat{K} = \sum_{i=1}^{N} K_i$ are weighted averages and for simplicity we assume $T = 2$ weeks, instead of 2.18 [1]. Similarly to above, we define $R_0 := R_e(0) \approx e^{rT}$.

3   Comparison Methods

We use district data from the World Health Organization (WHO) patient database, which contains weekly reported confirmed, suspected and probable infections from Liberia, Sierra Leone and Guinea [24]. Data ranges from Mar. 1, 2014 to Aug. 5, 2015.

Table 1 Number of parameters for each model

	Logistic	Two-patch (H)	Two-patch	Three-patch (H)	Three-patch (S)	Three-patch
Number of parameters	2	5	6	7	9	12

Table 1 lists the number of parameters of each model. By studying the special cases of the patch models we reduce the number of parameters that need to be estimated, which constrains model fits and reduces the likelihood of over-fitting the data.

We use Matlab's built-in function, fminsearch, to help locate optimized parameter values for data fitting. fminsearch is a derivative-free method that is based on the Nelder-Mead Simplex [16] and searches for minimums, but does not guarantee global minimums. We are searching for a biologically reasonable parameter set that minimizes the error between the simulations and the observed data. To this end, we define the weighted error function:

$$E_w = \frac{1}{N - P} \sum_{i=1}^{N} |y_i - \hat{y}_i| \, e^{-0.1(t_f - t_i)}, \tag{6}$$

where t_f is the final date that we have an observation for, P is the number of parameters and N is the number of observations. $\hat{y}_i$ denotes the observation at time t_i and y_i the value of our model at the ith observation. We make the assumption that recent data has higher significance for forecasting future cases, as reflected by the exponential factor. The value of 0.1 in the exponential term is used because it gave a reasonable temporal-weight to the data points.

3.1 Ranking Models by Fitting and Forecasting Errors

To compare the models, we use absolute and relative errors that penalize models that have more parameters. The absolute error is calculated using the following equation,

$$E_{abs} = \frac{1}{\sqrt{N - P}} \sqrt{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{7}$$

and the relative error is given by,

$$E_{rel} = \frac{1}{\sqrt{N - P}} \sqrt{\sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{\hat{y}_i}\right)^2}. \tag{8}$$

Since we are interested in assessing and ranking the forecasting performance of all models, we define the *forecasting error* as follows:

$$E_{fcst} = \frac{1}{\sqrt{N - \hat{N} - P}} \sqrt{\sum_{i=i^*}^{N} \left[y(t_i) - \hat{y}(t_i) \right]^2},\tag{9}$$

where i^* corresponds to the temporal index at which we start forecasting our models, N is the total number of observations and $\hat{N}$ is the total number of observations used for model calibration and P is the number of parameters. If i^* was not an integer value, we took its floor value.

3.2 Parameters and Confidence Interval Assessment

To further compare and assess the models we compute 95 % confidence intervals for the logistic, two-patch (H) and three-patch (H) models. Only these models were considered, because they have the least number of parameters which reduces the likelihood of overfitting the models to data. Bootstrapping can be used as a way to estimate standard errors of parameter estimates in statistical models. The basic idea is to fit the model to data, find the residuals and add them to the data. Next, randomly sample with replacement B times, where B is large and fit the model to each of these newly created data sets to obtain B different parameter sets from the fitted model. This allows one to obtain a distribution of the parameters without assuming anything prior about them. For further details see [9, 10, 20].

Recall a statistical model, with $y = (y_1, \ldots, y_n)$ being explained by k explanatory variables $x = (x_1, \ldots, x_k)$ using p parameters $\theta = (\theta_1, \ldots, \theta_p)$:

$$y_i = g(x_i|\theta) + \varepsilon_i$$

for $i = 1, \ldots, n$. Where g is a mathematical model such as an ordinary differential equation model, partial differential equation model, algebraic model, etc. ε is the error and is a random variable and y is another random variable. Let G be the partial derivative matrix with respect to θ and the leverages, $h_1, \ldots, h_n$ be the diagonal elements of the $G(G^{\dagger}G)^{-1}G^{\dagger}$ matrix, where $\dagger$ denotes matrix transpose.

The bootstrapping method is described below.

1. Fit the model to the original data with an initial parameter set, $\hat{\theta}$, and for each x_i, compute the corresponding residual $\hat{\varepsilon}_i = y_i - \hat{y}_i$ for $i = 1, 2, \ldots n$, where n is the total number of data points and $\hat{y}_i = g(x_i, \hat{\theta})$.
2. Correct for the potential heteroscedasticity in the residual variances by computing the **modified residuals**: $\hat{r}_i = \frac{\hat{\varepsilon}_i}{\sqrt{1-h_i}}$ and compute the centered residuals $r_i^* = \hat{\varepsilon}_i - \hat{r}_i$, where h_i are the leverages.
3. Sample with replacement from the n modified and centered residuals.
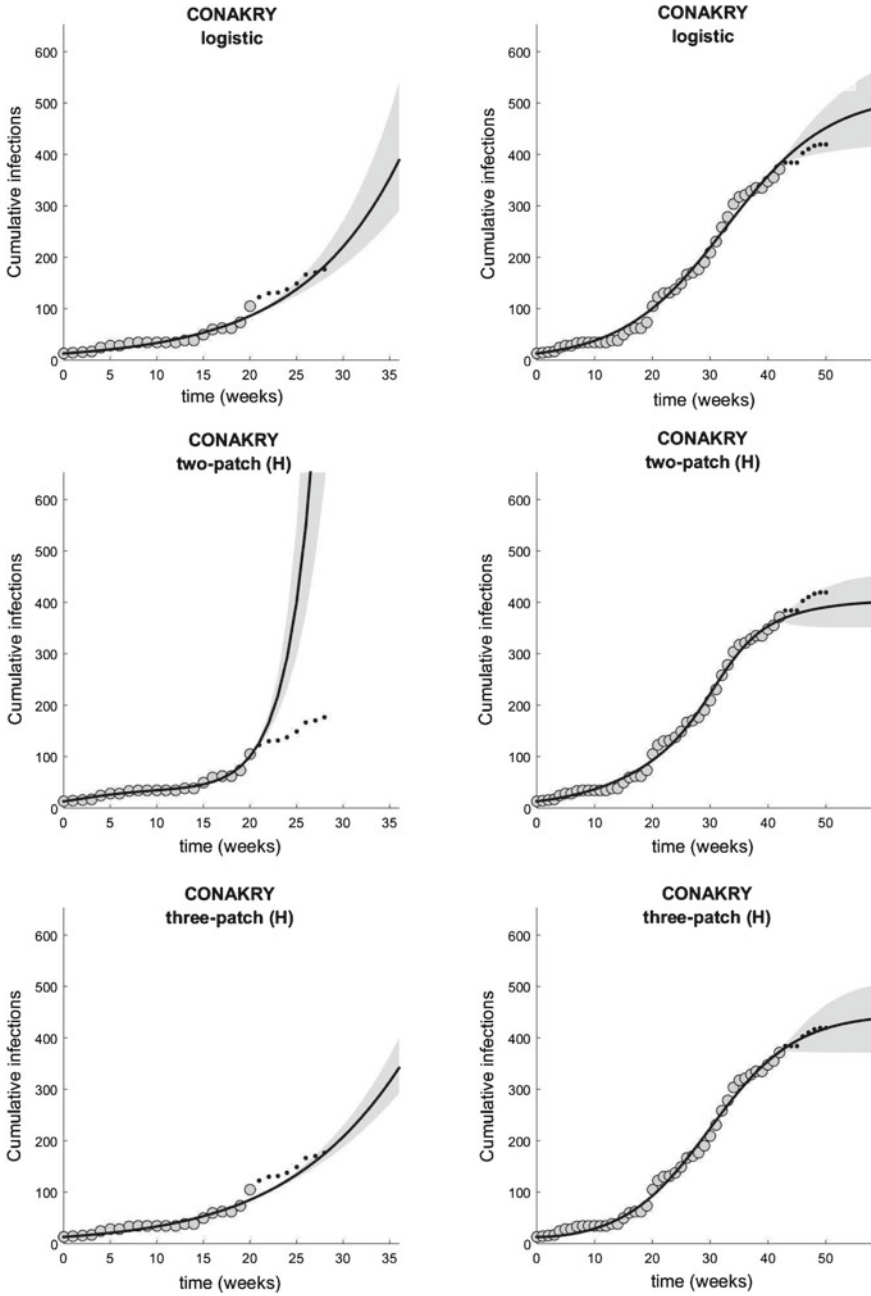
Fig. 2 Illustration of the model fitting and forecasting for Conakry, the capital and largest city of Guinea. *Left column* models trained on the first one-third data. *Right column* models trained on the first two-thirds of data. *Gray shaded region* represents 95 % prediction bands

4. Generate bootstrap sample, $\bar{y}_i := \hat{y}_i + r_j^*$, for all i and where j is random.
5. Fit the model to these new $\bar{y}_i$ values and obtain a new set of parameter values, $\bar{\boldsymbol{\theta}}$.
6. Repeat steps 3, 4 and 5 a large number of times[1] of times (say 2,000). This generates 2,000 bootstrap samples and corresponding sets of parameter estimations.
7. Use the 2,000 parameter estimates to generate distributions to find confidence intervals.

3.3 Challenges

Whenever fitting a mathematical model to time series data that ranges from small values to very large values, deciding at what time to initiate the model can seriously influence its forecasting ability. For example, training a model on a large set of data that is relatively near zero except for the last couple of points will force the fitting to be heavily biased by the large amount of initial points near zero, thus not providing a good forecast. We remedied this by starting the models after there were no three consecutive weeks that had no infections and by using the weighted error (Eq. 6) for fitting. This was done due to the fact that smaller outbreak waves happened before the main wave of infections appeared.

Forecasting an ongoing disease outbreak in real-time brings many challenges. New data being available means that computer programs must be designed to process and incorporate new data sets with ease and in a timely fashion. In our case, fitting six models (including special cases) to forty-one data sets requires a significant amount of computing resources.

4 Results

4.1 Data Based Model Validation

To validate the patch models for epidemic modeling, we fit all models to all data sets and compare model fits and errors. To illustrate this fitting process, Fig. 2 shows model fits of the logistic, two-patch and three-patch models with homogenous migration to cumulative reported case data from Conakry, Guinea.

We report the means for the weighted, relative and absolute error (respectively Eqs. 6, 7 and 8) for all 39 data sets in Table 2. Observe that the patch models show an improvement over the logistic model when fitting the data. Additionally, we see that the homogeneous migration models perform better than their free migration versions.

---

[1]Results from Efron and Tibshirani [10] suggest that accurate results for confidence intervals can be obtained from 1000 bootstrap samples. For standard errors this number is reduced to 200.

Table 2 Mean error statistics

Model	Weighted error	Relative error	Absolute
Logistic	82.2822	1.3387	102.198
2-Patch (H)	53.6764	1.1271	63.1193
2-Patch	58.6311	1.2124	72.7197
3-Patch (H)	48.709	1.1256	59.3391
3-Patch (S)	55.0951	1.1515	65.1694
3-Patch	54.215	1.1727	66.0885

In what follows, we summarize the different fitting and forecasting cases. Let FTG be the fitting error from Eq. 6 and FCST be the forecasting error from Eq. 9. We use the following convention to denote the different errors: FTG-Δ and FCST-Δ-Ω, where Δ is the fraction of data used for fitting and Ω is the number of weeks forecasted ahead (Table 5).

Fitting errors were calculated using Eq. 6 and the first one-third and the first two-thirds of each data set. All fitting errors are provided in Table 5 given in the appendix. From Fig. 3, most of the patch models had smaller mean fitting error than the logistic model.

Four and eight week forecasts were made after training all models to the first one-third and first two-thirds of the data set. Figure 3 shows that in all cases, the patch models had smaller mean forecasting errors. This supports the hypothesis that modeling spatial structure within the district improved forecasting error. Additionally, all models perform better when forecasting the short-term rather than long-term epidemic trajectory. Forecasting error variance was lowest with FCST-2/3-4. In contrast, the variance was the largest with FCST-1/3-8.
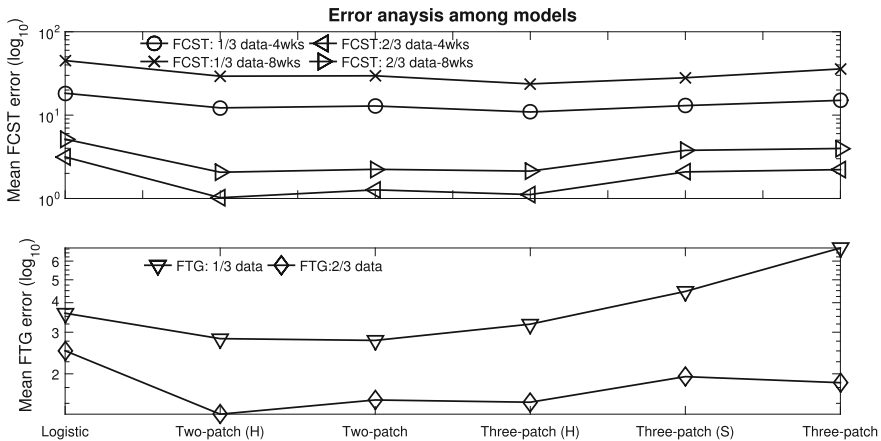
Fig. 3 Mean forecasting and fitting errors. Models are along the x-axis and variance is along the y-axis. We connect points for aesthetic purposes

Table 3 p-values of the Kruskal–Wallis test show forecasting errors do not significantly differ across models

Case	p-value
FCST-3-4	0.9806
FCST-3-8	0.9872
FCST-23-4	0.9933
FCST-23-8	0.9894

Results of Kruskal-Wallis tests were not significant for FCST-1/3-4, FCST-1/3-8, FCST-2/3-4 and FCST-2/3-8; the mean ranks for all forecasting cases did not significantly differ. We include the p-values (95 %) in Table 3 for this test.

4.2 Forecasting Error as a Function of Forecasting Points

Forecasting error for Port Loko, Guinea, Liberia and Sierra Leone was calculated for varying amounts of forecasting points.
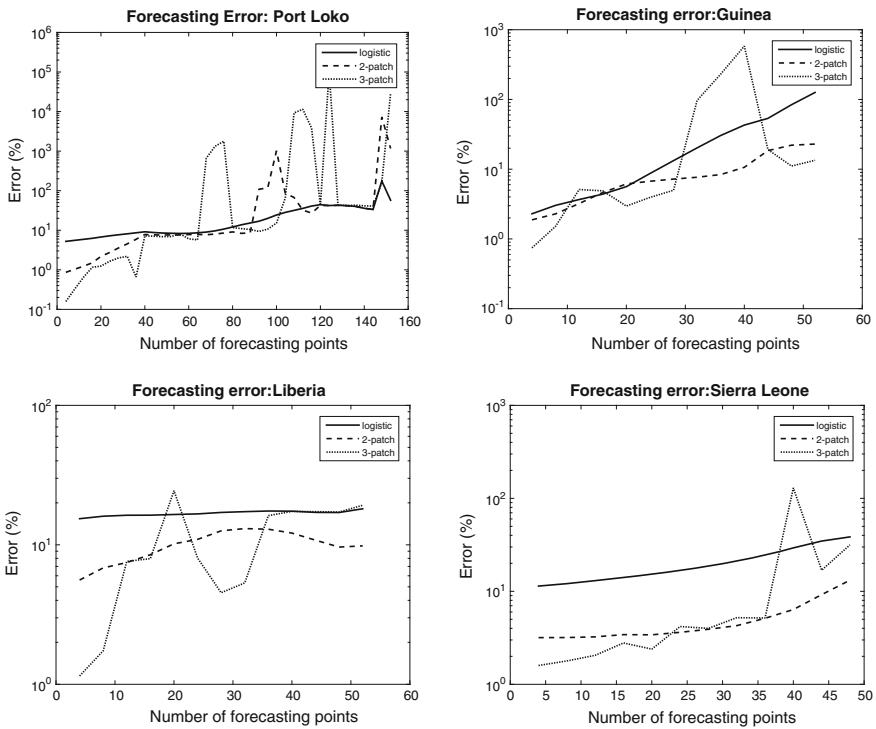

Fig. 4 Relative error of forecasting points for the logistic, two and three patch models with homogeneous migration rate: Port Loko, Guinea, Liberia and Sierra Leone

The forecasting error for Port Loko in Fig. 4, suggests that the patch models have smaller forecasting errors than the logistic equation for short-term forecasts (4–70 days). Additionally, it shows erratic long-term forecasting of the patch models for Port Loko, because they are not well constrained due to the limited data. Figure 4 further shows lower short-term error for Sierra Leone and Liberia by the two-patch model. We note that the three-patch model yielded the smallest error when forecasting ten prediction points or less (4–10 days).

4.3 Confidence Interval Assessment

Parameter confidence intervals for the logistic equation decrease in length as we decrease the number of prediction points (Fig. 5) for Port Loko. Similar assessments were done using data from Sierra Leone, Liberia and Guinea at the country level. Results were similar as the Port Loko case except for Liberia, where confidence interval lengths begin to increase when we forecast less data points. In summary, the logistic model shows well behaved parameter values when we fit to an increasing number of data points for three out of four data sets used.

The patch models tell a different story. The confidence intervals are larger and show erratic behavior when forecasting a large number of points. Indeed for the two-patch model (Fig. 6), the confidence intervals for r_1 actually increase when we are predicting a small number of data points from Port Loko. This variability is seen to be worse in the confidence intervals for the final epidemic sizes (K_i's) for both two and three patch models, but they are so erratic that they cannot be shown in a reasonable way and therefore are not included. The fact that the patch models have more parameters allows for different parameter sets that produce a well fit curve, but allow for large variability in the parameter sets. The same is seen in the confidence
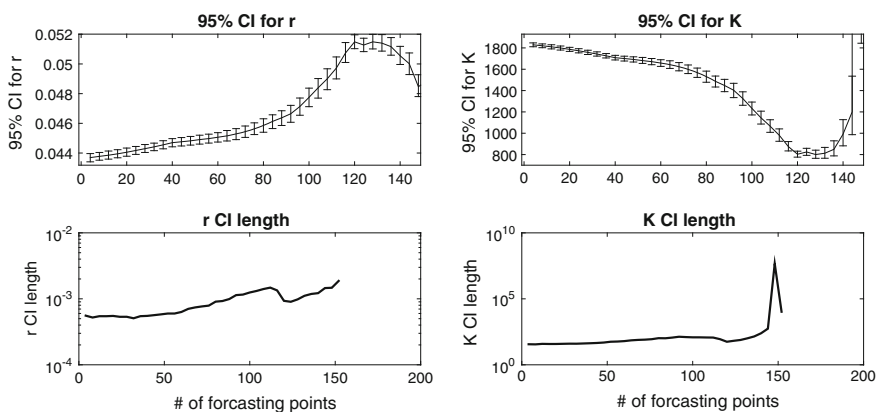
Fig. 5 95 % CI for r and K from (3). (*Bottom*) Plot of the length of the CI for r and K as a function of the number of forecasting points. District: Port Loko
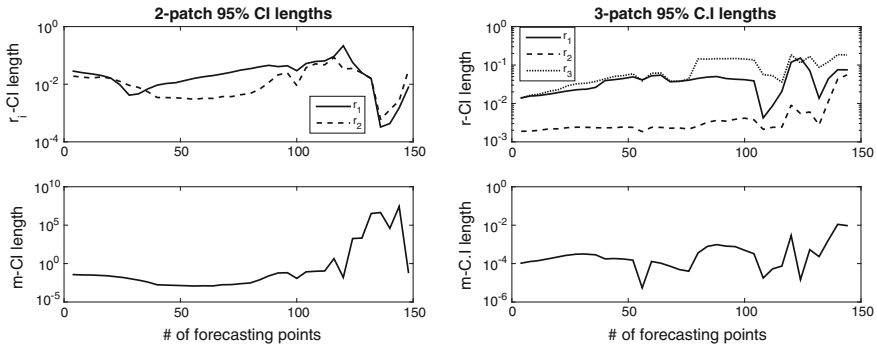
Fig. 6 95 % CI for r_i and m for $i = 1, 2, 3$ from (6) for Port Loko. (*Left*) two-patch confidence interval lengths for intrinsic infection rate and migration parameter. (*Right*) Three-patch 95 % confidence interval lengths for intrinsic infection rate and migration rate. Note the variability for high numbers of prediction points for both models and the high variability in r_1 for the two-patch model for low numbers of prediction points

interval assessment using data from Sierra Leone, Guinea and Liberia, (not shown here).

4.4 Implications for Liberia, Sierra Leone and Guinea: R_0

From the bootstrapping method, we calculated 95 % confidence intervals for R_0 in Guinea, Liberia and Sierra Leone (see Table 4).

5 Discussion

In this chapter, a family of logistic patch models were preliminarily evaluated for use in disease modeling and forecasting. An explicit formula for the cumulative number of infectious individuals was derived from a SI compartmental model which takes the form of the well known logistic model. This derivation follows from the behavior change assumption, Eq. (2). We then extended the logistic model to include spatial population heterogeneity by using multi-patch models that incorporate migration between patches and logistic growth within each patch. Each model's ability to forecast epidemic data was assessed by comparing model forecasting error, parameter distributions and parameter confidence intervals as functions of the number of data points used to calibrate the models. The patch models show an improvement over the logistic model in short-term forecasting, but naturally require the estimation of more parameters from limited data.

The models were tested by fitting them to the total reported case data from 39 districts in West Africa. In particular, the means of the weighted, relative and absolute

errors of the patch models are less than the logistic model's, suggesting that spatial structure improved the data fitting. Next, models were compared by their forecasting capabilities in two ways: comparing forecasting error and comparing parameter confidence intervals. These latter efforts were restricted to the logistic, two-patch and three-patch models with homogeneous migration. The forecasting errors from Fig. 3 show that the patch models forecast better than the logistic model. However, Fig. 4 shows long-term forecasting variability from the patch models, because of the limited data. In contrast to these results, the Kruskal-Wallis test showed no significant difference in the forecasting errors across the models.

The value of R_0 during the outbreak in Liberia, Guinea and Sierra Leone were estimated to be in the same range as previous studies that were based on compartmental models [3, 13, 14, 28]. In particular, from Table 4 the estimates from the two and three patch models for R_0 are similar with Althaus et al., but our confidence intervals are not as small [3]. This agreement further supports the reliability of the logistic and patch models with homogeneous migration.

In reality, early in the Ebola 2013–2015 epidemic, the public's behavior in Liberia, Sierra Leone and Guinea did not swiftly change in a manner that mediated disease transmission nor has there been any evidence supporting that the per-capita infection rate decreased linearly. Actually, the public's misunderstanding of the disease, lack of resources and fear fostered high-risk behaviors and resulted in an increased disease transmission in West Africa during the epidemic [19, 25]. However, health-care workers supplied valuable public awareness programs and medical resources that helped manage the spread. Our modeling assumptions approximate these notions and provide immediate behavior change in the spirit of Eq. (2), but this is modeled simultaneously everywhere in space and is one reason why the logistic model does not fit the data well. The patch-models overcome this issue by modeling behavior changes at different times, rates and locations, but require more data to be constrained. Indeed, an issue with the patch models is that the number of parameters increase quickly as more patches are introduced.

Table 4 R_0 and 95 % confidence intervals for R_0

	Althaus [3]	Team et al. [23]	Logistic	2-Patch (H)	3-Patch (H)
Guinea	1.51 (1.50–1.52)	1.71 (1.44–2.01)	1.252 (1.249, 1.255)	1.52 (1.42, 1.92)	1.45 (1.39, 1.51)
Liberia	1.59 (1.57–1.60)	1.83 (1.72–1.94)	2.11 (2.07, 2.15)	1.45 (1.12, 1.94)	1.43 (1.06, 2.199)
Sierra Leone	2.53 (2.41–2.67)	2.02 (1.79–2.26)	2.28 (2.25, 2.32)	2.27 (2, 2.62)	2.12 (1.87, 2.26)

Further work can be done with between-country and between-district scales. The latter would allow for more parameter constraint, but would have to be restricted to a small number of patches that represent a small number of neighboring districts. The problem with incorporating all districts is that it ultimately requires a high-dimensional patch model with many parameters on a complicated network. This may be remedied with a partial differential equation model or by using mobility data to constrain the migration parameters. In addition, exploring different behavior functions would be another direction to expand this work.

Although the logistic model is phenomenological, it is capable of fitting the sigmoid curves that usually result from plotting the cumulative reported cases of disease outbreaks. The logistic and the patch models provide a general framework for disease modeling, because they do not model specific disease transmission processes. Specifically, they are based on two fundamental mechanisms that influence disease outbreaks: behavior change in the community and movement of individuals within that community. We find that incorporating the latter mechanism decreased forecasting errors with respect to the logistic model, but also require more data for model calibration.

Appendix

Forecast and Fitting Error Tables

See Table 5.

Table 5 Fitting errors for all models

District	Logistic		Two-patch (H)		Two-patch		Three-patch (H)		Three-patch (S)		Three-patch	
	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds
BOMI	3.6147	1.198	1.7439	0.39968	2.23	0.49683	1.9891	0.4395	3.1711	0.52022	3.9724	0.47549
BONG	1.4365	0.54448	1.5116	0.42227	1.6447	0.46937	1.6753	0.5351	2.1265	0.49347	2.9081	0.58443
GBARPOLU	0.71881	0.078303	0.6592	0.087046	1.001	0.0887	1.1162	0.095164	1.1436	0.077254	2.5221	0.11758
GRAND BASSA	1.8093	0.68238	1.1774	0.29884	1.8118	0.41695	1.7176	0.38735	2.2839	0.44523	3.2918	0.58292
GRANDAPE MOUNT	3.333	0.5762	3.7319	0.63049	3.8035	0.65232	3.3711	0.67542	5.0092	0.73396	8.0373	0.83729
GRAND KRU	0.61518	0.17033	0.80231	0.099753	0.84577	0.17673	0.64961	0.12372	1.0854	0.1753	2.3581	0.2528
LOFA	5.2755	0.36486	5.6001	0.26833	6.9035	0.36263	7.9303	0.34862	6.4173	0.27562	11.1918	0.47298
MARGIBI	3.4293	0.73833	4.4185	0.67457	4.3363	0.72399	4.5471	0.72418	5.619	0.79608	7.7291	0.89117
MONTSERRADO	4.2503	8.3487	4.922	2.2296	5.2784	1.8897	5.6934	2.3507	6.7871	1.5	9.1043	2.424
NIMBA	1.2104	0.12813	1.351	0.13701	1.4561	0.14135	1.5792	0.1464	1.9673	0.15578	2.7993	0.17449
RIVER GEE	0.15977	0.01381	0.1785	0.015599	0.17445	0.015513	0.18504	0.014329	0.23398	0.019005	0.3597	0.019446
RIVERCESS	0.68386	0.077873	0.51155	0.093633	0.54034	0.027355	1.1207	0.1403	1.6564	0.10546	3.84	0.20521
SINOE	0.71361	0.074711	0.76264	0.069135	0.92105	0.089601	1.1015	0.095067	1.5029	0.099703	2.8564	0.13226
CONAKRY	2.4851	2.2749	2.4127	2.428	3.1721	2.443	3.4615	2.5474	4.0175	2.6153	7.2811	2.9771
COYAH	2.9429	3.1608	1.6338	3.4047	2.0082	2.9153	3.1834	3.1261	5.2135	3.2436	5.4088	4.6479
DUBREKA	1.8975	1.0283	0.84368	1.0297	2.5879	0.70316	5.2154	1.0243	3.0884	1.0003	5.7692	1.1607
FARANAH	2.1281	0.37701	2.1799	0.42643	2.4097	0.41307	3.8266	0.50515	5.0919	0.53024	25.1464	0.24839
FORECARIAH	1.6364	3.1111	2.0821	2.8257	2.2314	3.3816	1.6743	13.7987	2.9266	3.4089	5.3969	3.8571
KANKAN	1.1965	0.14143	2.3918	0.1566	1.4882	0.1561	3.15	0.49163	3.0055	0.16257	3.7905	0.23459
KINDIA	2.0281	1.136	1.1496	0.91487	1.6464	1.071	1.5116	0.38334	2.5737	1.0592	5.3894	1.2542
KISSIDOUGOU	2.9894	0.40939	2.3094	0.47537	3.0904	0.53037	2.3511	1.1683	5.3364	0.39528	21.0685	0.77644

(continued)

Table 5 (continued)

District	Logistic		Two-patch (H)		Two-patch		Three-patch (H)		Three-patch (S)		Three-patch	
	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds	One-third	Two-thirds
MACENTA	0.4115	5.2217	1.6022	5.5868	1.8795	5.3856	2.0131	0.62805	4.0898	5.9809	6.7702	6.1607
NZEREKORE	2.5579	0.38891	0.44257	0.40024	0.42032	0.41519	1.0254	5.7162	1.3283	0.49414	0.68875	0.53331
SIGUIRI	0.44141	0.30196	3.1856	0.32252	3.5061	0.25171	3.8516	0.4314	5.0128	0.35819	8.8683	0.41559
TELIMELE	0.10555	0.55761	0.42019	0.593	0.4517	0.62469	0.65173	0.34477	0.75666	0.6742	1.8585	0.72794
BO	5.0701	2.2253	0.12915	2.4011	0.13449	2.0631	0.175	0.61298	0.15866	2.4455	0.24843	2.6633
BOMBALI	8.8538	9.0596	1.9431	3.2113	1.7743	4.476	4.0898	2.2312	5.8781	3.8964	6.1361	4.9203
KAILAHUN	13.4441	2.3216	3.0016	0.67132	3.8811	0.66044	4.9654	3.3201	7.8443	2.2252	8.5021	1.2032
KAMBIA	3.0924	2.8205	5.1435	1.6991	3.718	1.1592	7.6597	1.0176	11.6425	1.4946	13.382	1.5634
KENEMA	10.8483	2.4777	1.3423	1.2226	2.1451	0.83164	3.4772	1.4973	4.8896	0.85422	4.7483	1.0399
KOINADUGU	2.18	0.38465	8.1016	0.25254	4.9844	0.44878	6.3207	1.2681	15.0373	0.35363	6.2745	0.40977
KONO	5.3272	3.2919	2.7568	3.3897	3.1007	3.3525	1.7213	0.4554	4.7125	3.2659	7.5391	3.6874
MOYAMBA	2.18	0.38465	6.7662	0.25254	6.8282	0.44878	7.4665	3.6681	9.9778	0.35363	15.8279	0.40977
PORT LOKO	15.4465	15.1163	2.7568	2.2217	3.1007	6.4625	1.7213	0.4554	4.7125	14.2096	7.5391	2.7181
PUJEHUN	0.22985	0.02164	16.4268	0.017028	11.313	0.017084	13.17	3.6144	12.6916	0.017823	17.257	0.048621
TONKOLILI	11.6003	3.3944	0.18359	1.598	0.29053	1.9964	0.2574	0.016134	0.42985	4.0286	0.62986	1.8315
PUJEHUN	0.22985	0.02164	7.3516	0.017028	4.5375	0.017084	4.5832	1.6187	6.898	0.017823	10.2676	0.048621
WESTERN AREA RURAL	8.5967	11.4656	0.18359	3.0142	0.29053	5.9014	0.2574	0.016134	0.42985	7.0712	0.62986	4.0112
WESTERN AREA URBAN	5.639	13.5544	5.7848	8.6758	6.0817	8.771	6.0987	3.1066	7.3591	10.3357	8.545	16.817

Models were trained on one-third and two-thirds of each district data set

References

1. Agnandji, S.T., Huttner, A., Zinser, M.E., Njuguna, P., Dahlke, C., Fernandes, J.F., Yerly, S., Dayer, J.A., Kraehling, V., Kasonta, R., et al.: Phase 1 trials of rVSV Ebola vaccine in Africa and Europe–preliminary report. N. Engl. J. Med. (2015)
2. Agusto, F.B., Teboh-Ewungkem, M.I., Gumel, A.B.: Mathematical assessment of the effect of traditional beliefs and customs on the transmission dynamics of the 2014 Ebola outbreaks. BMC Med. **13**(1), 96 (2015)
3. Althaus, C.L.: Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. PLoS currents **6** (2014)
4. Anderson, R.M., May, R.M.: Infectious Diseases of Humans, vol. 1. Oxford University Press, Oxford (1991)
5. Bickel, P.J., Doksum, K.A.: Mathematical Statistics: Basic Ideas and Selected Topics, Volume I, vol. 117. CRC Press, Boca Raton (2015)
6. Center for Disease Control: Outbreaks Chronology: Ebola Virus Disease. Website (2015). http://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html
7. Chowell, G., Nishiura, H.: Transmission dynamics and control of Ebola virus disease (EVD): a review. BMC Med. **12**(1), 196 (2014)
8. Chowell, G., Simonsen, L., Viboud, C., Kuang, Y.: Is West Africa approaching a catastrophic phase or is the 2014 Ebola epidemic slowing down? Different models yield different answers for Liberia. PLoS currents **6** (2014)
9. Davison, A.C., Hinkley, D.V.: Bootstrap Methods and Their Application, vol. 1. Cambridge university press, Cambridge (1997)
10. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. CRC press, Boca Raton (1994)
11. Frieden, T.R., Damon, I., Bell, B.P., Kenyon, T., Nichol, S.: Ebola 2014 – new challenges, new global response and responsibility. N. Engl. J. Med. **371**(13), 1177–1180 (2014)
12. Gao, D., Ruan, S.: A multipatch malaria model with logistic growth populations. SIAM J. Appl. Math. **72**(3), 819–841 (2012)
13. Gomes, M.F., y Piontti, A.P., Rossi, L., Chao, D., Longini, I., Halloran, M.E., Vespignani, A.: Assessing the international spreading risk associated with the 2014 West African Ebola outbreak. PLOS Curr. Outbreaks **1** (2014)
14. Khan, A., Naveed, M., Dur-e Ahmad, M., Imran, M.: Estimating the basic reproductive ratio for the Ebola outbreak in Liberia and Sierra Leone. Infect. Dis. Poverty **4**(1), 13 (2015)
15. Kiskowski, M.A.: A three-scale network model for the early growth dynamics of 2014 West Africa Ebola epidemic. PLoS Curr. **6** (2014)
16. Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E.: Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J. Optim. **9**(1), 112–147 (1998)
17. Meltzer, M.I., Atkins, C.Y., Santibanez, S., Knust, B., Petersen, B.W., Ervin, E.D., Nichol, S.T., Damon, I.K., Washington, M.L.: Estimating the future number of cases in the Ebola epidemic–Liberia and Sierra Leone, 2014–2015. MMWR Surveill Summ **63**(suppl 3), 1–14 (2014)
18. Merler, S., Ajelli, M., Fumanelli, L., Gomes, M.F., y Piontti, A.P., Rossi, L., Chao, D.L., Longini, I.M., Halloran, M.E., Vespignani, A.: 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 (2015)
19. Nielsen, C.F., Kidd, S., Sillah, A., Davis, E., Mermin, J., Kilmarx, P.H.: Improving burial practices and cemetery management during an ebola virus disease epidemic-Sierra Leone, 2014. MMWR Surveill Summ **64**, 1–8 (2015)
20. Pardoe, I., Weisberg, S.: An Introduction to bootstrap methods using Arc. Unpublished Report available at www.stat.umn.edu/arc/bootmethREV.pdf (2001)
21. Shaman, J., Yang, W., Kandula, S.: Inference and forecast of the current West African Ebola outbreak in Guinea, Sierra Leone and Liberia. PLoS Curr. **6** (2014)
22. Valdez, L., Rêgo, H.H.A., Stanley, H., Braunstein, L.: Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies. Scientific Reports **5**, Article no. 12172

23. WHO Ebola virus disease in West Africa–the first 9 months of the epidemic and forward projections. N. Engl. J. Med. **371**(16), 1481–1495 (2014)
24. World Health Organization: Ebola Response Roadmap Situation report 03-05-2015 (2015). http://www.who.int/csr/disease/ebola/situation-reports/en/
25. World Health Organization: Ebola response: What needs to happen in 2015 (2015). http://www.who.int/csr/disease/ebola/one-year-report/response-in-2015/en/
26. World Health Organization: Ebola vaccines, therapies, and diagnostics (2015). http://www.who.int/medicines/emp_ebola_q_as/en
27. World Health Organization: Guidance for Safe Handling of Human Remains of Ebola Patients in U.S. Hospitals and Mortuaries (2015). http://www.cdc.gov/vhf/ebola/healthcare-us/hospitals/handling-human-remains.html
28. Yamin, D., Gertler, S., Ndeffo-Mbah, M.L., Skrip, L.A., Fallah, M., Nyenswah, T.G., Altice, F.L., Galvani, A.P.: Effect of Ebola progression on transmission and control in Liberia. Ann. Intern. Med. **162**(1), 11–17 (2015)