Mathematical Biosciences 301 (2018) 83–92

Contents lists available at ScienceDirect

Mathematical Biosciences

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

# Simple multi-scale modeling of the transmission dynamics of the 1905 plague epidemic in Bombay

Bruce Pell[1,a], Tin Phan[*,1,b], Erica M. Rutter[c], Gerardo Chowell[d], Yang Kuang[b]

[a] Department of Mathematics, Statistics, and Computer Science, St. Olaf College, Minnesota, USA
[b] School of Mathematical and Statistical Sciences, Arizona State University, Arizona, USA
[c] Department of Mathematics, North Carolina State University, North Carolina, USA
[d] School of Public Health, Georgia State University, Georgia, USA

ARTICLE INFO

*Keywords:*
Multi-scale modeling
Sub-exponential growth
Plague epidemic
1905 Bombay plague
Mathematical modeling
Disease transmission dynamics

ABSTRACT

The first few disease generations of an infectious disease outbreak is the most critical phase to implement control interventions. The lack of accurate data and information during the early transmission phase hinders the application of complex compartmental models to make predictions and forecasts about important epidemic quantities. Thus, simpler models are often times better tools to understand the early dynamics of an outbreak particularly in the context of limited data. In this paper we mechanistically derive and fit a family of logistic models to spatial-temporal data of the 1905 plague epidemic in Bombay, India. We systematically compare parameter estimates, reproduction numbers, model fit, and short-term forecasts across models at different spatial resolutions. At the same time, we also assess the presence of sub-exponential growth dynamics at different spatial scales and investigate the role of spatial structure and data resolution (district level data and city level data) using simple structured models. Our results for the 1905 plague epidemic in Bombay indicates that it is possible for the growth of an epidemic in the early phase to be sub-exponential at sub-city level, while maintaining near exponential growth at an aggregated city level. We also show that the rate of movement between districts can have a significant effect on the final epidemic size.

## 1. Introduction

Compartmental dynamic models have become a standard tool to investigate mechanisms of infectious disease transmission and control [1]. This approach is particularly successful in improving our understanding of the spread of infectious diseases when sufficient outbreak data or key epidemiological parameters such as transmission rates are available. However, in reality each new outbreak emerges with factors that may differ from previously well-studied outbreaks, e.g. geographical, socio-economical, strain variation. Thus, we can argue that during the early stage of a new outbreak, when data is scarce and is subject to significant uncertainty, it is often difficult to statistically connect complex mathematical models to partial data of an unfolding outbreak.

On the other hand, simple models are often capable of capturing the qualitative behavior of an epidemic with only a small number of parameters. However, these models often provide a good fit to data and perform well in short-term forecasts, but do not necessarily inform the exact mechanisms behind the transmission process or the effects of control interventions. For instance, standard compartmental models often assume early exponential growth in the absence of susceptible depletion and the effects of control interventions or behavior changes. Although this assumption is appropriate in some situations, it could be the result from the aggregation of local exponential and sub-exponential growth patterns at finer spatial scales [2,3]. In other words, model fit may depend on the spatial scale at work [4].

In this paper we mechanistically derive (or derive in the context of population contact structure and human behavioral change) and fit a family of logistic models to spatial-temporal data of the 1905 plague epidemic in Bombay, India. We systematically compare parameter estimates, reproduction numbers, model fit, and short-term forecasts across models at different spatial resolutions. At the same time, we also assess the presence of sub-exponential growth dynamics at different spatial scales [3,4] and investigate the role of spatial structure and data resolution (district level data and city level data) using simple structured models.

* Corresponding author.
*E-mail addresses:* pell1@stolaf.edu (B. Pell), tin.t.phan@asu.edu (T. Phan), erutter@ncsu.edu (E.M. Rutter), gchowell@gsu.edu (G. Chowell), kuang@asu.edu (Y. Kuang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.mbs.2018.04.003
Received 9 February 2018; Received in revised form 10 April 2018; Accepted 10 April 2018
Available online 16 April 2018

### 1.1. Background on the 1905–1906 plague in Bombay

Bubonic plague is an infectious disease that is caused by the bacterium *Yersinia pestis*. Bubonic plague is spread through inflected flea bites, as well as exposure to the tissue or fluids of an animal infected with the plague, but not directly from human to human, unless it is the uncommon pneumonic plague [5]. Untreated, bubonic plague mortality is estimated at 30–60%.

It is hypothesized that the flea transmits bacteria to a host through the blocked flea paradigm [6]. The flea consumes the blood of an infected host, and the bacteria multiply until they form a blockage between the fleas' esophagus and stomach. This blockage can take anywhere from 5 days post infection to 2–3 weeks post infection [7]. The blockage prevents the flea from digesting new blood, until the blockage is resolved, implanting plague bacteria into a new host. The factors that influence the efficiency of transmission between flea and host are still not well understood [8].

When considering the epidemics of bubonic plague in India, spread of the disease was markedly seasonal. Early attempts at forecasting plague were informed by weather: low hot weather and monsoon season temperatures were expected to favor high incidence of plague [9]. Some have suggested the seasonality stems from the ability of the flea to transmit the bacteria to the new host during inclement temperatures, although new research suggests that temperature does not significantly affect efficiency of transmission [10]. Regardless, weather variation is a major factor in the spread of vector-borne diseases [11].

Plague was first observed in the Mandvi region of Bombay in 1896 [12]. Year after year, the plague epidemic appeared to persist in Bombay in a seasonal fashion [13]. Roughly half the city attempted to flee in order to escape the plague, but the plague spread throughout India in the following years. By 1914, more than 8.5 million people had perished, and some areas of India reported plague deaths until the 1940s [14]. A commission was formed in 1905 to investigate the mechanisms responsible for the high burden of plague. As part of this effort, the number of deaths due to plague on a bi-weekly basis were collected from 12 separate districts in Bombay by a team of scientists and medical professionals [15]. In 1906, the plague in Bombay exhibited a case fatality rate of approximately 90% [15].

### 1.2. Existing work on the 1905–1906 plague in Bombay

The first influential mathematical model investigating the spread of bubonic plague in Bombay was constructed by Kermack and McKendrick [16], an ODE model that included three compartments: susceptible individuals, infected individuals, and immune (or recovered) individuals. While this model was able to fit the curve of human deaths due to plague in 1906, it was not designed to explain multi-year plague outbreaks. Additionally, as direct human-human infection due to pneumonic plague was rare, this homogeneous mixing modeling framework was not well suited to explain the underlying disease dynamics of this plague epidemic. A more recent model of plague transmission dynamics [17] includes 5 compartments: susceptible rats, infected rats, recovered rats, average number of fleas per rat, and free fleas. Using this epizootic model, it was possible to estimate the number of human cases. Their stochastic metapopulation model was used to predict the probability that the epidemic would persist in the rodent population for more than 1, 2, and 10 years [18]. Other researchers have furthered this model. For instance, in ref. [19], an extended 9-compartment model includes the human population (susceptible, infected, recovered and victim) while removing the emphasis on inherited resistance in the rat population previously given in ref. [17]. They also used a Holling type 2 search strategy for the rat population. With this model, they were able to show agreement with the Bombay 1905–1906 epidemic data. Recently, Bacaer extended the model to consider the seasonality of the plague by linking temperature with the transmission rate between fleas and humans [20]. The resulting model successfully explains multiple years of data. In addition to mechanistic transmission models of plague, a number of studies have analyzed spatiotemporal transmission patterns of plague [21,22]. A notable example of a long term study in ref. [23] used statistical global modeling techniques to study the correlation between outbreaks among humans and the epizootics in two species of rats (*M. decumanus* and *M. rattus*).

While previous modeling efforts are useful to understand the mechanisms of the plague, they do not address the problem in real time, especially in the early phases. Additionally, plague outbreak data is often limited to the number of new cases or deaths in a particular area, it complicates the use of complex multi-host mechanistic models that involve many parameters. For this reason, it is often more appropriate to start outbreak investigations using simple phenomenological models. Here we build on previous work [2,3,24] to investigate the 1905 Bombay plague epidemic.

### 1.3. Models employed in this study

The generalized-growth model (GGM) has been previously shown to be a useful tool to characterize the early growth trajectory of an infectious disease outbreak [25]. The GGM takes the following form,

$$x' = rx^p \tag{1}$$

where $x(t)$ is the cumulative number of infections at time $t$, $r$ is the intrinsic rate of infection and $p$ $(0 \leq p \leq 1)$ is the scaling of growth parameter [25]. When $p = 1$ we recover the exponential growth model and when $0 \leq p \leq 1$, the early growth phase is following sub-exponential growth.

To incorporate sub-exponential growth into a model for forecasting past the ascending phase of an epidemic, we employ the generalized Richards equation (GRM), which is a generalized version of the Richard model [26]:

$$x' = rx^p\left(1 - \left(\frac{x}{K}\right)^a\right). \tag{2}$$

Here, $x(t)$ is the cumulative number of infections at time $t$, $r$ is the intrinsic rate of infection, $K$ is the final epidemic size, $p$ $(0 \leq p \leq 1)$ is the scaling of growth parameter and "the exponent of deviation" $a$ is a parameter that modulates the peak-time of incidence [25,27]. We obtain the original Richard model when $p = 1$ and the logistic equation when $p = 1$ and $a = 1$.

The generalized-Richards model has been previously used to fit and forecast single-epidemic outbreaks. Recently, both the GRM and GGM were used to study the 2001 epidemic foot-and-mouth disease in the UK, showing the ability of the models to generate short term forecasts in absence of extensive data [28]. In a similar way, the GRM was used to make forecasts in the recent Ebola and Zika outbreaks [29,30]. Special cases of this model have also been used to fit past outbreaks of several infectious diseases. For instance, a multi-phase Richards model was successful in describing the trajectory of the SARS outbreak in Toronto including estimates of turning points of the epidemic curve [31]. Another study has used the Richards equation to model four separate cases of epidemics: H1N1 in Canada, SARS in the greater Toronto area, dengue in Singapore, and SARS in Taiwan [27], and directly related the parameters in the compartmental SIR model with those of the Richards equation.

Lastly, we incorporate spatial structure explicitly by using a logistic growth patch-model as studied in ref. [24]. Specifically, we study the effect that movement and quarantine strategies have on the final epidemic size. Based on the layout in Fig. 1, it is natural to consider a three-patch model where patch one includes Fort North and South, patch two includes Mandvim, Chakla, Market, Oomarkhadi, and Dongri and patch three includes Dhobi Talao, Bhuleshwar, Fanaswadi, Khara Talao, Khumbharwada, and Khetwadi. The three patch model takes the following form,

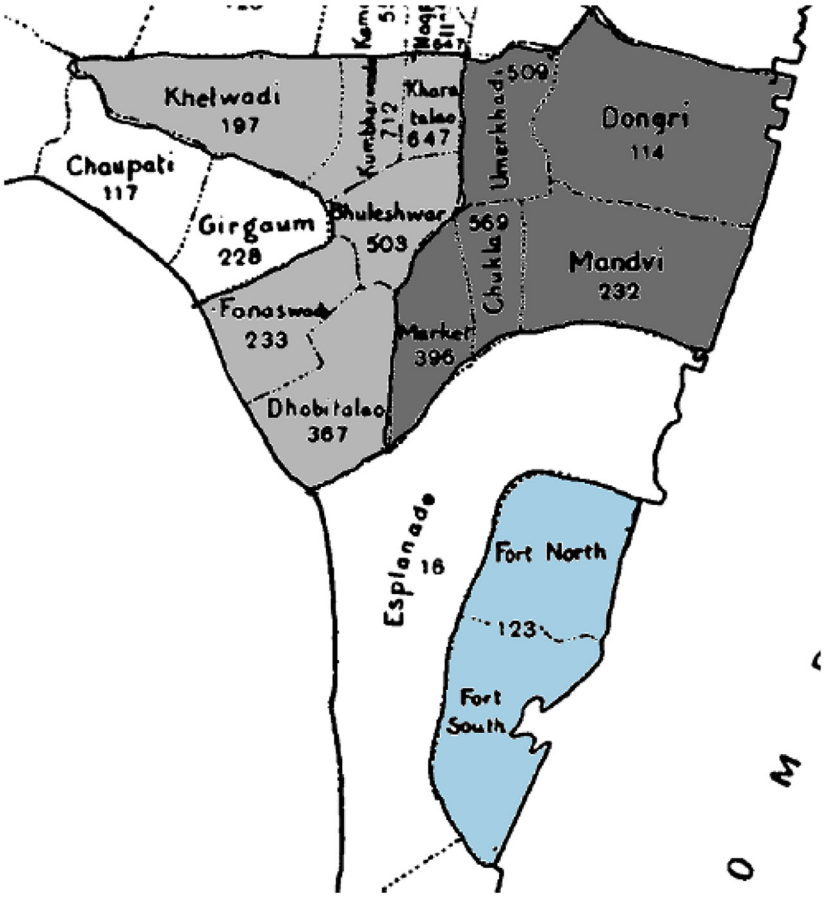



**Fig. 1.** Zoomed in view of the study area taken from [15]. Region/patch 1, 2 and 3 are the blue, dark grey and grey area respectively. The numbers show the population per acre. To study the spatial component, we consider the three regions as divided here. Since we do not have data for Esplanade, this region is not used directly in the modeling. However, taking advantage of the fact that Esplanade creates a natural barrier between Fort North and South and the other two regions, we choose the corresponding regions in a way that allows for dynamics where two patches were right next to each other and another was "disconnected". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$x_i' = r_i x_i\left(1 - \frac{x_i}{K_i}\right) + \sum_{j\neq i}^{3} m_{ji}x_j - \sum_{j\neq i}^{3} m_{ij}x_i \tag{3}$$

where $x_i$ is the cumulative number of infections in patch $i$, $m_{i,j}$ represents the rate at which cumulative infections travel from patch $i$ to patch $j$ and $K_i$ is the final epidemic size in patch $i$th if it were isolated from the rest.

### 1.4. Data

An advisory committee was formed in 1905 to collect extensive data on rat, flea and human cases through laboratory testing. These data were later used to infer the connection between the outbreak in the rat population with the outbreak in the human population. In this paper we employ incidence data collected from this commission which involved bi-weekly reports of plague mortality for each of the twelve districts that made up Bombay [15]. To resolve the non-uniformity in the starting time of the epidemic district-wise for the global fitting, we selected data points after which the outbreak has taken place in all districts.

## 2. Methods

### 2.1. Parameter estimation and confidence interval generation

We use the built-in MATLAB function LSQCURVEFIT to obtain parameter estimates within a predetermined lower and upper bounds via nonlinear least-squares. Parametric bootstrap with Poisson error structure was implemented to generate 200 model realizations based on the best fitted curve as described in ref. [32]. From these 200 realizations, we calculated 95% confidence intervals for model parameters and other epidemiological quantities [33].

### 2.2. Forecasting comparison

For the twelve districts of Bombay that we are studying, the onset week corresponds to the start of the monotone increase in the number of new cases. All models were initialized using the reported incidence at this point. Models were then trained on 4, 6 and 8 data points (corresponding to 8, 12 and 16 weeks) from this initial point. Incidence forecasts were made 2, 3 and 4 data points into the future (4, 6 and 8 weeks respectively).

### 2.3. Classification of exponential growth and the effective reproduction number

We define near-exponential growth to be when the confidence interval of the $p$ estimate contains values greater than 0.9. Near-exponential growth in this case means the exponential growth assumption in the early stage is acceptable. As in previous studies, we use the GGM to analyze the initial growth phase of incidence cases. Specifically, if the confidence interval for $p$ contains only values lower than 0.9, we categorize it as sub-exponential (polynomial) growth.

The effective reproduction number, $R_e(t)$, is defined as the average number of new infections generated by one infectious individual at time $t$. It can be estimated by using incidence data that is generated by the generalized-growth model and information about the probability distribution of the generation interval [25]. We numerically estimated the effective reproduction number by generating case incidence data from the GGM and using the discretized renewal equation:

$$R_e(t_i) = \frac{I_i}{\sum_{j=0}^{i} I_{i-j}\rho_j}, \tag{4}$$

where $I_i$ denotes incidence at calendar time $t_i$, $\rho_j$ denotes the discretized

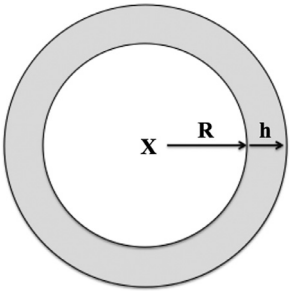



**Fig. 2.** Spatial representation of the system in 2 dimensions.

probability distribution of the generation interval. We assumed a generation interval with a Gamma distribution of 7 days and a standard deviation of 2 days [34].

Additionally, the effective reproduction number can be directly computed for the GGM with the following formula [25]:

$$R_g = \left[1 + \frac{r(1-p)T_g}{r(1-p)gT_g + A}\right]^{\frac{p}{1-p}}, \tag{5}$$

where $g$ is the disease generation, $T_g$ is the generation interval taken to be 7 days for the 1905–1906 Bombay plague and $A = x(0)^{1-p}$ [34]. We train the GGM using the first 3, 4 and 5 data points (6, 8 and 10 weeks) to obtain estimations for $r$ and $p$.

## 3. Results

### 3.1. Model derivations

The usual presentation of logistic models implies the notion that the intrinsic growth rate, $r$, and the carrying capacity, $K$, are independent, which neglects the possibility of a connection between the two as shown in chapter 6 of [35]. Here, we provide two mechanistic derivations of the GRM (and thus the logistic equation) in an epidemic setting.

Our first derivation of the Richards Model is adapted from a similar derivation found in [24]. Assuming 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 system of differential equations that model the susceptible ($S$) and infectious individuals ($I$):

$$\begin{aligned} S'(t) &= -\frac{\beta S(t) I(t)}{S(t) + I(t)}, \\ I'(t) &= \frac{\beta S(t) I(t)}{S(t) + I(t)} - \mu I(t), \end{aligned} \tag{6}$$

where $\beta$ is the infection rate and $\mu$ is the disease induced death rate. Note that we assume there is no recovery because the equation is derived specifically for the early phase of an epidemic. From system (6) 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 $x'(t) = \beta I$.

**Behavior assumption**: During an epidemic, a change in behavior in the community that mitigates the transmission rates is expected as an epidemic unfolds [36]. This might include better access to health care or, in the case of the Bombay plague epidemic, attempting to clean the streets and sewers to control the rat population [37]. This response is modeled by a function of the total reported cases and has a decreasing effect on the per-capita infection rate. That is,

$$\frac{I'(t)}{I(t)} = f(x(t)) \tag{7}$$

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

In the following, we assume that $f(x(t)) = r(1 - (bx(t))^a)$ for some positive constants $r$, $b$ and $a$. The chosen $f(x(t))$ has the simplest function form that reflects the desired property on the per-capita infection rate. Hence,

$$I'(t) = rI(t)(1 - (bx(t))^a) = \frac{r}{\beta}x'(t)(1 - (bx(t))^a).$$

Therefore,

$$I(t) - I(0) = \frac{r}{\beta}\left(x(t) - \frac{b^a}{a+1}[x(t)]^{a+1}\right) - \frac{r}{\beta}\left(x(0) - \frac{b^a}{a+1}[x(0)]^{a+1}\right).$$

Since $I(0) = x(0) \approx 0$, we see that $I(t)$ can be approximated by $\frac{r}{\beta}\left(x(t) - \frac{b^a}{a+1}[x(t)]^{a+1}\right)$. Therefore,

$$x'(t) = \beta I(t) = r\left(x(t) - \frac{b^a}{a+1}[x(t)]^{a+1}\right) = rx(t)\left(1 - \left(\frac{x(t)}{K}\right)^a\right), \tag{8}$$

where $K = \frac{(a+1)^{1/a}}{b}$. Here we interpret $r$ as the *intrinsic infection rate*, $b$ is a proportionality constant that corresponds to strength and effectiveness of disease interventions and preventive strategies, $K$ is the *final epidemic size* and the parameter $a$ corresponds to a nonlinear change in the behavior response which can be interpreted as the intensity of the behavior response as more cases are reported. Note that this is the GRM with $p = 1$.

The mechanistic derivation of the GRM follows from a geometric interpretation. Let $X$ be the population of total infected individuals living in the circle of radius $R$, see Fig. 2. We assume homogeneous mixing of the population within this circle, so $X$ is proportional to the area of this circle, e.g. $X = a\pi R^2$ for some proportional constant $a$. This is equivalent to $R = \sqrt{X/a\pi}$.

Since the infection spreads via contact, the growth of $X$ is proportional to the immediate surrounding (grey) area with a small radius extension $h$. Note that this is a reasonable assumption due to the chance of encountering a susceptible person from outside the infected circle is higher than one from the inside. This is especially true if the outbreak is in its early stage (the circle is small), so the ratio between the immediate surrounding and the area of the circle is relatively high. Additionally, we assume the death rate of $X$ is proportional to some power of its size. Thus we have

$$X' = \beta 2\pi Rh - dX^q$$

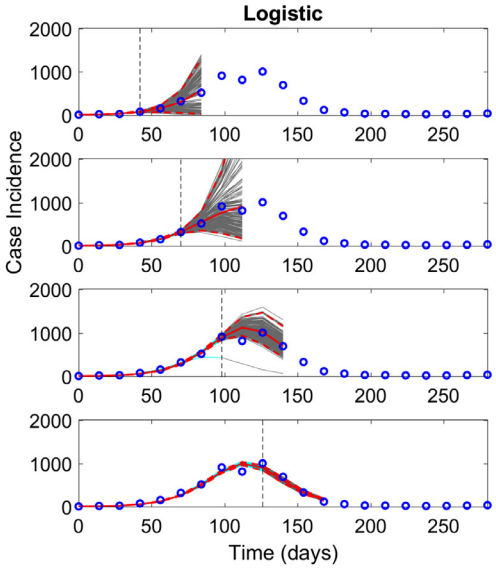
where $\beta$ and $d$ are corresponding proportional constants. Using the relationship of $R$ and $X$, this equation becomes

$$X' = 2\pi h\beta(a\pi)^{-1/2}X^{1/2} - dX^q = rX^{1/2}\left[1 - \left(\frac{X}{K}\right)^{q-\frac{1}{2}}\right]$$

where $r$: $=2\ h\beta\left(\frac{\pi}{a}\right)^{1/2}$ and $K$: $=\left[2\ h\beta\left(\frac{\pi}{ad^2}\right)^{1/2}\right]^{1/(q-1/2)}$. The value of $p$ in the 2 dimensional case is 1/2. However, as the dimension increases, the value of $p$ increases strictly and approaches 1, e.g. in 3 dimensions $p = 2/3$. The dependence of the value of $p$ on the dimension of the problem is a geometrical property of the derivation. The higher dimension perhaps links to the more number of ways for the outbreak to spread, which would make sense given the higher number of dimension corresponds to the higher value of $p$.

### 3.2. Forecasting comparison

The Logistic model and GRM incidence forecasts for the district Fanaswadi and the entire city of Bombay are shown in Figs. 3 and 5, respectively. There are three clear trends that should be noted. First, the confidence interval is smaller when more data are used for fitting. Secondly, global forecast and fitting are better than local forecast and fitting. Lastly, forecasting confidence intervals, calibrated on data past the peak of the epidemic, become significantly smaller in comparison to



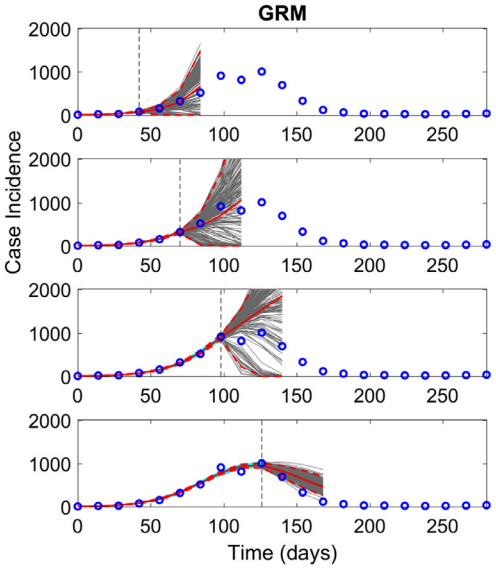
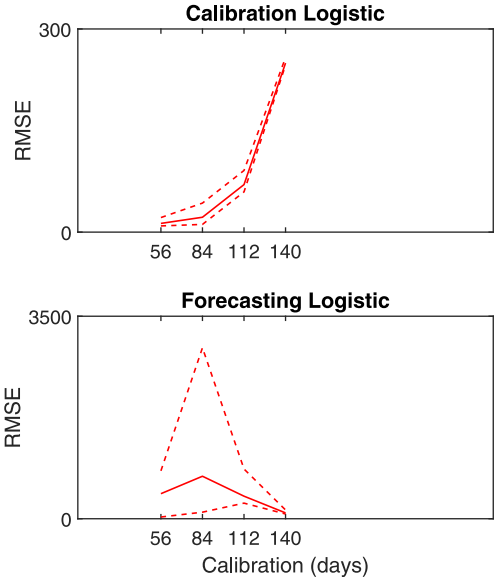




**Fig. 3.** Six–week ahead incidence forecasts of Bombay city. Left column: Logistic model fitted using an increasing amount of incidence data: 8, 12, 16 and 20 epidemic weeks. Right column: The generalized Richards model fitted using an increasing amount of incidence data: 8, 12, 16 and 20 epidemic weeks. Gray curves correspond to the ensemble of 200 realizations for the model forecast. The mean (solid red line) and 95% CIs (dashed red lines) of the model fit ensembles (gray curves) are also shown. The vertical dashed line separates the calibration and forecasting periods. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



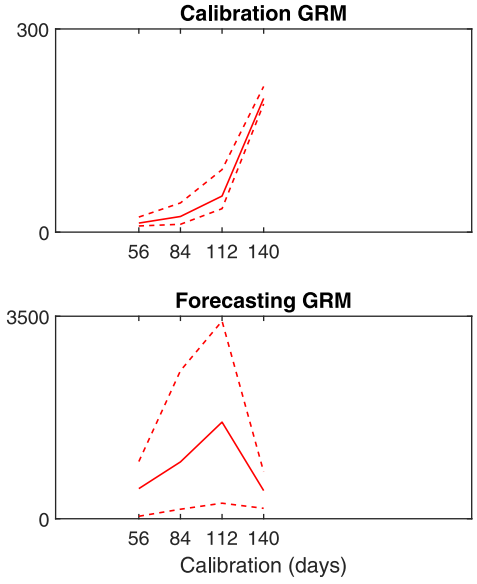
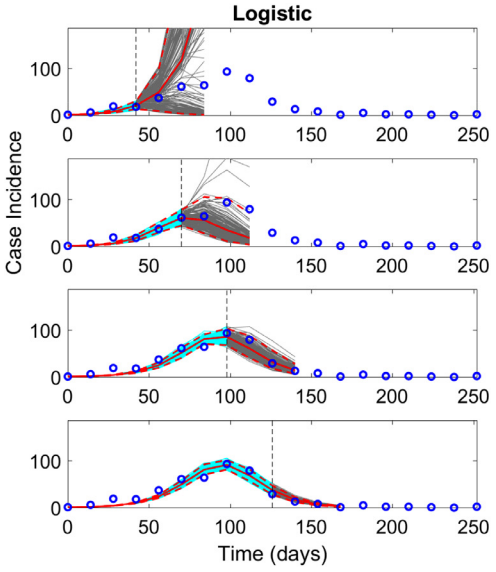




**Fig. 4.** RMSE for calibration and Six–week ahead incidence forecasts of the 1905 Plague epidemic in Bombay city. Left column: RMSE of logistic model. Right column: RMSE error of the generalized Richards model. The mean (solid red line) and 95% CIs (dashed red lines) are also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



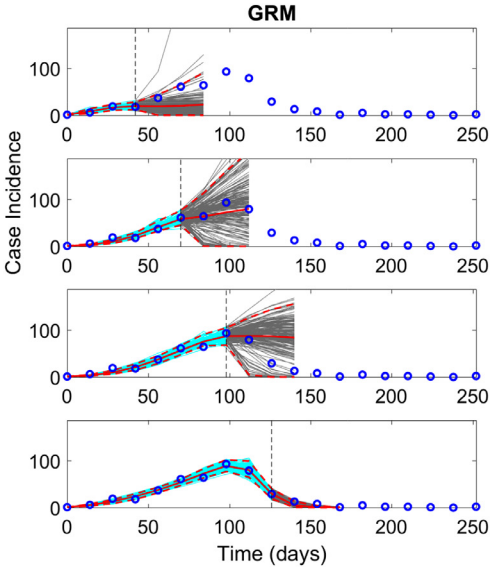




**Fig. 5.** Six–week ahead incidence forecasts of Fanaswadi. Left column: Logistic model fitted using an increasing amount of incidence data: 8, 12, 16 and 20 epidemic weeks. Right column: The generalized Richards model fitted using an increasing amount of incidence data: 8, 12, 16 and 20 epidemic weeks. Gray curves correspond to the ensemble of 200 realizations for the model forecast. The mean (solid red line) and 95% CIs (dashed red lines) of the model fit ensembles (gray curves) are also shown. The vertical dashed line separates the calibration and forecasting periods. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

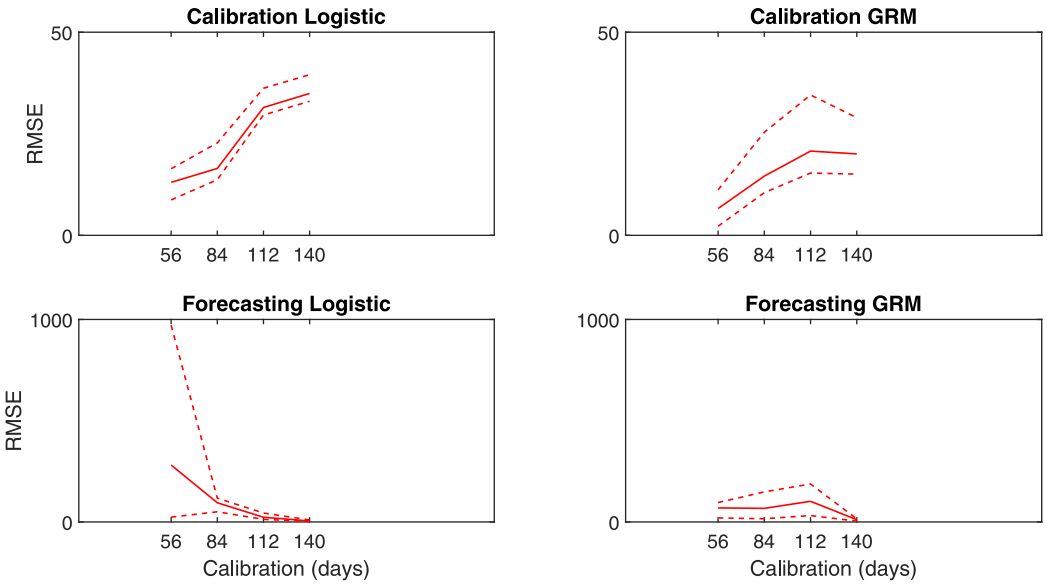



**Fig. 6.** RMSE for calibration and Six–week ahead incidence forecasts of Fanaswadi. Left column: RMSE of logistic model. Right column: RMSE error of the generalized Richards model. The mean (solid red line) and 95% CIs (dashed red lines) are also shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**
Summary table of key epidemiological quantity estimates and forecasting errors provided by the GGM and GRM. Both models were trained on 8 weeks of incidence data. Forecasting error is based on short-term forecasts 4 weeks into the future. Quantities inside bracket represent the 95% confidence interval.

| District | $p$ (GGM) | $R$ Renewal Eq. | $R$ Eq. (5) | GRM Forecasting Error (RMSE) | Logistic Forecasting Error (RMSE) |
|---|---|---|---|---|---|
| Fort North & South | 0.60 [0, 1] | 1.20 [0.40, 2.01] | 1.41 [1.00, 1.65] | 31.02 [3.87, 42.49] | 24.46 [1.85, 42.22] |
| Mandvi | 0.89 [0.31, 1] | 1.38 [0.89, 1.77] | 1.38 [1.22, 1.48] | 23.11 [2.20, 57.50] | 20.00 [1.85, 54.09] |
| Chakla | 0.99 [0.87, 1] | 1.56 [1.40, 1.73] | 1.68 [1.62, 1.73] | 317 [64.66, 570] | 242 [28.7, 544] |
| Market | 0.61 [0, 1] | 1.20 [0.73, 1.69] | 1.43 [1.00, 1.60] | 60.51 [14.33, 79.38] | 55.81 [10.37, 77.99] |
| Oomarkhadi | 0.74 [0.54, 0.98] | 1.37 [1.16, 1.57] | 2.10 [1.94, 2.24] | 131 [31.00, 344] | 98.36 [52.61, 117] |
| Dongri | 0.74 [0.32, 1] | 1.30 [0.95, 1.63] | 1.61 [1.40, 1.74] | 52.10 [4.01, 149] | 56.53 [13.70, 138] |
| Dhobi Talao | 0.85 [0.53, 1] | 1.46 [1.11, 1.75] | 1.80 [1.64, 1.99] | 76.16 [5.52, 250] | 82.53 [15.64, 258] |
| Bhuleshwar | 0.89 [0.66, 1] | 1.45 [1.20, 1.70] | 1.75 [1.65, 1.87] | 109 [11.71, 270] | 116 [14.11, 270] |
| Fanaswadi | 0.39 [0.05, 0.73] | 1.13 [0.82, 1.40] | 1.74 [1.11, 2.08] | 47.61 [7.69, 71.34] | 91.57 [9.40, 256] |
| Khara Talao | 0.89 [0.63, 1] | 1.64 [1.38, 1.86] | 2.47 [2.11, 2.74] | 242 [43.16, 479] | 169 [24.51, 427] |
| Khumbharwada | 0.65 [0.38, 0.91] | 1.30 [1.02, 1.56] | 2.10 [1.84, 2.26] | 99.02 [87.23, 98.60] | 88.83 [54.38, 95.66] |
| Khetwadi | 0.67 [0.22, 1] | 1.28 [0.86, 1.63] | 1.72 [1.36, 1.92] | 67.49 [25.10, 127] | 66.08 [25.96, 129] |
| Global | 0.96 [0.78, 1] | 1.44 [1.27, 1.60] | 1.57 [1.53, 1.62] | 175 [15.69, 358] | 149 [14.64, 290] |

forecasting using only data before the peak of the epidemic. It should be noted that long term forecast (several months) can have high variability, especially when model calibration uses 8 and 12 weeks of incidence data. Additionally, in general, the logistic model has comparable forecasting properties to the Generalized Richard model, for example see Figs. 4 and 6. More generally, training the models on 12-week incidence data and forecasting 8 weeks in to the future, gives a root mean square error (RMSE) mean of 296 (median = 104, IQR: [70.6, 185]) for the generalized Richards model and a RMSE mean of 304 (median = 103, IQR: [67.1, 183]) for the logistic model. We refer the reader to Table 1 and supplemental Tables S1 and S2 for more details concerning the comparison of forecasting errors.

In a previous study, the deceleration of growth parameter ($p$) during the epidemic in Bombay was estimated in the range of [0.68, 1] with mean value of 0.86 [3]. Our results at the city level agree with this estimation (95% CI: [0.78, 1] with mean 0.96), yet are slightly higher when using 8 weeks of incidence data. However, estimations at the sub-city level show that sub-exponential growth is quite common. Indeed, Fig. 7 illustrates that with 8 and 10 weeks of data the 95% confidence intervals of $p$ shrink around values consistent with sub-exponential growth. In general, $p$ varies across geographic locations at the sub-city level (median = 0.64, IQR: [0.42, 0.74], using 6 weeks of incidence data). Estimations of $p$ using 8 weeks of epidemic data using the GGM are summarized in Table 1. Supplementary Tables S1 and S2 provide $p$ estimations for 12 and 16 weeks. The difference in estimations of $p$ at the district and city level support the hypothesis that exponential growth at the global level may be a result of multiple sub-exponential growth profiles aggregated at the local level.

The estimations of the effective reproduction number in Fig. 8 which are summarized in Table 1 and supplementary Tables S1 and S2 show four clear trends. First, the estimations using Eq. (5) tend to have narrower confidence intervals than the ones obtained from the Renewal Equation (Eq. (4)). Second, the more data points used for fitting parameters, the smaller the confidence intervals are. Third, the confidence intervals for global estimates are smaller in comparison to the estimates for district. Finally, most of the estimations hover around the value of 1.5 and within the range between 1 and 2, with the exception of Khumbharwada. This shows consistency between the two methods of $R$ estimation. Indeed, the effective reproduction number is fairly uniform across geographic locations at the sub-city level (median = 1.23 IQR: [1.15, 1.33], using 12 weeks of incidence data and the renewal equation). Furthermore, our global estimates are similar to the estimate found in literature with median = 1.37 and IQR: [1.19, 1.61] [25]. This result shows interesting similarities and contrasts between the two methods for estimating the reproduction number with respect to the number of data points used for fitting and the geographical variations.

### 3.3. Patch model fitting

As a numerical exercise to investigate how migration may change

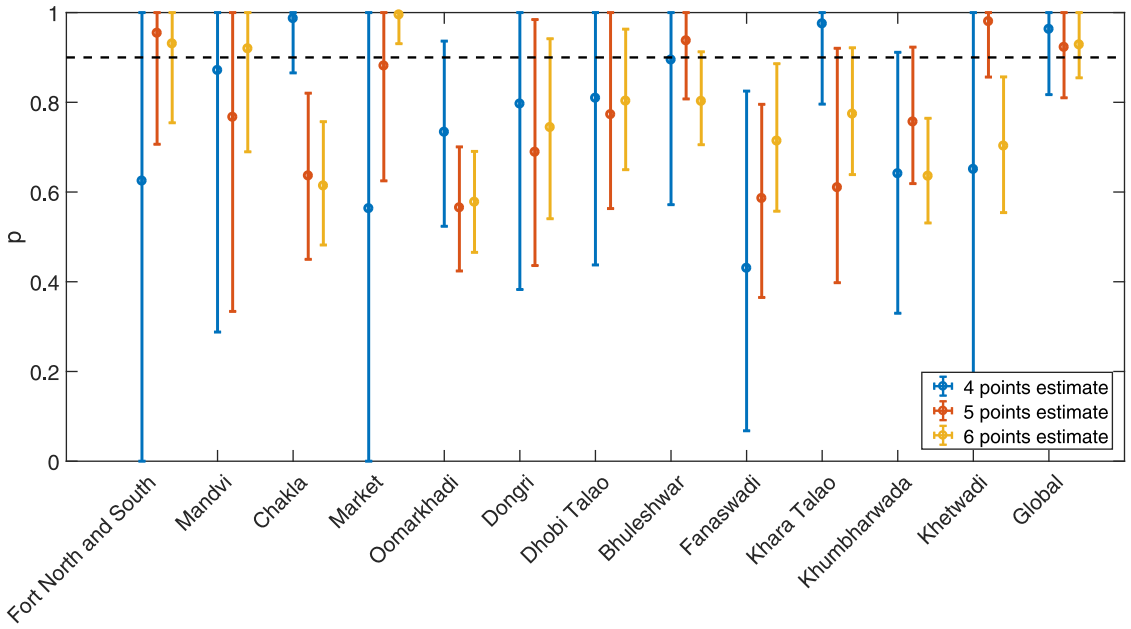



**Fig. 7.** 95% confidence intervals for parameter $p$ obtained by nonlinear least square fitting of the generalized growth model (GGM) to an increasing amount of data (8–12 weeks) across the twelve different districts of Bombay. The global estimation of $p$ corresponds to when all district data is aggregated together. Below the horizontal dashed line is the region that classified as sub-exponential growth.

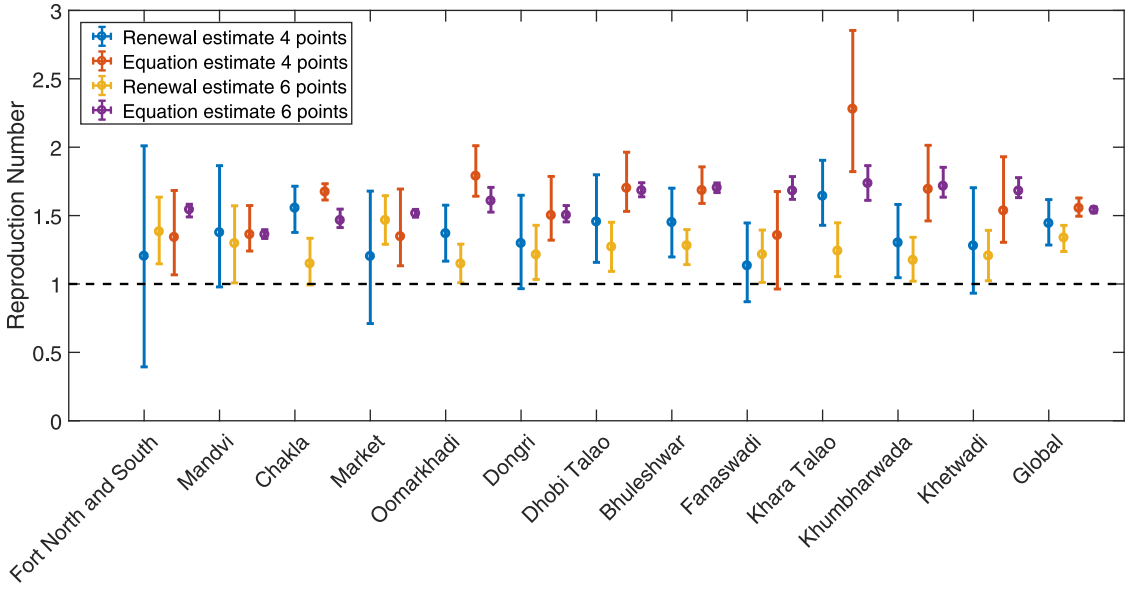



**Fig. 8.** $R$ estimates using 8 and 12 weeks of incidence data. Estimates are generated in two ways: 1) the Renewal equation (Eq. (4)) coupled with 200 GGM ensemble incidence forecasts and 2) parameter estimations from 200 GGM ensemble incidence forecasts used in Eq. (5). Estimations are at the district and global level. The black dash line is reference for the value 1.

the final epidemic size of an outbreak, we fit the 3-patch model simultaneously to three data sets comprised of 3-patches in Bombay as shown in Fig. 1. Parameter estimations and model fits are shown in Fig. 9 and migration bifurcation plots are shown in Figs. 10 and 11. Bifurcation plots with respect to migration between patches 2 and 3 show that certain levels of movement between the two patches can produce final epidemic sizes that are well over the final epidemic size given when the patches are isolated from each other ($K_1 + K_2 + K_3$). A possible explanation as to why this occurs is because the movement of a patch composed of many smaller districts will be significantly larger than the individual rate of movement of each district. Thus, if each district has a substantial rate of movement corresponding to sub-exponential growth, we can expect that the effective rate of movement of the global patch is large enough to account for near exponential growth. The left panel in Fig. 10 shows the importance of sending people to a less volatile patch, in this case, patch 1. That is, increasing migration into patch 1 from the neighboring patches effectively reduced the overall final epidemic size. On the other hand, migration out of patch 1 and into the more volatile patches 2 and 3 ultimately decreases the overall final epidemic size, but less effectively.

Comparison between the rate of movement with relation to patch 1 (Fort North and South) and rates of movement between patch 2 and 3 show agreement with our understanding of the situation during Bombay plague. Individuals have limited movement and geographical barrier presents difficulty in transition to patch 1.

Result on global asymptotic stability and existence of a unique

| Param. | Value | Param. | Value |
|---|---|---|---|
| $r_1$ | 0.088 | $m_{12}$ | 0.055 |
| $r_2$ | 0.062 | $m_{21}$ | 0.004 |
| $r_3$ | 0.189 | $m_{13}$ | 0.010 |
| $K_1$ | 333 | $m_{31}$ | 0.055 |
| $K_2$ | 2367 | $m_{23}$ | 0.607 |
| $K_3$ | 2520 | $m_{32}$ | 0.602 |

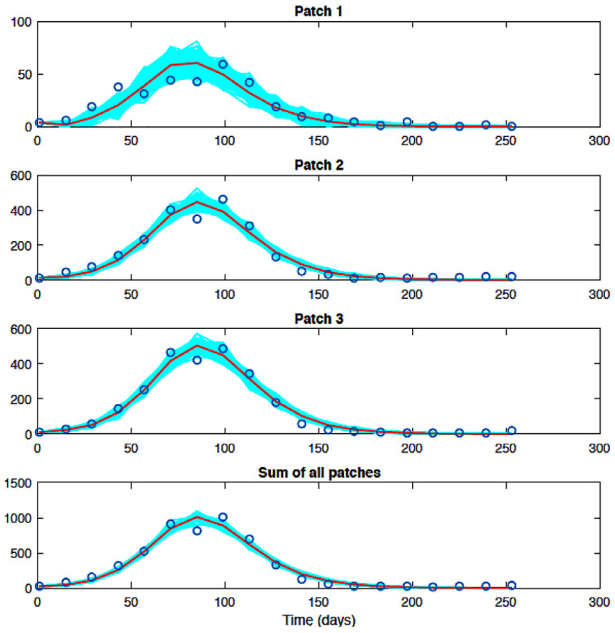



**Fig. 9.** (Right) Model fit to incidence data from Bombay. Patch 1 consists of Fort North and South. Patch 2 consists of Mandvim, Chakla, Market, Oomarkhadi, and Dongri. Lastly, patch 3 comprises of Dhobi Talao, Bhuleshwar, Fanaswadi, Khara Talao, Khumbharwada and Khetwadi. (Left) The parameter estimates of the best fit for patch model.

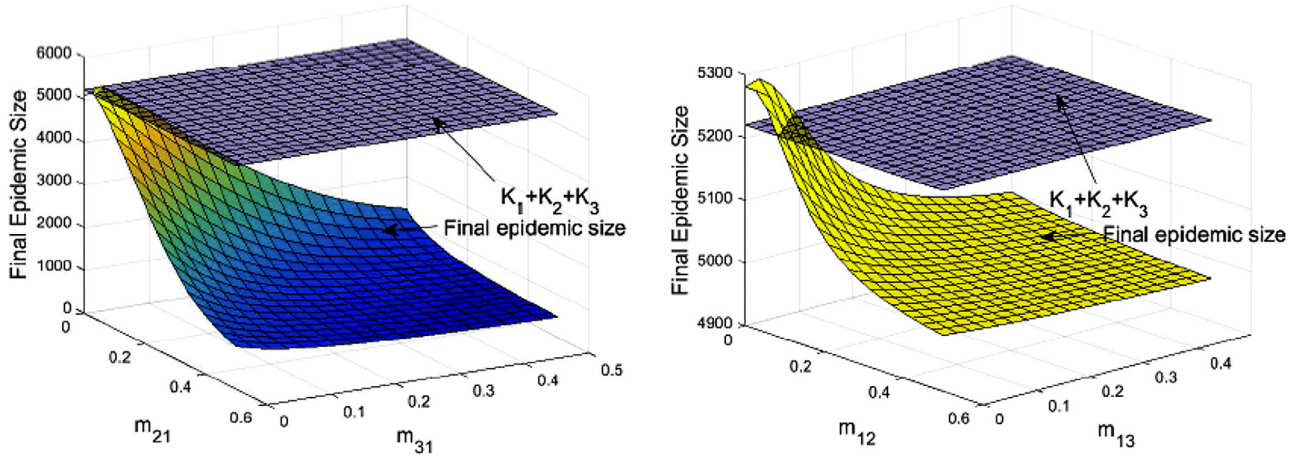



**Fig. 10.** The effect of controlling migration rates in and out of Fort North and Fort South and how they influence the final epidemic size. **Recall the notation:** $m_{ij}$ **means migration from patch i to patch j**. (**Left**: migration into Fort North and South). In this case, the model shows that the final epidemic size is a decreasing function of $m_{21}$ and $m_{31}$. (**Right**: migration out of Fort North and South). Increasing the rate that people from Fort North and South entering into the rest of the city has a decreasing effect on the final epidemic size; however, this may not be true when the rate of movement is very small. Note that the impact of migration is not as significant as in the previous case.

positive steady state for an n-dimensional patch system is shown in Lemma 3.1 in ref. [38], which is proved by applying Theorem 6.1 in Hirsch [39]. This would in fact imply the boundedness of the cumulative infection. However, the boundedness of the logistic model is often given from the corresponding carrying capacity. On the other hand, our simulation shows that it is possible for the population size to grow beyond the cumulative carrying capacity of the patches in the logistic patch model.

Interestingly, to the best of our knowledge, there has not been any results establishing the conditions for the final epidemic size to surpass the cumulative carrying capacity (e.g. $K_1 + K_2 + K3$). This may yield insightful consequences with potential public health implications. A meaningful inequality is not trivial in this case, as shown below.

Fig. 12 suggests a direct way to obtain such condition in the case of two patch logistic model, or precisely, $x_1^* + x_2^* > K_1 + K_2$ if and only if $x_1^b > x_1^a$. However, solving algebraically for something tractable from the inequality $x_1^b > x_1^a$ seems unlikely. A clear mathematical result would probably require a case by case consideration with respect to the relation of *r, m* and *K*. We will explore this mathematical problem in subsequent work.

## 4. Discussion

In this paper, we derived the generalized Richards model from first principles and related the deceleration parameter *p* to the spatial dimension in which the disease is spreading. In addition, we used the generalized-growth model to detect sub-exponential growth trajectories at the sub-city level and estimate the effective reproduction number at the sub-city and city level. We also assessed short-term forecast trajectories of the logistic equation and the GRM. Results suggest that the epidemic growth at the global level is near-exponential with *p* near 1 (95% CI: [0.78, 1] with mean 0.96), while the scaling of growth at the

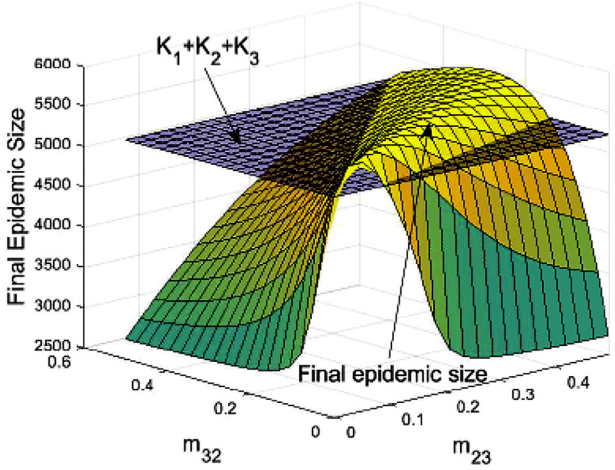



**Fig. 11.** Final epidemic size as a function of $m_{23}$ and $m_{32}$. **Recall the notation: $m_{ij}$ means migration from patch i to patch j**. The model predicts that it is possible for the overall final epidemic size to be larger than $K_1 + K_2 + K_3$.

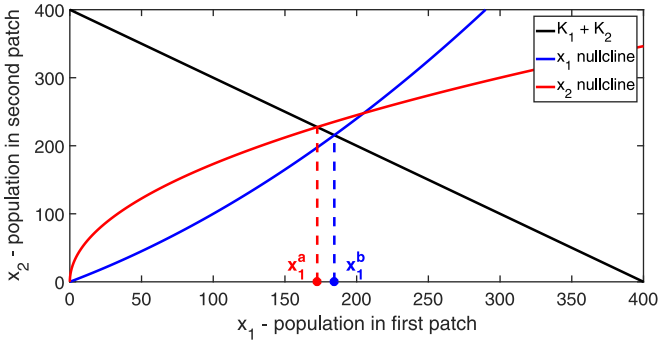



**Fig. 12.** $x_1$ and $x_2$ nullclines plotted with the cumulative carrying capacity $K_1 + K_2$. The intersections of the $x_1$ and $x_2$ nullclines with the cumulative carrying capacity have projections onto the $x_1$-axis denoted by $x_1^a$ and $x_1^b$ respectively.

local level is mostly in the range of sub-exponential.

The general assumption of exponential growth is not always optimal (perhaps not even appropriate) in the early phase of an epidemic [2,3]. Exponential growth is a phenomenon well associated with cell growth in an unlimited nutrient environment, but in a complex situation such as the spread of an outbreak, it is overly simplistic. Intuitively, even in the case of patient zero, where we could argue for unlimited supply of "nutrient", the rate of growth should still be dependent on other factors such as the range of movement of the individuals within the region [4]. Additionally, this early exponential stage of an outbreak can become short-lived once the disease has spread to more than a few individuals. The initial growth rate is influenced by many factors that need to be captured for a realistic understanding of the situation.

For the majority of the districts in Bombay city, our results suggest that sub-exponential growth is quite common, but at the city scale, growth is found to be near-exponential. Exponential growth at the city scale is also consistent with what has been reported by another study [25]. Ultimately, our results are preliminary and sub-exponential growth is most likely associated with multiple mechanisms (e.g., complex population mixing structures, reactive behavior changes, control interventions, local customs).

A possible explanation of sub-exponential growth may be the network structure of the susceptible population. In this case, we consider the small-world network interpretation [40]. If we suppose that each person has a close circle of friends/family members of size $F$. Then in a small world, $F$ can be approximated as $x^{1/n}$ where $n$ is the degree of separation (ranging between 2 and 8) and $x$ is the population. We can then partition the population $x$ into $X_i$, $i = 1, \ldots, n$. Here we define $X_1$ to be the initial set of infected individuals, $X_2 = x_1 F$ to be the close friends of $X_1$, $X_3$ to be the close friends of $X_2$ excluding those already in $X_2$ and $X_1$ and so forth. We assume that once the disease has spread beyond the $X_2$ population into the $X_3$, most of the total population would have been aware of this development and take counter measures to protect themselves from getting infected - recall the behavioral assumption in the first derivation. In other words, the disease is not likely to affect the population in $X_4$ and higher as effectively. This reasoning suggests that the real susceptible population to be $X_1 + X_2 + X_3$, which is approximately $x^{3/n}$, which implies $p = 3/n$.

Our analysis shows that the growth rate during the ascending phase of an epidemic can depend on the spatial scale of the data being used. In other words, global spatial scales can mask valuable information that can only be obtained by looking at local spatial scales. Indeed, we have shown that sub-exponential growth rates at the sub-city level could aggregate together and produce near exponential growth at the city level during the 1905 Bombay epidemic. This phenomenon has been observed during the Western Africa Ebola epidemic of 2014 and the early dynamics of HIV/AIDS epidemics in Brazil [2,41]. Future studies of how different spatial scales change the estimation of epidemiological quantities and their underlying physical mechanisms are warranted. The GGM and the GRM are important tools for gaining valuable information before details of the transmission mechanisms of the pathogen are known and have proven to be indispensable theoretical tools to inform and guide the future direction of the modeling community [34]. More importantly, these models can prove to be quite useful in quickly informing the allocation of public health resources and intervention strategies by identifying locations that yield near-exponential epidemic growth.

The bifurcation diagrams of the final epidemic size based on the rate of movement are only meaningful in the early stage of the outbreak, because as the outbreak progresses, the "intrinsic rate of growth" changes. Furthermore, isolating or over-packing the infectious population into a patch will cause an increase in the "intrinsic rate of growth" or the "effective carrying capacity", so it is inappropriate to fix these parameters for an analysis in the long term. However, the bifurcation plots serve to show the significance of the rate of movement. Realistically, however, when the rate of movement between each patch increases, this causes a higher chance for a larger final epidemic size. This perhaps suggests that a higher rate of movement is closely connected to the exponential growth assumption. Additionally, we note that all movement parameters relating to patch 1, which represents a single district is much smaller in comparison to the movement between patch 2 and 3. The rate of movement between patch 2 and 3 is a form of an effective rate of movement between the districts in the two patches. This can be explained via the natural barrier of Esplanada that separates Fort North and South from the rest as seen in Fig. 1. This is because the rate of movement between patch 2 and 3 in the model is a form of effective rate of movement obtained through fitting, or an approximation of the total movement of the districts that belong to the patch.

It is worth noting that global stability analysis of dynamical systems is a dominating feature in the mathematical study of epidemiological models. However, attention should also be paid to non-standard aspects of mathematical analysis, for example the condition at which the cumulative infected population surpasses the total carrying capacity in a patch model. This observation not only defies common sense, but also questions the phenomenological definition of "carrying capacity". Our result suggests a more general and comprehensive mechanism linking the "intrinsic rate of growth" and the "carrying capacity" in standard logistic model, which we will explore in future work.

Parameter identifiability is a problem that can arise from complex phenomenological models and limited data. In this sense, simpler models often hold the advantage since they tend to be identifiable, especially in the case of mechanistic models. We derived the Logistic

family mechanistically and obtained explicit relations between parameters. Thus, this supports its identifiability, e.g. the models can be reduced to a simpler form with identifiable parameter. In the future, perhaps a mechanistic relation between $r$ and $p$ that needs to be explored to possibly reach an identifiable stage for the GGM. It should be noted that structural identifiability can still remain an issue as there are examples where a simple model remains non-identifiable.

Due to the devastating nature of the Black Death, there have been multiple studies and mathematical models addressing the spread of Black Death through Europe in Medieval times. However, it has been argued by historians that the Black Death may have been vastly different from modern plague in terms of transmission, spread, and acquired immunity [42]. While the Bombay Plague Commission asserted that the method of spread was due to the flea-rat-human vector [15], others have argued that the Black Death may have been spread through other means. Examining the rate at which disease spread through Europe, and given the unseasonable temperatures for flea reproduction and survival, ref. [43] demonstrates better fits to historical epidemic data with a human-lice vector rather than the human-rat-flea vector. Another study in ref. [44] used spatiotemporal stochastic modeling to compare and contrast characteristics of the Black Death in Europe with the bubonic plague spread in India. They concluded that the Black Death and the modern plague were different in many key aspects: Black Death moved much faster, seasonality is more important in modern plague, and Black Death mortality was much larger. Other researchers have also concluded that modern plague and Black Death were inherently different [45]. Therefore, it would be interesting to discover whether, despite the many differences between the Bombay plague and the Black Death, the simple model presented in this work would also apply to the Black Death.

For future work, we would like to expand on our analyses by comparing the predictive ability of each model under different scaling of growth rate assumptions to further strengthen our conclusion relating spatial scale and scaling of growth. Additionally, while we fix the parameters in our analysis assessing the effect of migration, it is more reasonable to let them change over time. Thus, it would be interesting to derive a mechanistic relationship between $m$, $r$ and $K$ from first principles with respect to other parameters in order to comprehensively study the impact of immigration on final epidemic size.

## Acknowledgments

This work was partially supported by the NSF grant DMS-1518529. In addition, the authors would like to thank the reviewers for many helpful comments that improved the presentation of this manuscript.

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mbs.2018.04.003.

## References

[1] R.M. Anderson, R.M. May, B. Anderson, Infectious Diseases of Humans: Dynamics and Control, 28 Wiley Online Library, 1992.
[2] G. Chowell, C. Viboud, J.M. Hyman, L. Simonsen, The Western Africa Ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates, PLoS Curr. 7 (2015).
[3] C. Viboud, L. Simonsen, G. Chowell, A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks, Epidemics 15 (2016) 27–37.
[4] M.A. Kiskowski, A three-scale network model for the early growth dynamics of 2014 west africa ebola epidemic, PLoS Curr. 6 (2014).
[5] N.C. Stenseth, B.B. Atshabar, M. Begon, S.R. Belmain, E. Bertherat, E. Carniel, K.L. Gage, H. Leirs, L. Rahalison, Plague: past, present, and future, PLoS Med. 5 (1) (2008) e3.
[6] A.W. Bacot, C. Martin, Lxvii. observations on the mechanism of the transmission of plague by fleas, J. Hyg. 13 (Suppl) (1914) 423.
[7] R.J. Eisen, K.L. Gage, Adaptive strategies of yersinia pestis to persist during inter-epizootic and epizootic periods, Vet. Res. 40 (2) (2009) 1.
[8] R.J. Eisen, L. Eisen, K.L. Gage, Studies of vector competency and efficiency of north american fleas for yersinia pestis: state of the field and future research needs, J. Med. Entomol. 46 (4) (2009) 737–744.
[9] L. Rogers, The methods and results of forecasting the incidence of cholera, smallpox and plague in india, Trans. R. Soc. Trop. Med. Hyg. 27 (1933) 217–238.
[10] A.M. Schotthoefer, S.W. Bearden, J.L. Holmes, S.M. Vetter, J.A. Montenieri, S.K. Williams, C.B. Graham, M.E. Woods, R.J. Eisen, K.L. Gage, Effects of temperature on the transmission of yersinia pestis by the flea, xenopsylla cheopis, in the late phase period, Parasites Vectors 4 (1) (2011) 191.
[11] K.L. Gage, T.R. Burkot, R.J. Eisen, E.B. Hayes, Climate and vectorborne diseases, Am. J. Prev. Med. 35 (5) (2008) 436–450.
[12] W. Gatacre, Report: Bubonie Plague in Bombay for 1896–97, Times of India Steam Press, Bombay, 1897.
[13] I. Klein, Urban development and death: Bombay city, 1870–1914, Mod. Asian Stud. 20 (4) (1986) 725–754.
[14] D. Arnold, Colonizing the Body: State Medicine and Epidemic Disease in Nineteenth-century India, Univ. of California Press, 1993.
[15] P.R. Commission, The epidemiological observations made by the commission in bombay city, J. Hyg. 7 (1907) 724–798.
[16] W.O. Kermack, A.G. McKendrick, A Contribution to the Mathematical Theory of Epidemics, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 115 The Royal Society, 1927, pp. 700–721.
[17] M. Keeling, C. Gilligan, Metapopulation dynamics of bubonic plague, Nature 407 (6806) (2000) 903.
[18] M. Keeling, C. Gilligan, Bubonic plague: a metapopulation model of a zoonosis, Proc. R. Soc. London B 267 (1458) (2000) 2219–2230.
[19] S. Monecke, H. Monecke, J. Monecke, Modelling the black death. a historical case study and implications for the epidemiology of bubonic plague, Int. J. Med. Microbiol. 299 (8) (2009) 582–593.
[20] N. Bacaër, The model of kermack and mckendrick for the plague epidemic in bombay and the type reproduction number with seasonality, J. Math. Biol. 64 (3) (2012) 403–422.
[21] H.L. Yu, G. Christakos, Spatiotemporal modelling and mapping of the bubonic plague epidemic in india, Int. J. Health Geogr. 5 (1) (2006) 12.
[22] H.L. Yu, A. Kolovos, G. Christakos, J.C. Chen, S. Warmerdam, B. Dev, Interactive spatiotemporal modelling of health systems: the seks–gui framework, Stochastic Environ. Res. Risk Assess. 21 (5) (2007) 555–572.
[23] S. Mangiarotti, Low dimensional chaotic models for the plague epidemic in bombay (1896–1911), Chaos Solitons Fractals 81 (2015) 184–196.
[24] B. Pell, J. Baez, T. Phan, D. Gao, G. Chowell, Y. Kuang, Patch Models of EVD Transmission Dynamics, Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases, Springer, 2016, pp. 147–167.
[25] G. Chowell, C. Viboud, L. Simonsen, S.M. Moghadas, Characterizing the reproduction number of epidemics with early subexponential growth dynamics, J. R. Soc. Interface 13 (123) (2016) 20160659.
[26] F. Richards, A flexible growth function for empirical use, J. Exp. Bot. 10 (2) (1959) 290–301.
[27] X.S. Wang, J. Wu, Y. Yang, Richards model revisited: validation by and application to infection dynamics, J. Theor. Biol. 313 (2012) 12–19.
[28] D.W. Shanafelt, G. Jones, M. Lima, C. Perrings, G. Chowell, Forecasting the 2001 foot-and-mouth disease epidemic in the uk, Ecohealth (2017) 1–10.
[29] B. Pell, Y. Kuang, C. Viboud, G. Chowell, Using phenomenological models for forecasting the 2015 ebola challenge, Epidemics (2018).
[30] G. Chowell, D. Hincapie-Palacio, J. Ospina, B. Pell, A. Tariq, S. Dahal, S. Moghadas, A. Smirnova, L. Simonsen, C. Viboud, Using phenomenological models to characterize transmissibility and forecast patterns and final burden of zika epidemics, PLoS Curr. 8 (2016).
[31] Y.H. Hsieh, Y.S. Cheng, Real-time forecast of multiphase outbreak, Emerging Infect. Dis. 12 (1) (2006) 122.
[32] G. Chowell, Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts, Infect. Dis. Modell. 2 (3) (2017) 379–398.
[33] G. Chowell, C. Ammon, N. Hengartner, J. Hyman, Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: assessing the effects of hypothetical interventions, J. Theor. Biol. 241 (2) (2006) 193–204.
[34] G. Chowell, L. Sattenspiel, S. Bansal, C. Viboud, Mathematical models to characterize early epidemic growth: a review, Phys. Life Rev. 18 (2016) 66–97.
[35] Y. Kuang, J.D. Nagy, S.E. Eikenberry, Introduction to Mathematical Oncology, 59 CRC Press, 2016.
[36] S. Funk, M. Salathé, V.A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: a review, J. R. Soc. Interface (2010). rsif20100142.
[37] G.C. Kohn, Encyclopedia of Plague and Pestilence: From Ancient Times to the Present, Infobase Publishing, 2007.
[38] D. Gao, S. Ruan, A multipatch malaria model with logistic growth populations, SIAM J. Appl. Math. 72 (3) (2012) 819–841.
[39] M.W. Hirsch, The dynamical systems approach to differential equations, Bull. Am. Math. Soc. 11 (1) (1984) 1–64.
[40] D.J. Watts, S.H. Strogatz, Collective dynamics of small-worldnetworks, Nature 393 (6684) (1998) 440–442.
[41] L. Dinh, G. Chowell, R. Rothenberg, Growth scaling for the early dynamics of hiv/aids epidemics in brazil and the influence of socio-demographic factors, J. Theor. Biol. (2018).
[42] S.K.Jr. Cohn, The black death: end of a paradigm, Am. Hist. Rev. 107 (3) (2002) 703–738.
[43] K.R. Dean, F. Krauer, L. Walløe, O.C. Lingjærde, B. Bramanti, N.C. Stenseth, B.V. Schmid, Human ectoparasites and the spread of plague in europe during the second pandemic, Proc. Natl. Acad. Sci. (2018) 201715640.
[44] G. Christakos, R.A. Olea, H.L. Yu, Recent results on the spatiotemporal modelling and comparative analysis of black death and bubonic plague epidemics, Public Health 121 (9) (2007) 700–720.
[45] M.R. Welford, B.H. Bossak, Validation of inverse seasonal peak mortality in medieval plagues, including the black death, in comparison to modern yersinia pestis-variant diseases, PloS One 4 (12) (2009) e8401.