Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 142 (2017) 113–129
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original articles

Analysis of bias in an Ebola epidemic model by extended Kalman filter approach

Denis Ndanguza[a,*], Isambi S. Mbalawata[b,c], Heikki Haario[d], Jean M. Tchuenche[e]

[a] *Department of Mathematics, College of Science and Technology, University of Rwanda, P.O. Box 3900 Kigali, Rwanda*
[b] *Department of Mathematics, University of Dar-es-Salaam, P.O. Box 35062, Dar-es-Salaam, Tanzania*
[c] *African Institute for Mathematical Sciences Tanzania, P.O. Box 176 Bagamoyo, Tanzania*
[d] *Department of Computational Engineering and Physics, Lappeenranta University of Technology, P.O. Box 20, FI-53851 Lappeenranta, Finland*
[e] *School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa*

Received 16 May 2016; received in revised form 15 May 2017; accepted 15 May 2017
Available online 26 May 2017

Abstract

Ebola is a highly infectious disease generally characterized by sporadic outbreaks. A deterministic Ebola model is formulated and converted into Itô stochastic differential equations by adding noise on each compartment. In order to estimate the model parameter values, we use the extended Kalman filter technique as the filtering method and sum of square of errors to compute an approximation of the likelihood. From the obtained likelihood function, the maximum likelihood and MCMC methods for parameters estimation are then used. These parameter estimates provide useful information on quantities of epidemiological interest. Two cases are analyzed: (1) the model error covariance is set to be zero and (2) the bias is fully incorporated into the model. A comparison between these two cases is carried out to assess whether the bias is having a measure effect on parameters and states estimation. Finally, we investigate whether an estimate obtained from a biased study differs systematically from the true source population of the study. Our results indicate that the more the increase of bias, the more the noise in states simulation and parameters estimation compared to the deterministic model.

Keywords: Ebola; Estimation of parameters; Uncertainty; Extended Kalman Filter; Markov Chain Monte Carlo

1. Introduction

Many phenomena in nature are affected by noise. Therefore, modeling such phenomena using stochastic differential equations (SDEs) is potentially one of the best ways to capture features that cannot be gleaned from using deterministic models [5,24,33]. Various mathematical models of infectious diseases are largely deterministic and the literature on this topic is quite extensive [1,2,7,20–22,28,29,32,37,40,44,45,47] to name but a few. Ordinary

* Corresponding author.
E-mail address: dndanguza@gmail.com (D. Ndanguza).

http://dx.doi.org/10.1016/j.matcom.2017.05.005

differential equations (ODEs) describe how systems change over time and justify the effect of the starting point to the initial solution. However, such models often do not take into consideration uncertainties which may occur. The objective of the proposed Ebola model is to address this limitation by formulating a Itô stochastic differential Ebola model. Using this approach, uncertainties are incorporated into the Ebola model studied in [32] in order to capture the random fluctuations which may be observed over time.

Deterministic and stochastic models have advantages and disadvantages. Sometimes stochastic models are preferred over deterministic ones: (1) most natural way of studying the spread of disease is stochastic since it defines the probability of transmission of disease between individuals [3], (2) stochastic models converge to deterministic when the population size becomes large [24], (3) if we consider a small community with an epidemic outbreak, it seems reasonable to assume some uncertainty in the final number infected [2], (4) deterministic models are not the most relevant for modeling the start of an epidemic where the number of infectious individuals is small [34], (5) estimation of model parameters and states require the knowledge about uncertainty around the estimates. Stochastic modeling enables estimation of parameters from disease outbreak data to be equipped with standard errors [29], and finally (6) stochastic models are, in general, more realistic since the spread of diseases is stochastic in nature [29]. We revisit and analyze the Ebola model proposed in [32], which has only onset and death data. The 1995 Ebola outbreak in Congo has been extensively studied [10,18,23,32]. The data consist of two time series recorded from March 1 to July 16,1995 namely, daily account of Ebola cases by date of symptoms onset with a total of 291 cases, and daily account of deaths from Ebola with a total of 236 deaths [23,32]. It is documented that the first case became ill on January 6, 1995 and the last case died on July 16 [18,23,25,26].

Ndanguza et al. [32] slightly modified the model proposed in [10] by splitting the removed compartment into two: Recovered and Death and used both the least squares and MCMC methods to estimate the model parameters. Although both deterministic and stochastic mathematical models of Ebola abounds in the literature [4,7,10,11,14,17,27,35,36,38,42,43,46], none of these models has considered estimation of model parameters using the extended Kalman filter (EKF) or contrasted results of unbiased and biased Ebola models. It is not always possible or desirable to measure every variable that one would like to control. However, the filtering method may provide a means for inferring the missing information from indirect and noisy measurements [9]. The known advantage of filtering techniques is the simultaneous computation of states and likelihood function of parameters. The likelihood function can be used in Bayesian method like MCMC to estimate the parameters. For instance, the extended Kalman filter is used for predicting the likely future courses of dynamic systems that people are not likely to control [9,12].

Herein, we revisit the model proposed in [32] and analyze it using the extended Kalman filter method for parameters estimation. This approach requires converting the deterministic Ebola model into a Itô SDE Ebola model by adding a bias or noise on each compartment. For parameters estimation, the EKF algorithm computes the likelihood function and then we use the maximum likelihood and MCMC algorithms. Finally, we compare results from the unbiased Ebola model with those from the biased Ebola model.

The rest of this paper is organized as follows: The SDE model is formulated in Section 2. Analysis of unbiased and biased Ebola model is carried out respectively in Sections 3 and 4. Section 5 is the discussion followed by conclusion in Section 6.

2. Model framework

We assume the population is closed, that is, the effect of demographic changes (birth and natural death) is minimal during the course of the epidemic. This assumption of a closed community makes sense in the context of the Congo data because of the relatively short course of the disease. The total population size N is divided in four compartments: $S(t)$ susceptible individuals at time t, the exposed class $E(t)$ with an average incubation period of $1/k$ days before progressing to the infectious class $I(t)$, and the removed (death $D(t)$ or recovered $R(t)$) class $R(t)$ (where without any ambiguity of notation the R will henceforth be referred to as recovered class). The rate of new infections is modeled using the frequency-dependent transmission function, also known as the standard incidence [31]. In the absence of treatment, the R-class is termed removed because individuals reaching it will never have the chance to re-join the process. The model flowchart is depicted in Fig. 1. It should be noted that $C(t)$ is not a compartment, but serves to keep track of cumulative number of Ebola cases from the onset of symptoms [10].