Infectious Disease Modelling 2 (2017) 379–398

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts

Gerardo Chowell [a, b, *]

[a] Division of Epidemiology & Biostatistics, School of Public Health, Georgia State University, Atlanta, GA, USA
[b] Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

ARTICLE INFO

Article history:
Received 4 June 2017
Accepted 8 August 2017
Available online 12 August 2017

Keywords:
Parameter estimation
Uncertainty quantification
Bootstrap
Parameter identifiability
Model performance
Forecasts
Uncertainty propagation

ABSTRACT

Mathematical models provide a quantitative framework with which scientists can assess hypotheses on the potential underlying mechanisms that explain patterns in observed data at different spatial and temporal scales, generate estimates of key kinetic parameters, assess the impact of interventions, optimize the impact of control strategies, and generate forecasts. We review and illustrate a simple data assimilation framework for calibrating mathematical models based on ordinary differential equation models using time series data describing the temporal progression of case counts relating, for instance, to population growth or infectious disease transmission dynamics. In contrast to Bayesian estimation approaches that always raise the question of how to set priors for the parameters, this frequentist approach relies on modeling the error structure in the data. We discuss issues related to parameter identifiability, uncertainty quantification and propagation as well as model performance and forecasts along examples based on phenomenological and mechanistic models parameterized using simulated and real datasets.

1. Introduction

Emerging and re-emerging infectious diseases are undoubtedly one of humankind's most important health and security risks (Fauci & Morens, 2016). As epidemic threats increase so is the potential impact of mathematical and statistical inference and simulation approaches to guide prevention and mitigation plans. As the recent 2013–2016 Ebola epidemic exemplified, an infectious disease outbreak often forces public health officials to make key decisions to mitigate the outbreak in a changing environment where multiple factors positively or negatively impact local disease transmission (Chowell et al., 2017). Hence, public health officials are often interested in practical yet mathematically rigorous and computationally efficient approaches that comprehensively assimilate data and model uncertainty to 1) generate

* School of Public Health, Georgia State University, Atlanta, GA, USA.
E-mail address: gchowell@gsu.edu.
Peer review under responsibility of KeAi Communications Co., Ltd.

http://dx.doi.org/10.1016/j.idm.2017.08.001

estimates of key transmission parameters, 2) assess the impact of control interventions (vaccination campaigns, behavior changes), 3) test hypotheses, 4) evaluate how behavior changes affect transmission dynamics, 5) gain insight to the contribution of different transmission pathways, 6) optimize the impact of control strategies, and 7) generate short and long-term forecasts, just to name a few.

Mathematical models provide a quantitative framework with which scientists can assess hypotheses on the potential underlying mechanisms that explain patterns in the observed data at different spatial and temporal scales. Models vary in their complexity in terms of the number of variables and parameters that characterize the dynamic states of the system, in their spatial and temporal resolution (e.g., discrete vs. continuous time), and in their design (e.g., deterministic or stochastic). While agent-based models, formulated in terms of characteristics and interactions among individual agents, have become increasingly used to model detailed processes often occurring at multiple scales (e.g., within host vs. population level), models based on systems of ordinary differential equations are widely used in the biological and social sciences. These dynamic models are specified by a set of equations and their parameters that together quantify the spatial-temporal states of the system via a set of interrelated dynamic quantities (e.g, viral load, susceptibility levels, disease prevalence) (Banks et al., 2009).

In this paper we review and illustrate a simple data assimilation framework for connecting ordinary differential equation models to time series data describing the temporal progression of case counts relating to population growth or infectious disease transmission dynamics (e.g, daily incident cases). This frequentist approach relies on modeling the error structure in the data unlike Bayesian approaches which always raise the question of how to set priors for the parameters. We present examples based on phenomenological and mechanistic models of disease transmission dynamics together with simulated and real datasets. We discuss issues related to parameter identifiability, uncertainty quantification and propagation as well as model performance and forecasts.

2. Mathematical models

The general form of a dynamic model composed by a system of h ordinary differential equations is given by

$$\begin{aligned}\dot{x}_1(t) &= f_1(x_1, x_2, \ldots, x_h, \Theta)\\ \dot{x}_2(t) &= f_2(x_1, x_2, \ldots, x_h, \Theta)\\ &\vdots\\ \dot{x}_h(t) &= f_h(x_1, x_2, \ldots, x_h, \Theta)\end{aligned}$$

where $\dot{x}_i$ denotes the rate of change of the system states x_i where $i = 1, 2, \ldots h$ and $\Theta = (\theta_1, \theta_2, \ldots, \theta_m)$ is the set of model parameters.

In general, the complexity of a model is a function of the number of parameters that are needed to characterize the states of the system and the spectrum of the dynamics that can be recovered from the model (e.g., number of equilibrium points, oscillations, bifurcations, chaos). A trade-off exists between the level of model complexity and the ability to reliably parameterize the model with available data.

In the next sections we briefly discuss differences between phenomenological and mechanistic models along specific examples that will become useful to illustrate methodology in the subsequent sections.

2.1. Phenomenological models

Phenomenological models provide an empirical approach without a specific basis on the physical laws or mechanisms that give rise to the observed patterns in the data (Chowell et al., 2016a). Thus, these types of models emphasize the reproducibility of empirical observations using simple models. Next, we describe two useful models to characterize epidemic growth patterns namely the generalized-growth model (GGM) and the generalized Richards model (GRM).

2.1.1. The generalized-growth mode (GGM)

This is a phenomenological model that has proved useful to characterize and forecast early epidemic growth patterns (Chowell & Viboud, 2016; Viboud, Simonsen, & Chowell, 2016). In particular, previous analyses highlighted the presence of early sub-exponential growth patterns in infectious disease data across a diversity of disease outbreaks (Viboud et al., 2016). The generalized-growth model allows relaxing the assumption of exponential growth via a "scaling of growth" parameter, p. The model is given by the following differential equation:

$$C'(t) = rC^p(t) \tag{1}$$

where $C'(t)$ describes the incidence growth phase over time t, the solution $C(t)$ describes the cumulative number of cases at time t, r is a positive parameter denoting the growth rate, and p, the "deceleration of growth" parameter varied between 0 and 1. If $p = 0$, this equation describes constant incidence over time and the cumulative number of cases grows linearly while $p =$