Journal of Biological Systems, Vol. 19, No. 1 (2011) 71–100

DOI: 10.1142/S0218339011003798

World Scientific
www.worldscientific.com

A DETERMINISTIC METHODOLOGY FOR ESTIMATION OF PARAMETERS IN DYNAMIC MARKOV CHAIN MODELS

A. R. ORTIZ*, H. T. BANKS†, C. CASTILLO-CHAVEZ*,
G. CHOWELL* and X. WANG*

*School of Human Evolution and Social Change
Arizona State University
Tempe, Arizona 85287-2402, USA

†Center for Research in Scientific Computation
North Carolina State University
Box 8212, 2700 Stinson Drive
Raleigh, NC 27695-8212, USA
†htbanks@ncsu.edu

Received 3 May 2010
Accepted 18 July 2010

A method for estimating parameters in dynamic stochastic (Markov Chain) models based on Kurtz's limit theory coupled with inverse problem methods developed for deterministic dynamical systems is proposed and illustrated in the context of disease dynamics. This methodology relies on finding an approximate large-population behavior of an appropriate scaled stochastic system. The approach leads to a deterministic approximation obtained as solutions of rate equations (ordinary differential equations) in terms of the large sample size average over sample paths or trajectories (limits of pure jump Markov processes). Using the resulting deterministic model, we select parameter subset combinations that can be estimated using an ordinary-least-squares (OLS) or generalized-least-squares (GLS) inverse problem formulation with a given data set. The selection is based on two criteria of the sensitivity matrix: the degree of sensitivity measured in the form of its condition number and the degree of uncertainty measured in the form of its parameter selection score. We illustrate the ideas with a stochastic model for the transmission of vancomycin-resistant enterococcus (VRE) in hospitals and VRE surveillance data from an oncology unit.

Keywords: Markov Chain Stochastic Models; Inverse Problems; Parameter Estimation; Parameter Selection; Large Population Sample Path Approximations.

1. Introduction

Closely tied to the formulation of the mathematical models is the need to estimate the parameters (including initial conditions) involved as well as to provide uncertainty bounds for the estimates. Validating the mathematical models with empirical data for the system under investigation is a key step in gaining insight into the system process and evaluating the effectiveness of particular control strategies.[1–7]

†Corresponding author.

A number of advanced mathematical and statistical tools for parameter estimation in deterministic dynamic models are readily available. The key objective of this paper is to present a methodology to estimate parameters in a stochastic model using inverse problem methods developed for deterministic dynamical systems. In these inverse problem methods, parameter estimates along with uncertainty bounds (confidence intervals) are readily obtained from longitudinal data for a single realization of the observation process for the stochastic system. Moreover, sensitivity analysis along with parameter selection (determining which parameters are most "identifiable" with the given data) can be done without massive simulation studies.

It is difficult to carry out the above estimation related tasks directly with stochastic models and limited data. Even so, there is a long and substantial record[8–11] (along with the many references therein) of research efforts in this area. Much of this literature involves a Bayesian approach to Maximum Likelihood Estimators (MLE) for transition parameters in Markov Chain models for infectious disease progression. In addition to requiring additional assumptions (over those for least squares estimation[12] on the observation errors, it is well known that such an approach is computationally intensive even in the case of deterministic models.[13] The alternative methodology presented in this paper (which we employed in a slightly different setting[14]) is based on using an approximate large-population behavior of an appropriate scaled stochastic system using Kurtz's limit theory.[15,16] By scaling the stochastic system and applying the Strong Law of Large Numbers (SLLN) for the relevant Poisson process, we can derive the corresponding deterministic approximation as solutions of rate equations in terms of the large sample size average over sample paths or trajectories. Using the resulting deterministic model, we select parameter subset combinations (a recently developed methodology[17] for deterministic models) that can be accurately estimated using an ordinary-least-squares (OLS) or generalized-least-squares (GLS) inverse problem formulation with a given data set along with an appropriate statistical model for the observation process. As noted above (and discussed in detail elsewhere[12]) an MLE approach requires additional *a priori* assumptions (specifically one must assume a form for the observation error distributions in forming the likelihood function) and therefore for convenience and generality, we favor an OLS or GLS approach when appropriate.

Given an experimental data set, a mathematical model may be more sensitive to some parameters than others, and the dependence between the parameters can impact the well-posedness of an inverse problem. Therefore, it is of interest to limit the attempted estimation to subsets of parameters for which the mathematical model is most sensitive. The analysis used to select the type of inverse problem formulation and the subset of parameters to be estimated from a given data set is based on previous ideas in the literature,[12,17,18] and is reviewed in Sec. 4. The selection procedure is based on two criteria of the sensitivity matrix: the degree of sensitivity measured in the form of its condition number and the degree of uncertainty measured in the form of its parameter selection score. The idea is to first select

all parameter combinations with a full rank sensitivity matrix and then calculate the corresponding asymptotic standard errors and selection scores. Then, parameter subset combinations with small condition numbers and selection scores are considered as feasible (i.e., can be estimated with reasonable levels of uncertainty). The resulting parameter estimates for transition parameters with their associated uncertainty measures can then be used in original Markov Chain model for simulation/predictive investigations.

The motivation for this manuscript is derived from our interest in understanding the spread of infectious diseases in particular, nosocomial infections, in order to prevent major outbreaks in hospital settings. Thus, in Sec. 2, we introduce a stochastic model of the transmission of Vancomycin-Resistant Enterococcus (VRE) in hospitals that is used to illustrate the methodology introduced in this paper. This model was developed in our initial studies of VRE[19] (see also other studies[3,7,20–28] on VRE). In Sec. 2.2, we show in detail how to obtain the corresponding deterministic approximation using Kurtz's theory. We follow in Sec. 3 with a description of the surveillance data motivating our efforts and the parameters that can be estimated directly from the data. In Sec. 4, we review the inverse problem and parameter selection methodology used to estimate parameters and quantify uncertainty for problems with deterministic systems. Finally, in Secs. 5 and 6, we present some illustrative results and along with some summary conclusions.

2. A Motivating VRE Stochastic Model

For our motivating example model, patients in a hospital unit are classified by compartments or states as either uncolonized $U(t)$, VRE colonized $C(t)$, or VRE colonized in isolation $J(t)$, as depicted in the compartmental schematic of Fig. 1. A description of the variables and parameters are given in Table 1. Patients are admitted to the hospital unit at a rate Λ per day and some fraction m are already VRE colonized. The transition from one compartment to another follows an exponential distribution and the expected mean duration within a compartment is given by the inverse of the parameter of the exponential distribution. A hand-hygiene policy applied to health care workers on isolated VRE colonized patients reduces infectivity by a factor of γ $(0 < \gamma < 1)$. It is assumed at VRE colonization periods last from weeks to months and because spontaneous decolonization occurs infrequently, clearance of the bacteria is not considered in the model. VRE colonized patients are moved into isolation at a rate α.

2.1. *The VRE stochastic model*

The dynamics of the VRE colonization of patients in a hospital unit are modeled as a continuous time Markov Chain (MC) with discrete state space[14,29,30] embedded in $\mathbb{R}^3$. In this case, the population of patients is considered discrete (i.e., VRE colonization occurs in units of whole individuals) and the timing of events is a

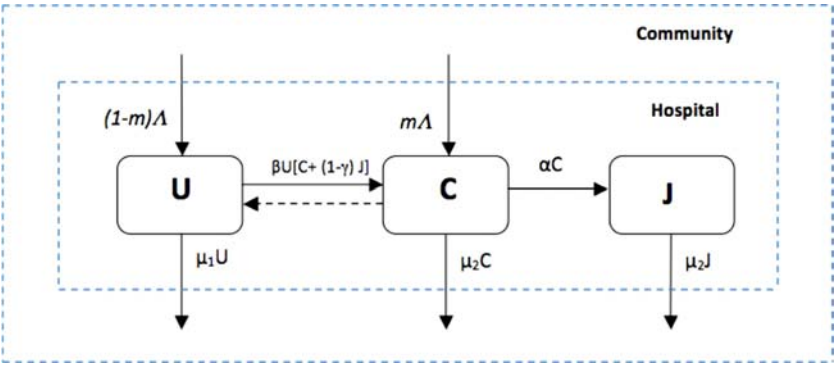


Fig. 1. Compartmental VRE model.

Table 1. Model parameters.

Variables	Description	Units
$U(t)$	Number of uncolonized patients	Individuals
$C(t)$	Number of VRE colonized patients	Individuals
$J(t)$	Number of VRE colonized patients in isolation	Individuals
Parameters	Description	Units
Λ	Patients admission rate	Individuals/day
m	VRE colonized patients on admission rate	Dimensionless
β	Effective contact rate	1/day
γ	HCW hand hygiene compliance rate	Dimensionless
α	Patient isolation rate	1/day
μ_1	Uncolonized patients discharged rate	1/day
μ_2	VRE colonized patients discharged rate	1/day

probabilistic process. The state of the MC at time t is denoted by $\{U(t) = i, C(t) = j, J(t) = k\}, t \geq 0$ and $i, j, k \in \{0, 1, \ldots, N\}$. The probability during a small time interval, dt of transiting from one state to another is described by

$$P\{U(t+dt) = i-1, C(t+dt) = j+1, J(t+dt) = k \,|\, U(t) = i, C(t) = j, J(t) = k\} = m\mu_1 U\,dt + \beta U[C + (1-\gamma)J]dt + o(dt), \tag{1}$$

$$P\{U(t+dt) = i, C(t+dt) = j+1, J(t+dt) = k-1 \,|\, U(t) = i, C(t) = j, J(t) = k\} = m\mu_2 J\,dt + o(dt), \tag{2}$$

$$P\{U(t+dt) = i+1, C(t+dt) = j-1, J(t+dt) = k \,|\, U(t) = i, C(t) = j, J(t) = k\} = (1-m)\mu_2 C\,dt + o(dt), \tag{3}$$

$$P\{U(t+dt) = i+1, C(t+dt) = j, J(t+dt) = k-1 \,|\, U(t) = i, C(t) = j, J(t) = k\} = (1-m)\mu_2 J\,dt + o(dt), \tag{4}$$

$$P\{U(t+dt) = i, C(t+dt) = j-1, J(t+dt) = k+1 \,|\, U(t) = i, C(t) = j, J(t) = k\} = \alpha C\,dt + o(dt), \tag{5}$$

$$P\{U(t+dt)=i, C(t+dt)=j, J(t+dt)=k \mid U(t)=i, C(t)=j, J(t)=k\}$$
$$=(1-m)\mu_1 U\,dt + m\mu_2 C\,dt + [1-(\Lambda+\beta U[C+(1-\gamma)J]+\alpha C)]\,dt + o(dt). \tag{6}$$

In the VRE epidemic model a constant population is assumed in which the hospital remains full for all t (i.e., overall admission rate equals overall discharge rate, $\Lambda = \mu_1 U + \mu_2(C+J)$). Hence, the admission of a patient in either compartments U or C are dependent events on the discharged in either compartment U or C or J (or vice versa). We assume that when a patient is discharged from the hospital, he/she is immediately replaced by an admission into the compartment U with probability $(1-m)$ or into the compartment C with probability m. Equation (1) is the probability of entering compartment C by either an admission (due to a discharge in compartment U) or by effective colonization. Equation (2) is the probability of entering compartment C by an admission due to a discharge in J. Equation (3) is the probability of admission to compartment U by a discharge in C. Equation (4) is the probability of admission into compartment U by a discharge in J. Equation (5) is the probability of moving a VRE colonized patient into isolation. Finally, Eq. (6) is the probability that none of the states changes due to: an uncolonized patient being discharged and replaced back into the U compartment, or a VRE colonized patient in C being discharged and replaced back into the C compartment, or no event occurs.

2.2. *The deterministic approximation*

When dividing Eqs. (1)–(6) by dt and taking the limit when dt tends to 0+, we obtain the rates of transitions as summarized in Table 2. In the stochastic model, the rates represent the mean number of transitions that can be expected in a given period, with the actual numbers distributed about these means. Hence, the rates determine the frequencies of occurrence through time for the transitions or events. We emphasize that it is these transition rates that

Table 2. Transition rates.

Event	Effect	Transition rate
Discharge of uncolonized patient	$(U,C,J)=(i-1,j,k)$	$\lambda_1=\mu_1 U$
Discharge of VRE colonized patient	$(U,C,J)=(i,j-1,k)$	$\lambda_2=\mu_2 C$
Discharge of VRE colonized patient in isolation	$(U,C,J)=(i,j,k-1)$	$\lambda_3=\mu_2 J$
Patient colonization due to VRE colonized patients	$(U,C,J)=(i-1,j+1,k)$	$\lambda_4=\beta UC$
Patient colonization due to VRE colonized patients in isolation	$(U,C,J)=(i-1,j+1,k)$	$\lambda_5=\beta(1-\gamma)UJ$
Isolation of VRE colonized patient	$(U,C,J)=(i,j-1,k+1))$	$\lambda_6=\alpha C$
Admission of uncolonized patient	$U=i+1$	$(1-m)(\lambda_1+\lambda_2+\lambda_3)$
Admission of VRE colonized patient	$C=j+1$	$m(\lambda_1+\lambda_2+\lambda_3)$

are the fundamental parameters of the Markov chain model. Thus, it is quite important to accurately estimate (by whatever means possible) these parameters. To facilitate estimation of these transition rates, we use a fundamental approximation idea due to Kurtz[15,16] (see also Chapter 5 of the text by Andersson and Britton[8]) which permits use of a wide range of deterministic estimation techniques.

We follow arguments we have used earlier in a similar situation[14] and let $R_i(t)$ for $i = 1, \ldots, 6$, be the number of times that the ith transition has occurred by time t. Then, the state of the system at time t can be written as

$$\begin{aligned} U(t) &= U(0) - R_1(t) - R_4(t) - R_5(t) + (1-m)(R_1(t) + R_2(t) + R_3(t)) \\ C(t) &= C(0) - R_2(t) + R_4(t) + R_5(t) - R_6(t) + m(R_1(t) + R_2(t) + R_3(t)) \\ J(t) &= J(0) - R_3(t) + R_6(t), \end{aligned} \tag{7}$$

where $R_i(t)$ is a counting process with intensity $\lambda_i(U(t), C(t), J(t))$ given by

$$R_i(t) = Y_i\left(\int_0^t \lambda_i(U(s), C(s), J(s))ds\right), \quad i = 1, \ldots, 6, \tag{8}$$

with Y_i as independent unit Poisson processes. Note that the state of the system is $(U(s), C(s), J(s))$ and hence each $\lambda_i(U(s), C(s), J(s))$ is constant between transition times. Also, note that sample paths $r_i(t)$ of $R_i(t)$ are given in terms of sample paths $(u(t), c(t), j(t))$ of $(U(t), C(t), J(t))$ by

$$r_i(t) = Y_i\left(\int_0^t \lambda_i(u(s), c(s), j(s))ds\right), \quad i = 1, \ldots, 6. \tag{9}$$

Let $U^N(t) = U(t)/N$, $C^N(t) = C(t)/N$, $J^N(t) = J(t)/N$ be the patient units per system size or the proportion in the stochastic process. The corresponding sample paths are $(u^N(t), c^N(t), j^N(t))$. We express the rates λ_i for $i = 1, \ldots, 6$ in terms of these scaled variables as follows:

$$\begin{aligned} \lambda_1 &= \mu_1 u(t) = N\mu_1 u^N(t), \quad & \lambda_4 &= \beta u(t)c(t) = N^2\beta u^N(t)c^N(t), \\ \lambda_2 &= \mu_2 c(t) = N\mu_2 c^N(t), \quad & \lambda_5 &= \beta(1-\gamma)u(t)j(t) = N^2\beta(1-\gamma)u^N(t)j^N(t), \\ \lambda_3 &= \mu_2 j(t) = N\mu_2 j^N(t), \quad & \lambda_6 &= \alpha c(t) = N\alpha c^N(t). \end{aligned}$$

Using these rates, we can obtain an approximation for $r_i^N(t)$, the averages of the $r_i(t)$ of (9) by applying the SLLN for the Poisson Process (i.e., $Y(N\mu)/N \approx \mu$). The process results in approximating for large N, the sample paths $(u^N(t), c^N(t), j^N(t))$ by a large sample size average approximation over paths $(\bar{u}(t), \bar{c}(t), \bar{j}(t))$ defined by a deterministic system. That is, we approximate integrals in the averaged sample path equations by integrals that are used as the defining equations for the deterministic

sample paths. This is done by the approximations

$$\begin{aligned} r_1^N(t) &= \frac{r_1(t)}{N} = \frac{1}{N} Y_1\left(\int_0^t \lambda_1(u(s))ds\right) \\ &= \frac{1}{N} Y_1\left(N\int_0^t \mu_1 u^N(s)ds\right) \approx \frac{1}{N} Y_1\left(N\int_0^t \mu_1 \bar{u}(s)ds\right) \\ &\approx \int_0^t \mu_1 \bar{u}(s)ds. \end{aligned} \tag{10}$$

The approximations for $r_i^N(t)$ for $i = 2, \ldots, 6$, can be obtained similarly. When dividing both sides of the sample path analogue of each equation in (7) by N and applying the approximations for $r_i^N(t)$, we can write the system of integral equations (i.e., rate equations) that approximate the stochastic Eqs. (7) via the SLLN. The rate approximation equations are given by

$$\begin{aligned} u^N(t) &= u^N(0) - r_1^N(t) - r_4^N(t) - r_5^N(t) + (1-m)(r_1^N(t) + r_2^N(t) + r_3^N(t)) \\ &\approx \bar{u}(0) - \int_0^t \mu_1 \bar{u}(s)ds - \int_0^t N\beta \bar{u}(s)\bar{c}(s)ds - \int_0^t N\beta(1-\gamma)\bar{u}(s)\bar{c}(s)ds \\ &\quad + \int_0^t (1-m)(\mu_1 \bar{u}(s) + \mu_2(\bar{c}(s) + \bar{j}(s)))ds \\ c^N(t) &= c^N(0) - r_2^N(t) + r_4^N(t) + r_5^N(t) - r_6^N(t) + m(r_1^N(t) + r_2^N(t) + r_3^N(t)) \\ &\approx \bar{c}(0) - \int_0^t \mu_2 \bar{c}(s)ds + \int_0^t N\beta \bar{u}(s)\bar{c}(s)ds + \int_0^t N\beta(1-\gamma)\bar{u}(s)\bar{j}(s)ds \\ &\quad - \int_0^t \alpha \bar{c}(s)ds + \int_0^t m(\mu_1 \bar{u}(s) + \mu_2(\bar{c}(s) + \bar{j}(s)))ds \\ j^N(t) &= j^N(0) - r_3^N(t) + r_6^N(t) \\ &\approx \bar{j}(0) - \int_0^t \mu_2 \bar{j}(s)ds + \int_0^t \alpha \bar{c}(s)ds. \end{aligned} \tag{11}$$

Upon approximating $(u^N(t), c^N(t), j^N(t))$ in the left side by $(\bar{u}(t), \bar{c}(t), \bar{j}(t))$ and differentiating the above equations, we obtain the defining deterministic ordinary differential equations for $(\bar{u}, \bar{c}, \bar{j})$ given by

$$\begin{aligned} \frac{d\bar{u}(t)}{dt} &= -\mu_1 \bar{u}(t) - \beta N \bar{u}(t)\bar{c}(t) - \beta N(1-\gamma)\bar{u}(t)\bar{j}(t) \\ &\quad + (1-m)(\mu_1 \bar{u}(t) + \mu_2(\bar{c}(t) + \bar{j}(t))) \end{aligned}$$

$$\begin{aligned}\frac{d\bar{c}(t)}{dt} &= -\mu_2\bar{c}(t) + \beta N\bar{u}(t)\bar{c}(t) + \beta N(1-\gamma)\bar{u}(t)\bar{j}(t) - \alpha\bar{c}(t) \\ &\quad + m(\mu_1\bar{u}(t) + \mu_2(\bar{c}(t) + \bar{j}(t))) \\ \frac{d\bar{j}(t)}{dt} &= -\mu_2\bar{j}(t) + \alpha\bar{c}(t), \end{aligned} \tag{12}$$

with initial conditions $\bar{u}(0) = U_0/N$, $\bar{c}(0) = C_0/N$, and $\bar{j}(0) = J_0/N$.

To facilitate comparison with the MC model, we let $\bar{U}(t) = N\bar{u}(t)$, $\bar{C}(t) = N\bar{c}(t)$, and $\bar{J}(t) = N\bar{j}(t)$. Then, the system of ordinary differential equations which provides an approximation to averages over sample paths of $U((t), C(t), J(t))$ is described by

$$\begin{aligned}\frac{d\bar{U}(t)}{dt} &= (1-m)[\mu_1\bar{U}(t) + \mu_2(\bar{C}(t) + \bar{J}(t))] - \beta\bar{U}(t)[\bar{C}(t) + (1-\gamma)\bar{J}(t)] - \mu_1\bar{U}(t) \\ \frac{d\bar{C}(t)}{dt} &= m[\mu_1\bar{U}(t) + \mu_2(\bar{C}(t) + \bar{J}(t))] + \beta\bar{U}(t)[\bar{C}(t) + (1-\gamma)\bar{J}(t)] - (\alpha + \mu_2)\bar{C}(t) \\ \frac{d\bar{J}(t)}{dt} &= \alpha\bar{C}(t) - \mu_2\bar{J}(t), \end{aligned} \tag{13}$$

with initial conditions $\bar{U}(0) = U_0$, $\bar{C}(0) = C_0$, and $\bar{J}(0) = J_0$.

2.3. *Simulation results*

We carried out simulations to compare the results of the stochastic and deterministic models. The deterministic system was numerically solved using *ode*45 in Matlab. Both deterministic and stochastic results are generated using the same parameter values and initial conditions as given in Table 3. We used a stochastic simulation algorithm proposed by Gillespie[31] that is standard for discrete state continuous time MC models. The algorithm is as follows:

(1) **Initialize** the state of the system;
(2) For a given state of the system calculate the transition rates, λ_i, for $i = 1, \ldots, n$, where n is the total types of transitions;
(3) Calculate the sum of all transition rates $\lambda = \sum_{i=1}^{n} \lambda_i$;
(4) **Monte Carlo Step:** Simulate the time until the next transition, τ, by drawing from an exponential distribution with mean $1/\lambda$;
(5) **Monte Carlo Step:** Simulate the transition type by drawing from the discrete distribution with $P(transition = i) = \lambda_i/\lambda$. Generate an uniformly distributed random number r_2. For $0 < r_2 < \lambda_1/\lambda$ transition 1 is chosen, for $\lambda_1/\lambda < r_2 < (\lambda_1 + \lambda_2)/\lambda$ transition 2 is chosen, and so on;
(6) **Update** the new time $t = t + \tau$ and the new system state;
(7) **Iterate** Steps 2–6 until $t \geq t_{\text{stop}}$.

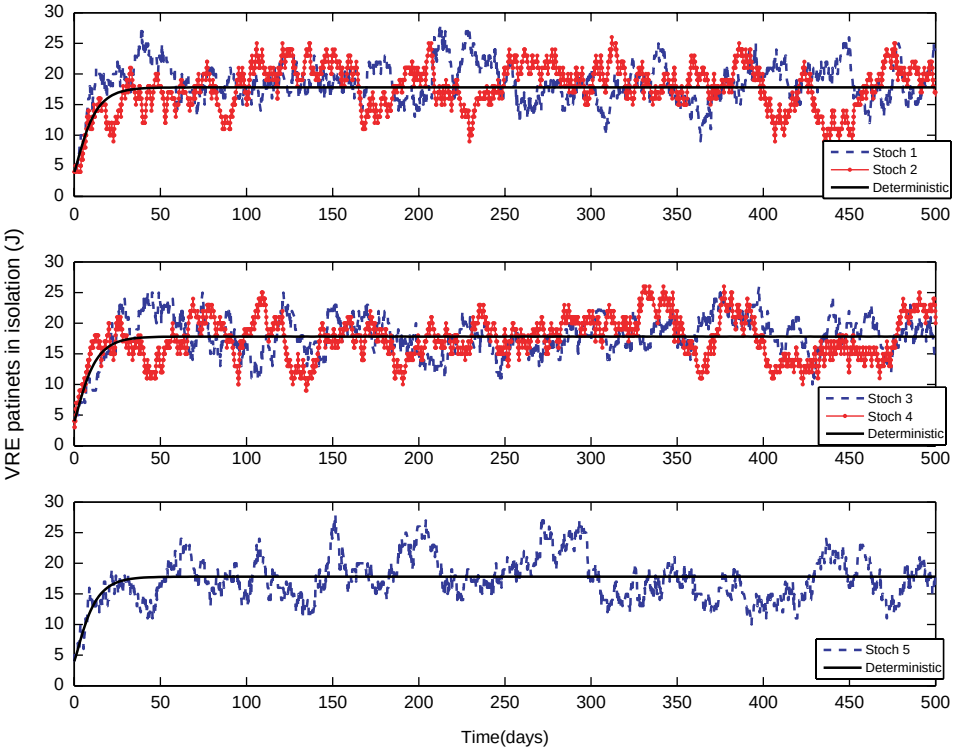
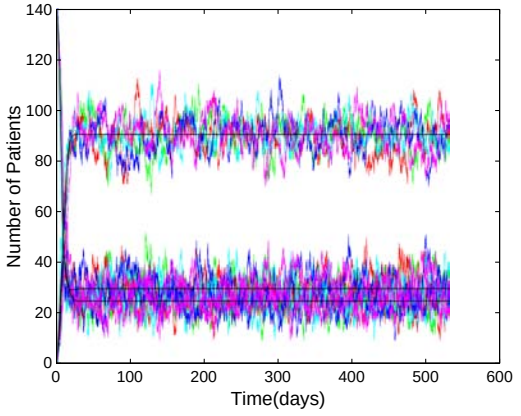
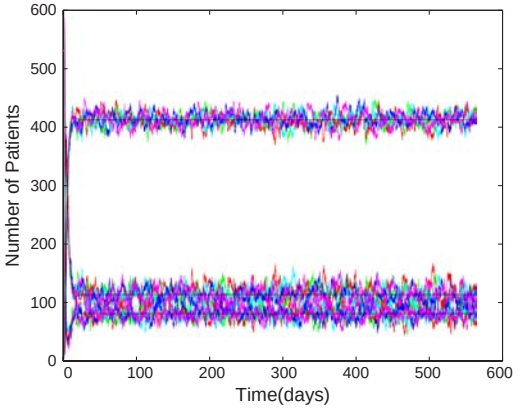


Fig. 2. Sample of 5 stochastic realizations of J in comparison to the numerical solution of this state for the deterministic model; $N = 37$ patients, $t_{\text{stop}} = 500$. Parameter values used are in Table 3.

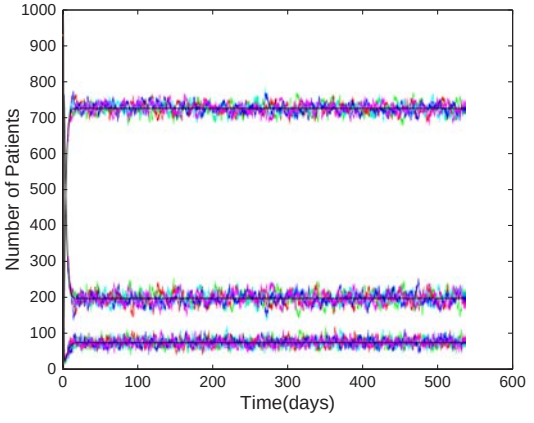
Figure 2 depicts a simulation of the state variable J of the stochastic model for a sample of 5 stochastic realizations ($N = 37$, $t_{\text{stop}} = 500$) plotted in comparison to the deterministic numerical solution for the J compartment (similar behavior is observed when plotting the other model compartments). Note that the stochastic realizations exhibit very large differences. However, when carrying out the simulations for larger values of N, the variation between the stochastic realizations decreases as the value of N increases and become quite close to the numerical solution of the deterministic model, as seen in Fig. 3. To quantitatively analyze how the variability of the stochastic realizations decreases as N increases, we calculated the coefficients of variation (CV) in the range $t \in [300, 400]$ using 100 stochastic realizations. The coefficients of variation (CV) (defined as the vector of normalized standard errors (see (38)) for (U, C, J) in the range $t \in [300, 400]$ using a sample of 100 stochastic realizations depicted in Fig. 3 for the different values of $N = 137,\ 537,\ 937,\ 2037$ are, (a) (0.064, 0.060, 0.027), (b) (0.036, 0.027, 0.008), (c) (0.031, 0.021, 0.006) and (d) (0.029, 0.014, 0.004) respectively. We note that the CV decrease as $N \to \infty$.



(a) $N = 137$

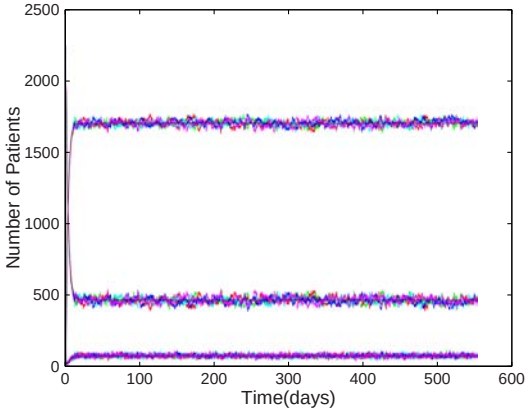


(b) $N = 537$



(c) $N = 937$

Fig. 3. Sample of 5 stochastic realizations for each compartment in comparison to the numerical solution of the deterministic model for $N = 137, 537, 937, 2037$, $t_{\text{stop}} = 500$.



(d) $N = 2037$

Fig. 3. (*Continued*)

3. VRE Surveillance Data

The motivating surveillance data is from an oncology unit, obtained from the VRE Infection Control database of the Department of Quality Improvement Support Service of Yale-New Haven Hospital. Data reports on the number of VRE cases occurred on admission (including patients transferred), the patients' length of stay, the daily number of patients in isolation due to VRE colonization, the compliance of swab culture administered on admission, and the healthcare worker contacts precautions compliance. Data collection occurred during the period of January 2005 to January 2007 with a mean number of 31 in-patients per day (with a total of 37 beds).

Ward protocol required rectal swabbing all patients on admission, and once a week (every Tuesday) for VRE surveillance. Compliance was not 100%, as the mean percentage of swab cultures taken on admitted patients per day was 77%. Swab-test results were usually returned 48 hours after admission. If a patient tested VRE positive, he/she was isolated. The isolation procedure consisted of contact precautions by the use of gowns, gloves, and the location of a patient in a single room or in a room with another patient who was also VRE positive. If a readmission patient had a positive VRE culture in the past, he/she did not get the rectal swab on admission but was isolated immediately. The isolation report was performed on weekdays (no weekends or holidays). The mean number of isolated VRE colonized patients per day was 9.39 (std = 2.90).

3.1. *Parameters estimated directly from the surveillance data*

Infection control measures were implemented in the form of healthcare worker hand-hygiene before and after contact with the patients by the use of gloves and gowns,

and washing the hands. In accordance with this measures, data collectors quarterly reported healthcare workers' hand-hygiene compliance before and after contact patients. For this study, we consider the reported mean of the healthcare with the worker before patient contact compliance as a better estimator for the parameter γ. In this case, $\gamma = 57.56\%$. In the oncology unit, VRE colonized patients had a mean length of stay of 13.15 days ($SE = 2.41 \times 10^{-2}$) compared with 6.27 ($SE = 3.0 \times 10^{-3}$) for the uncolonized patients. These means are statistically significant (p-value < 0.0001) supporting our model assumption of having different discharge rates. Hence, we take $1/\mu_1 = 6.27$ and $1/\mu_2 = 13.15$ giving $\mu_1 = 0.16$ and $\mu_2 = 0.08$.

In an attempt to estimate the fraction m of patients that are colonized on admission, we found inconsistencies in the reported prevalence of VRE on admission (the summaries of admitted patients did not match the actual data). In estimating the initial conditions (U_0, C_0, J_0) from the data reported on the first day of data collection (3 January 2005), only the number of VRE colonized patients in isolation were reported. Hence, the initial conditions for U_0 and C_0 classes cannot be easily estimated. Another parameter that is of interest and cannot be estimated directly from the data is the VRE transmission rate β. As a result, the fraction of patients that are colonized on admission, the initial conditions, and the transmission rate will be estimated using inverse problem methodology. In Table 3, we present the estimated and assumed values of the parameters and initial conditions taken as nominal values in inverse problem calculations.

4. Inverse Problem Methodology

We briefly summarize the statistical methodology for estimating parameters in dynamical systems such as (13) using observations of some of the states. These results are well established both theoretically and computationally and more details

Table 3. Parameters and initial conditions values (some values are assumed for optimization purposes).

Initial conditions	Oncology unit ($N = 37$)	Source	Units
$U(0)$	29	Assumed	Individuals
$C(0)$	4	Assumed	Individuals
$J(0)$	4	data	Individuals
Parameters	Oncology unit ($N = 37$)	Source	Units
Λ	$\mu_1 U(t) + \mu_2(C(t) + J(t))$	—	Individuals/day
m	0.04	Assumed	Dimensionless
β	0.001	Assumed	1/day
γ	0.58	data	Dimensionless
α	0.29	data	1/day
μ_1	0.16	data	1/day
μ_2	0.08	data	1/day

(including rigorous arguments with precise statements of underlying assumptions) can be found in numerous references.[12,18,32]

Let Y_j for $j = 1, \ldots, n$, be longitudinal data observations (which are random variables) corresponding to the experimental data for the observation process. In general, since Y_j is not assumed to be free of error (i.e., error in the data collection process), Y_j will not be exactly $f(t_j, \theta_0)$, the observed part of the true trajectory at time t_j. The statistical model for variability in observations is assumed given by

$$Y_j = f(t_j, \theta_0) + \mathcal{E}_j \quad \text{for } j = 1, \ldots, n, \tag{14}$$

in the case of absolute error in the measurements. Thus, we can envision experimental data as generally consisting of observations from a "perfect" model plus an error component, where θ_0 corresponds to the "true" parameter that generates the observations Y_j for $j = 1, \ldots, n$. We assume that the $\mathcal{E}_j$'s are generated from a generally unknown probability distribution P. They are assumed to satisfy the error assumptions.

(EA) *The random variables* $\mathcal{E}_j, j = 1, \ldots, n$, *are independent identically distributed with mean zero* (*i.e.*, $E(\mathcal{E}_j) = 0$) *and constant finite variance* (*i.e.*, $var(\mathcal{E}_j) = \sigma_0^2 < \infty$).

The observational process corresponding to the mathematical model (13) is denoted by

$$f(t_j, \theta_0) = J(t_j, \theta_0), \tag{15}$$

where the observation function $f(t_j, \theta)$ depends on the parameters θ in a nonlinear fashion.

4.1. *Ordinary least squares (OLS) estimation*

If the error distribution is unknown, an OLS optimization procedure is often employed. This method can be viewed as minimizing the distance between the data and the model where all observations are treated to be of equal importance. The OLS method defines "best" as when the norm square of the residuals is a minimum

$$\theta_{OLS} = \theta_{OLS}^n = \arg\min_{\theta \in \Theta} \sum_{j=1}^{n} [Y_j - f(t_j, \theta)]^2. \tag{16}$$

This corresponds to solving for θ in

$$\sum_{j=1}^{n} [Y_j - f(t_j, \theta)] \nabla f(t_j, \theta) = 0.$$

We do not know the distribution of the random variable θ_{OLS}, but under a number of regularity and sampling hypotheses, as $n \to \infty$, θ_{OLS} is approximately

distributed according to a multivariate normal distribution. More precisely, under the asymptotic theory[12,18,32] we have as $n \to \infty$ the approximation

$$\theta_{OLS} = \theta^n_{OLS} \sim \mathcal{N}_p(\theta_0, \Sigma^n_0), \tag{17}$$

where the covariance matrix Σ^n_0 is defined by

$$\Sigma^n_0 \equiv \sigma^2_0[n\Omega_0]^{-1}$$

with

$$\Omega_0 \equiv \lim_{n\to\infty} \frac{1}{n}\chi^n(\theta_0)^T\chi^n(\theta_0).$$

Here $\chi^n(\theta) = \{\chi_{jk}\}$ is the sensitivity matrix given by

$$\chi_{jk}(\theta) = \frac{\partial f(t_j, \theta)}{\partial \theta_k} \quad j = 1, \ldots, n \quad \text{and} \quad k = 1, \ldots, p.$$

The error variance σ^2_0 is approximated by

$$\hat{\sigma}^2_{OLS} = \frac{1}{n-p}\sum_{j=1}^{n}[y_j - f(t_j, \hat{\theta}_{OLS})]^2 \tag{18}$$

as the bias adjusted estimate for σ^2_0, where $\hat{\theta}_{OLS}$ is the realization of θ_{OLS} for a given realization $\{y_j\}$ of $\{Y_j\}$. The covariance matrix Σ^n_0 is approximated by

$$\hat{\Sigma}^n_{OLS} = \hat{\sigma}^2_{OLS}[\chi^T(\hat{\theta}_{OLS})\chi(\hat{\theta}_{OLS})]^{-1}. \tag{19}$$

Therefore, in practice one uses the approximation

$$\theta_{OLS} \sim \mathcal{N}_p(\theta_0, \Sigma^n_0) \approx \mathcal{N}_p(\hat{\theta}_{OLS}, \hat{\Sigma}^n_{OLS}). \tag{20}$$

Asymptotic standard errors for the parameter estimates are obtained by taking square roots of the diagonal elements of $\hat{\Sigma}^n_{OLS}$, i.e., $SE(\hat{\theta}_k) = \sqrt{(\hat{\Sigma}^n_{OLS})_{kk}}$, $k = 1, \ldots, p$. The sensitivity matrix can be calculated by solving the sensitivity equations

$$\frac{d}{dt}\frac{\partial x}{\partial \theta} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial g}{\partial \theta}, \tag{21}$$

where in our example (13), written as $\dot{x} = g(x(t), \theta)$, $\partial g/\partial x$ is a 3×3 matrix function and both $\partial x/\partial\theta$ and $\partial g/\partial\theta$ are $3 \times p$ matrix functions.

4.2. *Generalized least squares (GLS) estimation*

If the error distribution is unknown and we suspect that relative error is present in the measurement, then the assumption of constant variance of the error in the longitudinal data does not hold. In such cases, a generalized least square (GLS) optimization procedure should be employed. For this case, we need to formulate a

new statistical model to take into consideration the non-constant error variability. If we can assume that the size of the error depends linearly on the size of the observed quantity, the statistical model (i.e, relative error model) is given by

$$Y_j = f(t_j, \theta_0)(1 + \mathcal{E}_j) \quad \text{for } j = 1, \ldots, n, \tag{22}$$

where the $\mathcal{E}_j$ satisfy (EA). It follows that $Y_j \sim \mathcal{N}(f(t_j, \theta_0), \sigma_0^2 f^2(t_j, \theta_0))$. In this case, GLS can be viewed as minimizing the distance between the data and the model while taking into account a model dependency variance in the observations. The GLS method defines "best" estimator as θ_{GLS} obtained from solving

$$\sum_{j=1}^{n} f^{-2}(t_j, \theta_{GLS})[Y_j - f(t_j, \theta_{GLS})]\nabla f(t_j, \theta_{GLS}) = 0, \tag{23}$$

with the corresponding estimate $\hat{\theta}_{GLS}$ for a given realization $\{y_j\}$. From asymptotic theory[12,18] we find

$$\theta_{GLS} = \theta_{GLS}^n \sim \mathcal{N}_p(\theta_0, \Sigma_0^n) \tag{24}$$

where

$$\Sigma_0^n \approx \sigma_0^2[\chi^T(\theta_0)W(\theta_0)\chi(\theta_0)]^{-1}$$

with

$$\chi(\theta) = \begin{bmatrix} \frac{\partial f(t_1,\theta)}{\partial \theta_1} & \cdots & \frac{\partial f(t_1,\theta)}{\partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial f(t_n,\theta)}{\partial \theta_1} & \cdots & \frac{\partial f(t_n,\theta)}{\partial \theta_p} \end{bmatrix}$$

and $W^{-1}(\theta) = diag(f^2(t_1, \theta), \ldots, (f^2(t_n, \theta))$. Using the estimates, we have the covariance matrix approximation

$$\Sigma_0^n \approx \hat{\Sigma}_{GLS}^n = \hat{\sigma}_{GLS}^2[\chi^T(\hat{\theta}_{GLS})W(\hat{\theta}_{GLS})\chi(\hat{\theta}_{GLS})]^{-1} \tag{25}$$

and the error variance approximation

$$\hat{\sigma}_{GLS}^2 = \frac{1}{n-p}\sum_{j=1}^{n} \frac{1}{f^2(t_j, \hat{\theta}_{GLS})}[y_j - f(t_j, \hat{\theta}_{GLS})]^2. \tag{26}$$

Therefore, in practice, we use the approximation

$$\theta_{GLS} \sim \mathcal{N}_p(\theta_0, \Sigma_0^n) \approx \mathcal{N}_p(\hat{\theta}_{GLS}, \hat{\Sigma}_{GLS}^n). \tag{27}$$

We can also calculate the asymptotic standard errors for $\hat{\theta}_{GLS}$ by taking the square roots of the diagonal elements of the covariance matrix $\hat{\Sigma}_{GLS}^n$. Again the sensitivity matrix $\chi(\hat{\theta}_{GLS}) = \{\chi_{jk}\}$ can be calculated using the sensitivity equations in (21).

Typically, one does not attempt to solve (23) directly, but rather the estimate $\hat{\theta}_{GLS}$ for a given realization $\{y_j\}$ can be solved iteratively using the algorithm:

(1) Set $k = 0$. Estimate the initial $\hat{\theta}_{GLS}^{(k)}$ by using the OLS estimate with y_j in place of Y_j in (16);
(2) Form the weights $\hat{w}_j^k = f^{-2}(t_j, \hat{\theta}_{GLS}^{(k)})$;
(3) Find $\hat{\theta}_{GLS}^{(k+1)}$ by minimizing

$$J^k(\theta_{GLS}) = \sum_{j=1}^{n} \hat{w_j^k} |y_j - f(t_j, \theta_{GLS})|^2; \tag{28}$$

(4) Set $k = k + 1$ and return to 2. Terminate the process when two successive estimates for $\hat{\theta}_{GLS}$ are "close" to one another.

4.3. *Subset selection algorithm*

It is typical that in systems such as (13), some of the parameters (components of θ) are more readily estimated than others. The ability to accurately estimate a parameter is directly related to the sensitivity of the model output to a parameter. In order to identify the subset of parameters that has a high sensitivity to the mathematical model, we use the identifiability analysis recently developed.[17] The parameter selection or parameter identifiability algorithm consists of considering two criteria:

(1) Select the combinations of parameter vectors that have a full rank sensitivity matrix $\chi^n(\hat{\theta})$. The degree of sensitivity for the matrix is measured in the form of its *condition number* $\kappa(\chi^n(\hat{\theta}))$ defined below in (36);
(2) For each parameter vector selected in the first criterion, estimate its degree of uncertainty. Its degree of uncertainty is measured in the form of the *parameter selection score* $\upsilon(\hat{\theta})$ defined by (37).

The motivation behind the first criterion is as follows. If θ_0 is the true parameter, then $\Delta\theta = \theta - \theta_0$ denotes a local perturbation from θ_0. This gives rise to a local perturbation $\Delta y(t) = y(t, \theta) - y(t, \theta_0)$ in the model output. Then, by a first order Taylor approximation, we obtain the approximate relationship

$$\Delta y \approx \chi \Delta\theta. \tag{29}$$

A parameter vector is identifiable (locally) if the above equation can be solved uniquely for $\Delta\theta$. This is the case if $rank(\chi) = p$, or equivalently, if and only if the Fisher information matrix, $F = \chi^T(\hat{\theta})\chi(\hat{\theta})$ is nonsingular or

$$\det(\chi^T\chi) \neq 0. \tag{30}$$

The Fisher information matrix measures the amount of information that an observation process carries an unknown parameter θ. If near-singularity of F is present,

then the approximation of the covariance matrix and consequently the calculation of standard errors and confidence intervals for the corresponding estimated parameters are affected.

If one focuses on properties of the sensitivity matrix $\chi(\theta)$ rather than the Fisher information matrix, a singular value decomposition (SVD) of the sensitivity matrix plays a crucial role in uncertainty quantification. The SVD of the sensitivity matrix is denoted by

$$\chi(\theta) = U \begin{bmatrix} \Lambda \\ \mathbf{0} \end{bmatrix} V^T \tag{31}$$

where $U = [U_1\, U_2]$ and V are $n \times n$ and $p \times p$ orthogonal matrices, with U_1 containing the first p columns of U and U_2 containing the last $n - p$ columns. Λ is a $p \times p$ diagonal matrix defined as $\Lambda = diag(s_1, \ldots, s_p)$ with $s_1 \geq s_2 \geq \cdots \geq s_p \geq 0$, and $\mathbf{0}$ denotes an $(n - p) \times p$ matrix of zeros.

Suppose that $f(t, \theta)$ is well approximated for all $t = t_j$ by its linear Taylor expansion around θ_0 as

$$f(t, \theta) \approx f(t, \theta_0) + \frac{\partial f}{\partial \theta}(t, \theta_0)(\theta - \theta_0). \tag{32}$$

Then letting $f(\theta) = (f(t_1, \theta), \ldots, f(t_n, \theta))^T, Y = (Y_1, \ldots, Y_n)^T$ and $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_n)^T$, we have from (14)

$$Y - f(\theta) = -\chi(\theta_0)(\theta - \theta_0) + \mathcal{E}. \tag{33}$$

We can then define the estimator θ_{OLS} that minimizes $|Y - f(\theta)|^2$ and using the invariance property of the Euclidean norm (i.e., $|w|^2 = w^T w = w^T I w = w^T U U^T w = |U^T w|^2$) we have

$$\begin{aligned} |Y - f(\theta)|^2 &= |-\chi(\theta_0)(\theta - \theta_0) + \mathcal{E}|^2 \\ &= \left| U^T \left(-U \begin{bmatrix} \Lambda \\ \mathbf{0} \end{bmatrix} V^T(\theta - \theta_0) + \mathcal{E} \right) \right|^2 \\ &= \left| - \begin{bmatrix} \Lambda \\ \mathbf{0} \end{bmatrix} V^T(\theta - \theta_0) + \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \mathcal{E} \right|^2. \end{aligned} \tag{34}$$

Solving $|-\Lambda V^T(\theta - \theta_0) + U_1^T \mathcal{E}|^2 = 0$ for θ we have

$$\theta - \theta_0 = (\Lambda V^T)^{-1} U_1^T \mathcal{E}.$$

This implies

$$\begin{aligned} \hat{\theta}_{OLS} &= \theta_0 + V \Lambda^{-1} U_1^T \mathcal{E} \\ &= \theta_0 + \sum_{i=1}^{p} \frac{1}{s_i} v_i u_i^T \mathcal{E}, \end{aligned} \tag{35}$$

where u_i, v_i denote the ith column of U, V, respectively. Note that, if $s_i \to 0$, the estimator is particulary sensitive to $\mathcal{E}$.

If $\chi(\theta) \in \mathbb{R}^{n\times p}$ is a full rank sensitivity matrix (i.e., $rank(\chi(\theta)) = p$) its *condition number* κ is defined as the ratio of the largest to smallest singular value given by

$$\kappa(\chi(\theta)) = \frac{s_1}{s_p}. \tag{36}$$

which provides a degree of singularity due to perturbations and hence a criterion for parameter identifiability. If the columns of $\chi(\theta)$ are nearly dependent then (36) is large.

Motivation of the second criterion is the uncertainty in the parameters of a particular subset combination that can be quantified using the standard errors $SE(\theta)$. In order to compare the degree of variation from one parameter to another, the vector coefficient of variation $CV = SE(\theta)/\theta \in \mathbb{R}^p$ is used. The CV (defined in (38)) allows one to compare the parameters even if the parameter estimates are substantially different in units and scales. Hence, a second criterion can be established by considering the parameter *selection score*

$$\upsilon(\theta) = |CV(\theta)|, \tag{37}$$

where $|\cdot|$ is the norm in $\mathbb{R}^p$ and

$$CV_i(\theta) = \frac{SE(\hat{\theta}_i)}{\theta_i} = \frac{\sqrt{(\hat{\Sigma}^n)_{ii}}}{\theta_i}. \tag{38}$$

In (37), a value near zero indicates lower uncertainty possibilities in the estimation, while large values suggest a possibility of a wide uncertainty in at least some of the estimates. We note that of course these scores depend on the number p of parameters being estimated.

In general, the algorithm[17] that searches all possible parameter combinations and selects the ones satisfying criterias 1 and 2 is the following:

(1) **Combinatorial search.** For a fixed p, $1 \le p \le K$ (where K is total number of parameters and initial conditions that are candidates for estimation for our problem here $K = 8$), calculate the set

$$S_p = \{\theta = (q_1, \ldots, q_p) \in \mathbb{R}^p \,|\, q_k \in Q_K,\ q_k \neq q_l \text{ for all } k, l = 1, \ldots, p\}$$

where $Q_K = \{\alpha, \gamma, \mu_1, \mu_2, J_0, C_0, m, \beta\}$ and the set S_p collects all the parameter vectors obtained from the combinatorial search;

(2) **Full rank test.** Calculate the set of feasible parameters Θ_p as $\Theta_p = \{\theta \,|\, \theta \in S_p \subset \mathbb{R}^p, rank(\chi(\theta) = p)\}$.

Calculate the condition number defined by

$$\kappa(\chi(\theta)) = \frac{s_1}{s_p};$$

(3) **Standard error test**. For every $\theta \in \Theta_p$, calculate a vector of coefficients of variation $CV(\theta) \in \mathbb{R}^p$ by

$$CV_i = \frac{\sqrt{(\sum(\theta)_{ii}}}{\theta_i},$$

for $i = 1, \ldots, p$ and $\sum(\theta) = \sigma_0^2[\chi(\theta)^T\chi(\theta)]^{-1} \in \mathbb{R}^{p\times p}$ is the appropriate covariance matrix from the asymptotic error theory. Calculate the parameter selection score as $\upsilon(\theta) = |CV(\theta)|$.

5. Inverse Problem Results

5.1. *Optimization algorithm testing with synthetic data*

Before illustrating with the VRE surveillance data, we test and illustrate use of the optimization algorithm to investigate the convergence of the parameters estimates $\hat{\theta}$ to the known values θ_0. In order to do this, we construct a synthetic dataset $\{y_j\}$ for $j = 1, \ldots, n$, by adding variability to the corresponding model solution component $f(t_j, \theta_0) = J(t_j, \theta_0)$ in (13). The statistical model that captures the variability is taken as

$$y_j = f(t_j, \theta_0) + \sigma z_j \tag{39}$$

where z_j is a realization from a standard normal variable (i.e., $Z_j \sim \mathcal{N}(0, 1)$) and σ is the constant variability. The magnitude of σ determines how much noise is added. A low value indicates that the data points tend to be very close to the same value (the mean), while high values indicates that the data are "spread out" over a large range of values. Therefore, we can expect that 95% of the time, numbers generated from this distribution will fall in the interval $[-1.96\sigma, 1.96\sigma]$. To this end, we consider the standard error as one indication of the ability of the algorithm to estimate the parameters using the synthetic data set.

The OLS and GLS optimization were solved with MATLAB routine *lsqnonlin* for $n = 500$. Parameter upper bounds are taken as

$$(\alpha, \gamma, \mu_1, \mu_2, J_0, C_0, m, \beta) = (0.5, 1, 1, 1, N, N, 1, 1)$$

and lower bounds are set to zero. Note that the upper bound for α is 0.5 because the method for VRE detection is assumed to take at least 2 days. The model solutions $f(t_j, \theta_0) = J(t_j, \theta_0)$ are generated with initial conditions and parameter values θ_0 for the oncology unit as described in Table 3 (which are assumed to be the true values). By introducing variability levels such as $\sigma = 0$, $\sigma = 0.01$, $\sigma = 0.05$, and $\sigma = 0.1$ in the model solutions, the reliability of the algorithm and hence that of estimates are explored. Note that even though we are adding constant variability to the synthetic data, the GLS optimization algorithm is tested with this data. This is because, we wish to investigate how the noise affects the standard deviation and not how meaningful they are.

In Tables 6, 7, and 8 in the Appendix, we summarize the results for the inverse problems for $\theta = (J_0, C_0, \beta)$, $\theta = (J_0, C_0, m, \beta)$, and $\theta = (\alpha, J_0, C_0, m, \beta)$ using an OLS and a GLS optimization formulation. We note that, in these tables, in some cases, the resulting confidence intervals do not cover the true parameter θ_0. This is expected especially in the case of experimental data and models where no true value exists but can also occur in synthetic data examples where a true θ_0 does exist by construction. Moreover, in these particular examples with constant variance synthetic data, we can expect the GLS standard error values to perhaps be optimistic (there is a mismatch between the error model and the GLS asymptotic theory) as well as fail to cover the generating θ_0. Nonetheless, results indicate that both algorithms appear to be reliable for the estimation of the parameters since the estimated values are close to their true values. Finally, as depicted in Figs. 4, 5, and 6, we observe that as σ increases, the corresponding standard errors increase. This indicates that the reliability of both algorithms in estimating the parameters depend, of course, on the observational error in the data. Similar results were obtained for the other types of inverse problem formulations.

5.2. *Subset selection results using the oncology unit surveillance data*

To carry out the subset selection algorithm with the oncology unit surveillance data, we assumed nominal parameter values described in Table 3. Since we are interested in estimating the initial conditions, transmission rate, and the fraction of patients that are already colonized on admission, when $p = 4$ the only parameter

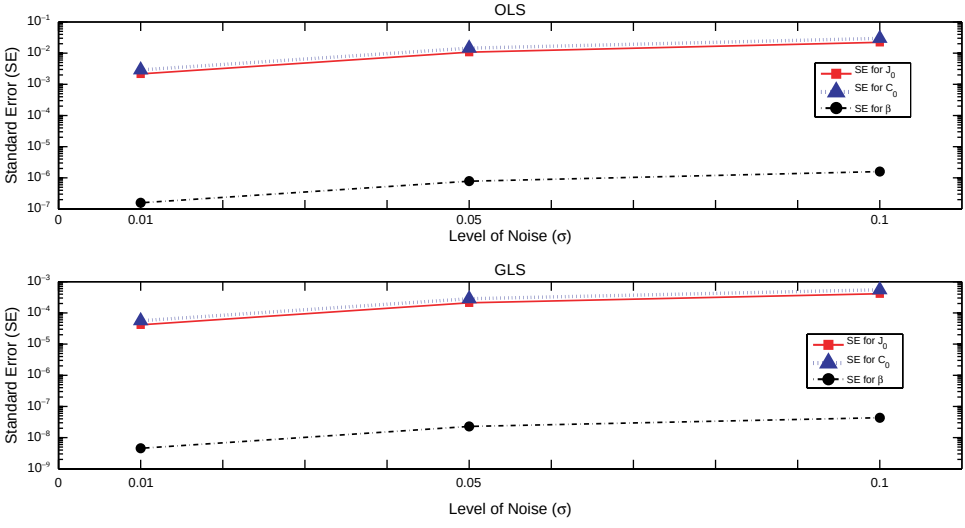


Fig. 4. *SE* vs. σ for OLS and GLS optimization for $\theta = (J_0, C_0, \beta)$ using synthetic data.

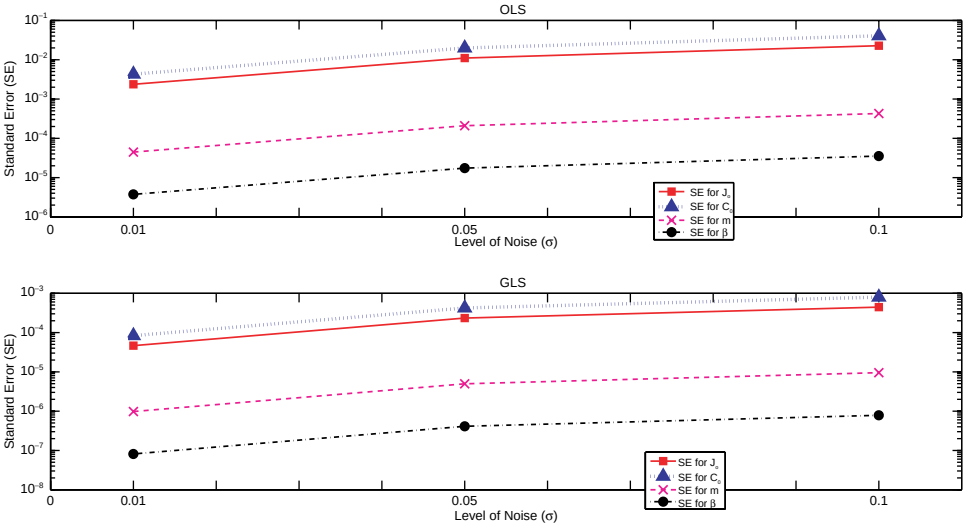


Fig. 5. *SE* vs. σ for OLS and GLS optimization for $\theta = (J_0, C_0, m, \beta)$ using synthetic data.

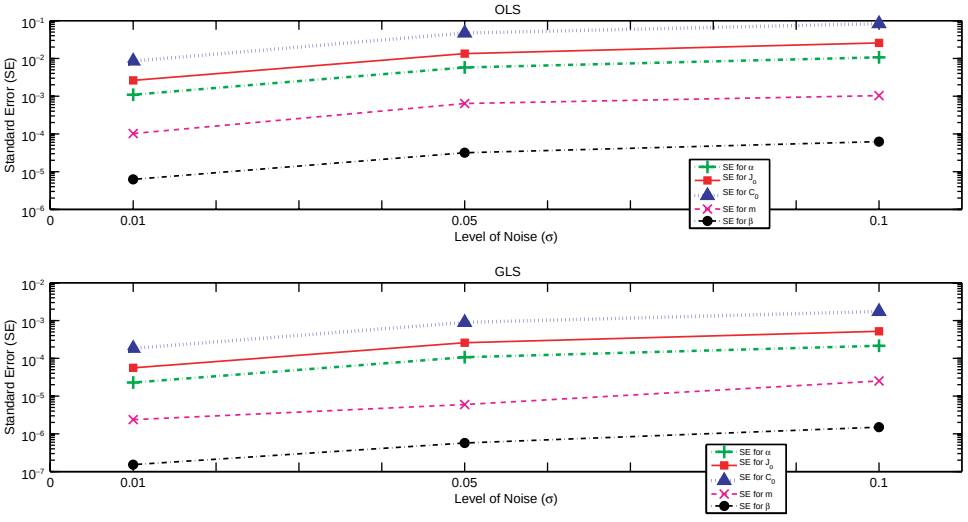


Fig. 6. *SE* vs. σ for OLS and GLS optimization for $\theta = (\alpha, J_0, C_0, m, \beta)$ using synthetic data.

combination considered is that of $\theta = (J_0, C_0, m, \beta)$. When $p = 1, 2, 3$, the only parameters considered are $\theta = (\beta)$, $\theta = (m, \beta)$, and $\theta = (J_0, C_0, \beta)$.

In Table 4, we present the resulting selection score $\upsilon(\theta)$ and condition number $k(\chi(\theta))$ for each subset of parameters when $p = 1, \ldots, 8$. Values that fall in the smallest selection score with the relative small condition number are considered the

Table 4. Subset parameter selected as a result of the selection algorithm for $p = 1, \ldots, 8$ using the oncology unit surveillance data with nominal parameter values described in Table 3 using the GLS optimization.

Parameter vector q	Selection score $\upsilon(q)$	Condition number $\kappa(\chi(q))$
(β)	1.975e−05	1.000e+00
(m, β)	2.358e−03	8.070e+02
(Jo, Co, β)	7.134e−03	8.236e+04
(Jo, Co, m, β)	1.815e−02	9.946e+04
$(\gamma, Jo, Co, m, \beta)$	1.539e−01	2.253e+05
$(\alpha, Jo, Co, m, \beta)$	1.597e−01	1.063e+06
$(\mu_1, Jo, Co, m, \beta)$	1.715e+01	1.308e+08
$(\mu_2, Jo, Co, m, \beta)$	6.123e+03	3.695e+05
$(\alpha, \mu_1, Jo, Co, m, \beta)$	6.225e+00	5.522e+06
$(\gamma, \mu_1, Jo, Co, m, \beta)$	1.741e+01	1.127e+08
$(\alpha, \mu_2, Jo, Co, m, \beta)$	6.315e+01	2.453e+05
$(\gamma, \mu_2, Jo, Co, m, \beta)$	8.472e+02	7.112e+05
$(\alpha, \gamma, Jo, Co, m, \beta)$	2.297e+03	2.852e+06
$(\mu_1, \mu_2, Jo, Co, m, \beta)$	7.475e+04	2.091e+05
$(\alpha, \gamma, \mu_1, Jo, Co, m, \beta)$	8.413e+02	2.074e+09
$(\alpha, \mu_1, \mu_2, Jo, Co, m, \beta)$	1.929e+03	3.760e+05
$(\alpha, \gamma, \mu_2, Jo, Co, m, \beta)$	3.447e+04	4.305e+06
$(\gamma, \mu_1, \mu_2, Jo, Co, m, \beta)$	4.589e+04	4.311e+07
$(\alpha, \gamma, \mu_1, \mu_2, Jo, Co, m, \beta)$	1.469e+04	1.967e+09

Table 5. Results of 4 inverse formulations solved with nominal values in Table 3 via GLS optimization for the oncology unit surveillance data.

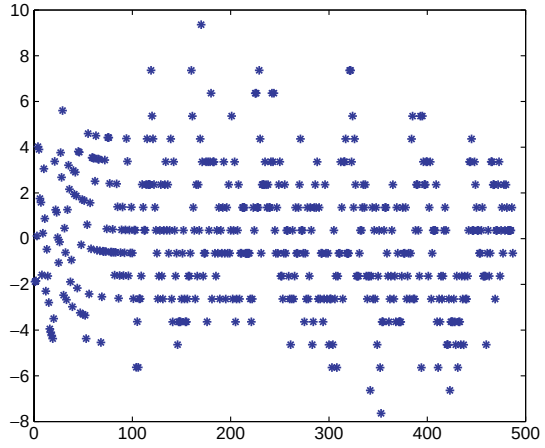
	γ	J_0	C_0	m	β
$\hat{\theta}$	6.392e−01	4.004e+00	1.092e+00	5.277e−02	4.770e−03
$SE(\hat{\theta})$	2.680e−02	1.811e−02	4.985e−02	6.007e−03	3.955e−04
$CV(\hat{\theta})$	4.192e−02	4.524e−03	4.567e−02	1.139e−01	8.291e−02
$\hat{\theta}$	–	**3.706e+00**	**1.966e+00**	**3.608e−02**	**4.865e−03**
$SE(\hat{\theta})$	–	1.499e−02	1.560e−02	5.616e−04	1.675e−05
$CV(\hat{\theta})$	–	4.044e−03	7.934e−03	1.556e−02	3.443e−03
$\hat{\theta}$	–	3.706e+00	1.966e+00	–	4.865e−03
$SE(\hat{\theta})$	–	1.419e−02	1.184e−02	–	9.945e−08
$CV(\hat{\theta})$	–	3.829e−03	6.020e−03	–	2.044e−05
$\hat{\theta}$	–	–	–	4.070e−02	4.725e−03
$SE(\hat{\theta})$	–	–	–	9.290e−05	2.802e−06
$CV(\hat{\theta})$	–	–	–	2.282e−03	5.931e−04

most feasible subset of parameters. Results indicate that the subsets of parameters $\theta = (J_0, C_0, m, \beta)$ have small condition numbers and relatively small selection scores indicating that these subsets might provide low uncertainty in the parameter estimates. In Table 5, we summarize the results of 4 inverse problems corresponding to the subsets with the lowest selection scores and small condition numbers.

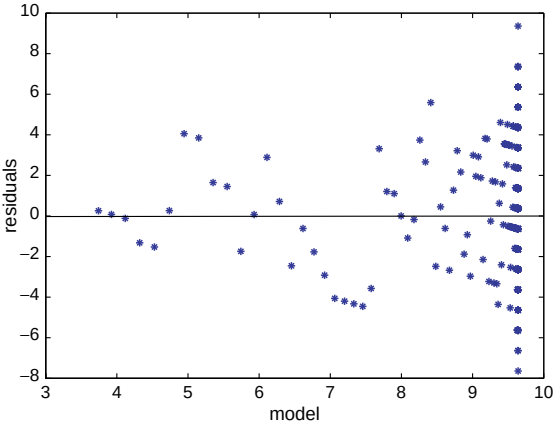
These subsets of parameters are:

$$\theta = (\gamma, J_0, C_0, m, \beta)$$
$$\theta = (J_0, C_0, m, \beta)$$
$$\theta = (J_0, C_0, \beta)$$
$$\theta = (m, \beta).$$

We analyze the results using the coefficient of variation (CV) by considering the effect that the inclusion or exclusion of parameters has on the vector $\theta = (J_0, C_0, m, \beta)$. In this subset, the standard errors for J_0 is about 0.4% of

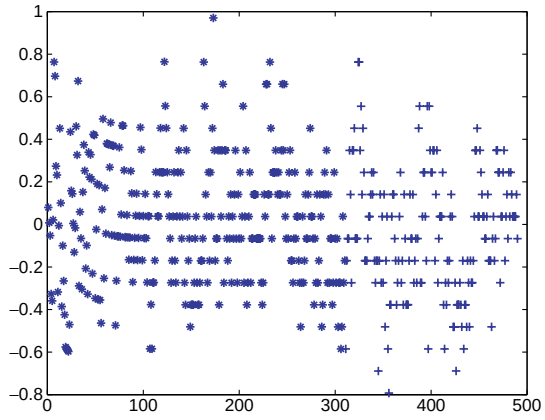


(a) OLS: Residuals vs. Time

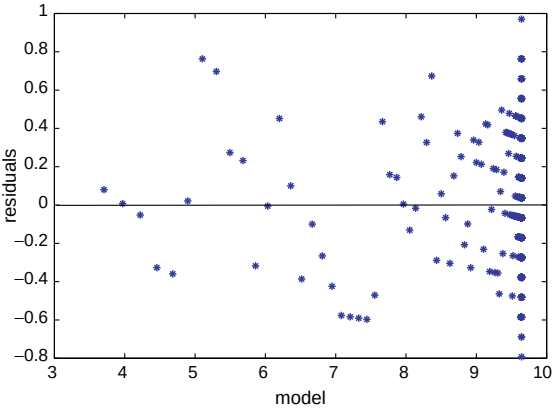


(b) OLS: Residuals vs. Model

Fig. 7. Residual analysis for the OLS and GLS optimization for $\theta = (J_0, C_0, m, \beta)$ using the oncology unit surveillance data. Note the difference in scales of axis in (a), (b) versus (c), (d).



(c) GLS: Residuals/Model vs. Time



(d) GLS: Residuals/Model vs. Model

Fig. 7. (*Continued*)

the estimate, for C_0 it is about 0.8% of the estimate, for m it is about 1.6% of the estimate, and for β it is 0.3% of the estimate. When including γ (i.e., $\theta = (\gamma, J_0, C_0, m, \beta)$), the CV increases for almost all parameters. On the other hand, when m is dropped or when the initial conditions are dropped, there is a reduction in the CV. Since this reduction is not significant, we can conclude that the subset $\theta = (J_0, C_0, m, \beta)$ is a good choice to be estimated from the oncology surveillance data since it provides estimates with low uncertainty.

Comparison of residual plots (details on the use of residual plots in such problems are readily available[12]) for all subsets of parameters combinations suggested that the assumptions of the relative error statistical model (22) corresponding to the GLS procedure are most suitable. In particular, the residual analysis for

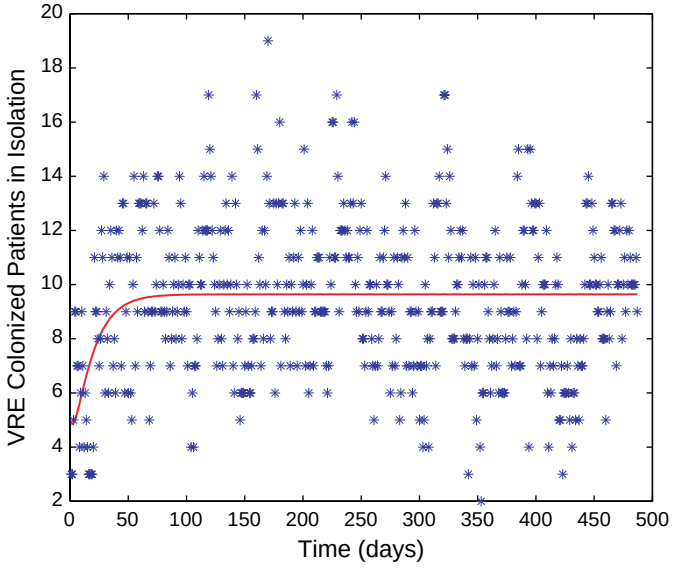


Fig. 8. Best fit model solutions to oncology unit surveillance data via GLS optimization, $(\hat{J}_0, \hat{C}_0, \hat{\mu_2}, \hat{\beta}) = (4, 2, 0.04, 0.0049)$.

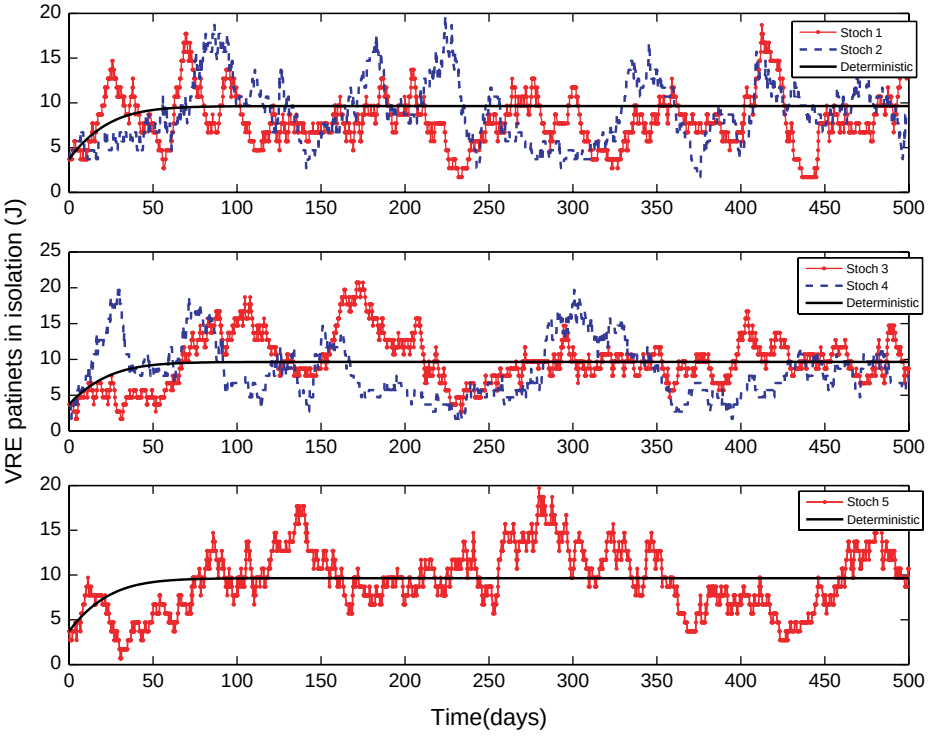


Fig. 9. Sample of 5 stochastic realizations in comparison to the numerical solution of the deterministic model using estimated parameters in Table 5; $N = 37$ patients, $t_{\text{stop}} = 500$.

estimating $\theta = (J_0, C_0, m, \beta)$ using OLS vs. GLS is presented in Fig. 7. Note the similarity between the plots of Figs. 7(a) and 7(c) which strongly supports the assumption of longitudinal independence. The OLS residual plot (b) in Fig. 7 reveals a slight fan shaped error structure which indicates the constant variance (absolute error) assumption is suspect. When GLS optimization is used instead, the residual plot (d) in Fig. 7 (note the difference in scales on the vertical axes of (b) and (d)) reveals a more random error structure, suggesting that the GLS procedure is most appropriate. A best fit of the model solution to the oncology surveillance data is plotted in Fig. 8. Also in Fig. 9, we plot 5 stochastic realizations with the original stochastic model using the best fit estimated parameters in Table 5 in comparison to the numerical solution of the deterministic model. We note that this is not a particularly encouraging fit of models to the experimental data, perhaps suggesting unaccounted for modeling error which is discussed below and is addressed in a technical report[19] and a forthcoming manuscript.[33]

6. Concluding Remarks

Over the past decade, efforts to connect models to data in the context of disease dynamics have grown dramatically. While many of these efforts have been carried out in the context of deterministic epidemic models,[34] as we have already noted, a significant number[8–11] involve stochastic models in an MLE Bayesian framework. It is not only the case that the use of stochastic Markov Chain models is often most appropriate, moreover the use of stochastic processes in epidemiology has had a long and distinguished history[35] going back to 1766.

The introduction of a methodology for parameter estimation within the context of a typical MC stochastic model through the use of a limit theory due to Kurtz provides a simple rigorous approximation approach for solution to an inverse problem of interest to epidemiologists. We have illustrated this approach with an example of nosocomial infections in hospital occupancy units. Since the number of beds in a typical hospital unit is small, it is natural to consider an integer valued stochastic model. Estimating epidemiological parameters in such a stochastic model can be a difficult task, particularly when the data is quite limited. The alternative approach involving the estimation of parameters from a corresponding deterministic approximation to the MC stochastic model, based on large sample size averages over sample paths, provides a reasonable first step. Once a deterministic approximation is obtained, one can readily apply standard as well as recently developed parameter estimation methods for deterministic systems which not only provide the parameter estimates but also the corresponding measures of uncertainty for the estimates.

Our presentation here amply illustrates the effective use of the proposed methodology. In fact, it permits one to evaluate the usefulness of the model formulated above and conclude that it is not adequate to describe the VRE hospital data

available to us. In revisiting the hospital protocols and carefully considering trends and patterns in the data, we found strong support for models with delays (there is a two day delay in reporting testing results) as well as for models with jump discrete events (infective individuals are all removed and put in isolation on the day each week that testing results are returned). In our subsequent efforts, we have developed models with delays as well as discontinuities for jump phenomena. These are reported elsewhere.[19,33]

Acknowledgments

This research was supported in part by the Alfred P. Sloan Foundation Graduate Scholarship (sponsored by the Sloan Foundation and NACME), the Louis Stokes Alliance for Minority Participation (LSAMP) Fellowship (sponsored by NSF/WAESO), and the More Graduate Education at Mountain State Alliance (MGE@MSA) Scholarship. It was also supported in part by the US. Air Force Office of Scientific Research under grant AFOSR-FA9550-09-1-0226 and in part by the National Institute of Allergy and Infectious Disease under grant NIAID 9R01AI071915-05. The authors are grateful to Dr. Carlos Torres-Viera for providing the sample data used to illustrate the methodology.

References

1. Brauer F, Castillo-Chavez C, *Mathematical Models in Population Biology and Epidemiology*, Texts in Applied Mathematics, Vol. 40, Springer-Verlag, New York, 2001.
2. Brauer F, van den Driessche P, Wu J, Allen L, *Mathematical Epidemiology,* Springer-Verlag, New York, 2008.
3. Byers KE, Anglim AM, Anneski CJ, Farr BM, Duration of colonization with vancomycin-resistant enterococcus, *Inf Control Hosp Epidemiol* **27**:271–278, 2002.
4. Castillo-Chavez C, Chow KC, Wang X, A mathematical model of nosocomial infection and antibiotic resistance: evaluating the efficacy of antimicrobial cycling programs and patient isolation on dual resistance, Mathematical and Theoretical Biology Institute Technical Reports, MTBI-04-05M, Arizonia State University, 2007.
5. Chowell G, Fenimore PW, Castillo-Garsow MA, Castillo-Chavez C, SARS outbreaks in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism, *J Theor Biol* **224**:1–8, 2003.
6. Grenfell B, Dobson A, *Ecology of Infectious Diseases in Natural Populations*, Cambridge University Press, Cambridge, 1995.
7. Montecalvo MA, Jarvis WR, Uman J, Shay DK, Petrullo C, Rodney K, Gedris C, Horowitz HW, Wormser GP, Infection-control measures reduce transmission of vancomycin-resistant enterococci in an endemic setting, *Annals Internal Med* **131**:269–272, 1999.
8. Andersson H, Britton T, *Stochastic Epidemic Models and Their Statisticsal Analysis*, Lec. Notes in Statistics, Vol. 51, Springer-Verlag, New York, 2000.
9. Becker NG, Britton T, Statistical studies of infectious disease incidence, *J Roy Statist Soc B* **61**:287–307, 1999.
10. Longini IM, Jr, Koopman JS, Household and community transmission parameters from final distributions of infections in households, *Biometrics* **38**:115–126, 1982.

11. Yang Y, Halloran ME, Longinis IM, A Bayesian model for evaluating influenza antiviral efficacy in household studies with asymptomatic infections, *Biostatistics* **10**:390–403, 2009.
12. Banks HT, Davidian M, Samuels JR, Jr, Sutton KL, An inverse problem statistical methodology summary, Center for Research in Scientic Computatione Technical Report, CRSC-TR08-01, 2008, North Carolina State University, in *Mathematical Statistical Estimation Approaches in Epidemiology*, Chowell G *et al.* (eds.), Springer, New York, 2009, pp. 249–302.
13. Banks HT, Grove SL, Hu S, Ma Y, A hierarchical Bayesian approach for parameter estimation in HIV models, CRSC-TR05-19, April, 2005; *Inverse Problems* **21**: 1803–1822, 2005.
14. Bai P, Banks HT, Dediu S, Govan AY, Last M, Lloyd A, Nguyen HK, Olufsen MS, Rempala G, Slenning BD, Stochastic and deterministic models for agricultural production networks, *Math Biosci Eng* **4**:373–402, 2007.
15. Ethier SN, Kurtz TG, *Markov Processes: Characterization and Convergence*, J. Wiley and Sons, New York, 1986.
16. Kurtz TG, Solutions of ordinary differential equations as limits of pure jump Markov processes, *J Applied Prob* **7**:49–58, 1970.
17. Cintron-Arias A, Banks HT, Capaldi A, Lloyd AL, A sensitivity matrix based methodology for inverse problem formulation, *J Inverse and Ill-Posed Problems* **15**:545–564, 2009.
18. Davidian M, Giltinan DM, *Nonlinear Models for Repeated Measurement Data*, Monographs on Statistics and Applied Probability, Vol. 62, Chapman and Hall/CRC Press, Boca Raton, FL, 1995.
19. Ortiz AR, Banks HT, Castillo-Chavez C, Chowell G, Torres-Viera C, Wang X, Modeling the transmission of Vancomycin-resistant enterococcus (VRE) in hospitals: a case study, Center for Research in Scientic Computation Technical Report, CRSC-TR10-05, February, 2010, North Carolina State University, Raleigh.
20. Austin DJ, Bonten MJM, Weinstein RA, Slaughter S and Anderson RM, Vancomycin-resistant enterococci in intensive-care hospital settings: transmission dynamics, persistence, and the impact of infection control programs, *Proc Natl Acad Sci USA* **96**:6908–6913, 1999.
21. Austin DJ, Anderson RM, Transmission dynamic of epidemic methicillin-resistant staphylococcus aureus and vancomycin-resistant enterococci in England and Wales, *J Infect Diseases* **179**:883–891, 1999.
22. Austin DJ, Kakehashi M, Anderson RM, The transmission dynamic of antibiotic-resistant bacteria: The relationship between resistance in commensal organisms and antibiotic consumption, *Proc Biol Sci* **264**:1629–1638, 1997.
23. Bonten MJM, Willems R, Weinstein RA, Vancomycin-Resistant Enterococci: Why are they here, and where do they come from?, *The Lancet Infectious Diseases* **1**(5):314–325, 2001.
24. Cooper BS, Medley GF, Scott GM, Preliminary analysis of the transmission dynamics of nosocomial infections: Stochastic and management effects, *J Hosp Infec* **43**:131–147, 1999.
25. D'Agata EMC, Horn MA, Webb GF, The impact of persistent gastrointestinal colonization on the transmission dynamics of vancomycin-resitant enterococci, *J Infec Diseases* **185**:766–773, 2002.
26. Lipsitch M, Bergstrom CT, Levin BR, The epidemiology of antibiotic resistance in hospital paradoxes and prescirptions, *Proc Natl Acad Sci USA* **97**:1938–1943, 2000.

27. Song JY, Cheong HJ, Jo YM, Choi WS, Noh JY, Heo JY, Kim WJ, Vancomycin-resistant Enterococcus colonization before admission to the intensive care unit: A clinico- epidemiologic analysis, *Amer J Infec Control* **37**:734–740, 2009.
28. Webb GF, D'Agata EMC, Magal P, Ruan S, A model of antibiotic-resistant bacterial epidemics in hospitals, *Proc Natl Acad Sci USA* **102**:13343–13348, 2005.
29. Allen L, *An Introduction to Stochastic Processes with Applications to Biology*, Pearson Education-Prentice Hall, New Jersey, 2003.
30. Renshaw E, *Modeling Biological Populations in Space and Time*, Cambridge Studies in Mathematical Biology (No. 11), Cambridge, 1991.
31. Gillespie DT, A General method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J Comput. Physics* **22**:403–434, 1976.
32. Seber GAF, Wild CJ, *Nonlinear Regression*, J. Wiley and Sons, New York, 1989.
33. Ortiz AR, Banks HT, Castillo-Chavez C, Chowell G, Torres-Viera C, Wang X, A discrete events delay differential system model for transmission of Vancomycin-resistant enterococcus (VRE) in hospitals, in preparation.
34. Chowell G, Hyman JM, Bettencourt LMA, Castillo-Chavez C (eds.), *Mathematical and Statistical Estimation Approaches in Epidemiology*, Springer, New York, 2009.
35. Dietz K, Heesterbeek JAP, Daniel Bernoullis epidemiological model revisited, *Math Biosci* **180**:1–21, 2002.

Appendix

Table 6. OLS and GLS optimization algorithm testing for $\theta = (J_0, C_0, \beta)$ using synthetic data. The model was fit to the synthetic data with levels of noise: $\sigma = 0, 0.01, 0.05$, and 0.1.

	σ	J_0	C_0	β
$\hat{\theta}^{OLS}$	0	4.000e+00	4.000e+00	1.000e−03
$SE(\hat{\theta}^{OLS})$	0	2.301e−13	3.097e−13	1.691e−17
$\hat{\theta}^{OLS}$	0.01	4.007e+00	3.998e+00	1.006e−03
$SE(\hat{\theta}^{OLS})$	0.01	2.162e−03	2.907e−03	1.584e−07
$\hat{\theta}^{OLS}$	0.05	4.022e+00	3.995e+00	1.032e−03
$SE(\hat{\theta}^{OLS})$	0.05	1.077e−02	1.444e−02	7.793e−07
$\hat{\theta}^{OLS}$	0.1	4.074e+00	3.954e+00	1.063e−03
$SE(\hat{\theta}^{OLS})$	0.1	2.222e−02	2.971e−02	1.585e−06
$\hat{\theta}^{GLS}$	0	4.000e+00	4.000e+00	1.000e−03
$SE(\hat{\theta}^{GLS})$	0	3.956e−15	5.346e−15	4.358e−19
$\hat{\theta}^{GLS}$	0.01	4.002e+00	4.000e+00	1.006e−03
$SE(\hat{\theta}^{GLS})$	0.01	4.150e−05	5.604e−05	4.553e−09
$\hat{\theta}^{GLS}$	0.05	4.040e+00	3.973e+00	1.032e−03
$SE(\hat{\theta}^{GLS})$	0.05	2.112e−04	2.847e−04	2.290e−08
$\hat{\theta}^{GLS}$	0.1	4.016e+00	4.015e+00	1.067e−03
$SE(\hat{\theta}^{GLS})$	0.1	4.119e−04	5.523e−04	4.343e−08

Table 7. OLS and GLS optimization algorithm testing for $\theta = (J_0, C_0, m, \beta)$ using synthetic data. The model was fit to the synthetic data with levels of noise: $\sigma = 0, 0.01, 0.05$, and 0.1.

	σ	J_0	C_0	m	β
$\hat{\theta}^{OLS}$	0	4.000e+00	4.000e+00	4.000e−02	1.000e−03
$SE(\hat{\theta}^{OLS})$	0	8.620e−12	1.557e−11	1.611e−13	1.352e−14
$\hat{\theta}^{OLS}$	0.01	4.004e+00	4.008e+00	4.013e−02	9.955e−04
$SE(\hat{\theta}^{OLS})$	0.01	2.372e−03	4.287e−03	4.457e−05	3.735e−06
$\hat{\theta}^{OLS}$	0.05	4.029e+00	3.992e+00	4.032e−02	1.004e−03
$SE(\hat{\theta}^{OLS})$	0.05	1.102e−02	1.990e−02	2.091e−04	1.741e−05
$\hat{\theta}^{OLS}$	0.1	4.074e+00	3.945e+00	3.987e−02	1.074e−03
$SE(\hat{\theta}^{OLS})$	0.1	2.271e−02	4.067e−02	4.273e−04	3.529e−05
$\hat{\theta}^{GLS}$	0	4.000e+00	4.000e+00	4.000e−02	1.000e−03
$SE(\hat{\theta}^{GLS})$	0	1.496e−13	2.696e−13	3.145e−15	2.626e−16
$\hat{\theta}^{GLS}$	0.01	4.003e+00	4.001e+00	4.003e−02	1.004e−03
$SE(\hat{\theta}^{GLS})$	0.01	4.636e−05	8.350e−05	9.762e−07	8.137e−08
$\hat{\theta}^{GLS}$	0.05	4.009e+00	4.013e+00	4.022e−02	1.014e−03
$SE(\hat{\theta}^{GLS})$	0.05	2.343e−04	4.214e−04	4.971e−06	4.118e−07
$\hat{\theta}^{GLS}$	0.1	4.050e+00	4.011e+00	4.046e−02	1.025e−03
$SE(\hat{\theta}^{GLS})$	0.1	4.434e−04	7.976e−04	9.545e−06	7.845e−07

Table 8. OLS and GLS optimization algorithm testing for $\theta = (\alpha, J_0, C_0, m, \beta)$ using synthetic data. The model was fit to the synthetic data with levels of noise: $\sigma = 0, 0.01, 0.05$, and 0.1.

	σ	α	J_0	C_0	m	β
$\hat{\theta}^{OLS}$	0	2.890e−01	4.003e+00	4.136e+00	4.451e−02	1.856e−03
$SE(\hat{\theta}^{OLS})$	0	2.145e−04	5.126e−04	1.674e−03	2.009e−05	1.220e−06
$\hat{\theta}^{OLS}$	0.01	2.895e−01	4.010e+00	4.120e+00	4.454e−02	1.872e−03
$SE(\hat{\theta}^{OLS})$	0.01	1.094e−03	2.620e−03	8.517e−03	1.025e−04	6.264e−06
$\hat{\theta}^{OLS}$	0.05	2.811e−01	4.023e+00	4.269e+00	5.080e−02	1.670e−03
$SE(\hat{\theta}^{OLS})$	0.05	5.755e−03	1.337e−02	4.696e−02	6.381e−04	3.168e−05
$\hat{\theta}^{OLS}$	0.1	2.899e−01	4.051e+00	4.109e+00	4.541e−02	1.880e−03
$SE(\hat{\theta}^{OLS})$	0.1	1.071e−02	2.572e−02	8.329e−02	1.033e−03	6.263e−05
$\hat{\theta}^{GLS}$	0	2.901e−01	4.000e+00	4.095e+00	4.325e−02	1.538e−03
$SE(\hat{\theta}^{GLS})$	0	3.606e−06	8.920e−06	2.854e−05	3.686e−07	2.277e−08
$\hat{\theta}^{GLS}$	0.01	2.894e−01	4.009e+00	4.130e+00	4.300e−02	1.869e−03
$SE(\hat{\theta}^{GLS})$	0.01	2.282e−05	5.581e−05	1.836e−04	2.373e−06	1.525e−07
$\hat{\theta}^{GLS}$	0.05	2.787e−01	4.030e+00	4.348e+00	2.360e−02	1.860e−03
$SE(\hat{\theta}^{GLS})$	0.05	1.068e−04	2.581e−04	8.901e−04	5.955e−06	5.713e−07
$\hat{\theta}^{GLS}$	0.1	2.900e−01	4.035e+00	4.189e+00	4.739e−02	1.882e−03
$SE(\hat{\theta}^{GLS})$	0.1	2.147e−04	5.191e−04	1.740e−03	2.506e−05	1.502e−06