

PARAMETER ESTIMATION IN RANDOM
DIFFERENTIAL EQUATION MODELS

Banks H.T., Joyner M.L.

Abstract We consider two distinct techniques for estimating random parameters in random differential equation (RDE) models. In one approach, the solution to a RDE is represented by a collection of solution trajectories in the form of sample deterministic equations. In a second approach we employ pointwise equivalent stochastic differential equation (SDE) representations for certain RDEs. Each of the approaches is tested using deterministic model comparison techniques for a logistic growth model which is viewed as a special case of a more general Bernoulli growth model. We demonstrate efficacy of the preferred method with experimental data using algae growth model comparisons.

Key words: parameter estimation, random differential equations, stochastic differential equation equivalents, model comparison techniques

AMS Mathematics Subject Classification: 34K29, 34K50, 65L09, 37H10, 49N45

1 Introduction

In this paper, we examine techniques for estimating random variable parameters in random differential equation (RDE) models. Our ultimate research effort lies in the development of model comparison techniques for RDE models which are presented in a separate article [5]. However, in that effort, it is necessary to estimate optimal parameters for use in the test statistic. Overall, the theory [4, 12] for RDE is much less advanced than that for stochastic differential equations (SDE). While the questions of existence and uniqueness of solutions are without question important, for this presentation we simply assume that the RDE we investigate have a unique solution, and focus on discussion of the equation for the probability density function of the solution. Due in part to their wide applicability [7, 11, 13], RDE have enjoyed considerable research efforts on computational methods in the past decade. Widely used approaches include Monte Carlo methods, stochastic Galerkin methods and probabilistic collocation methods (also called stochastic collocation methods) [9, 8, 14]. Specifically, both Monte Carlo methods and probabilistic collocation methods seek to solve deterministic realizations of the given RDE (and thus both methods were developed in the spirit of the sample function approach). The difference between these two methods rests primarily in the manner in which one chooses the "sampling" points. Monte Carlo methods are based on large sampling of the distribution of random input variables while probabilistic collocation methods are based on quadrature rules (or sparse quadrature rules in high-dimensional space). Stochastic Galerkin methods are based

on (generalized) polynomial chaos expansions, which express the unknown stochastic process by a convergent series of (global) orthogonal polynomials in terms of random input parameters. Interested readers can refer to [14] and the references therein for details.

Some difficulties which have arisen in dealing with RDE may be due in part to the inability to accurately estimate parameters in the models. Here we compare two different techniques for estimating these random variable parameters and determining the accuracy of each method. Since the model selection criteria we have developed [5] extends the techniques for deterministic systems, we only consider parameter estimation techniques which are also extensions of methods developed for estimation of parameters in deterministic systems.

2 A Brief Overview of Random Differential Equations

A general random ordinary differential equation (RDE) containing random parameter values can be written as

$$\frac{d\mathbf{x}}{dt} = g(t, \mathbf{x}, \mathbf{Q}), \quad \mathbf{x}(0) = \mathbf{x}_0 \quad (1)$$

where \mathbf{Q} is a m -dimensional random vector. For example, consider the logistic deterministic model given by

$$\frac{dx}{dt} = rx(t) \left(1 - \frac{x(t)}{\kappa} \right) \quad (2)$$

where r is the growth rate and κ is the limiting capacity. In this deterministic model, both r and κ are assumed to be constant parameter values. One may formulate a logistic RDE model by instead assuming that one or both of these parameter values are random variables which behave according to some known (or to be determined) distribution. For example, if $R \sim \mathcal{N}(\mu_R, \sigma_R^2)$ is a random variable parameter for the growth rate and one assumes the limiting capacity κ is a constant, then

$$\frac{dx(t; Q)}{dt} = Rx(t; Q) \left(1 - \frac{x(t; Q)}{\kappa} \right) \quad (3)$$

is a RDE with random variable parameter $Q = R$. If instead we let $R \sim \mathcal{N}(\mu_R, \sigma_R^2)$ and $K \sim \mathcal{N}(\mu_K, \sigma_K^2)$, then

$$\frac{dx(t; \mathbf{Q})}{dt} = Rx(t; \mathbf{Q}) \left(1 - \frac{x(t; \mathbf{Q})}{K} \right) \quad (4)$$

is a RDE with random variable parameter $\mathbf{Q} = [R, K]$.

Regardless of the type of mathematical model, parameter estimation is a vital step in the development of the model. The validation of a mathematical model with empirical data allows one to use the model to gain insights into the processes inherent in the system as well as investigate the potential effect of perturbations on or within the system. If we re-examine each of the above models, we note that the parameter estimation problem is slightly different for each of them. In the deterministic model (Eq. (2)), both r and κ are constants; therefore, it is necessary to estimate only two

parameters in this model. In Equation (3) in which κ is constant and the growth rate is now assumed to be a random variable, $R \sim \mathcal{N}(\mu_R, \sigma_R^2)$, there are effectively three values which must be estimated to fully determine the RDE model: κ , μ_R , and σ_R . If R was assumed to satisfy a different statistical distribution, then it would be necessary to estimate all the parameters to determine completely the assumed distribution. In the last model in Equation (4), both the growth rate, R , and limiting capacity, K , are assumed to be random variable parameters behaving according to a normal distribution. Hence, in this model, we must be able to estimate μ_R , σ_R , μ_K , and σ_K . Therefore, the parameter estimation problem is different depending on which variables are assumed to be random variables and the choice of any assumed distribution for each random variable parameter. In Section 3, we consider the solution to the RDE to be a collection of solution trajectories to a sample deterministic system. As such, we develop a method for parameter estimation which utilizes the sample deterministic system and methods for parameter estimation in deterministic systems. In Section 4, we also utilize deterministic methods for parameter estimation; however, the method developed in this section is based on the equivalence of an RDE model to a stochastic differential equation (SDE) model and the relationship between a SDE model and deterministic system for large population sizes.

3 Method 1: Parameter Estimation Method Using Sample Deterministic Equation

There are two common ways to approach RDE, the mean calculus approach and the sample function approach [2]. We use here the sample function approach in which one considers *individual realizations* of the RDE. Each realization of the RDE is a solution to a deterministic differential equation, called a sample deterministic differential equation, which is assumed here to have a unique solution [2]. For example, for every realization r of $R \sim \mathcal{N}(\mu_R, \sigma_R^2)$ in the RDE model, we obtain the deterministic differential equation given by Equation (2). In this approach to RDE models, the solution to an RDE is a collection of solution trajectories to the sample deterministic equations. Let $\mathbf{h}(t; \mathbf{Q})$ represent the observation process in the RDE model and $\mathbf{f}_d(t, \mathbf{q})$ be the observation process in the sample deterministic differential equation. Then we can assume data $\mathbf{z} = (z_1, z_2, \dots, z_N)^T$ is a realization of a random variable \mathbf{Z} which is generated from a stochastic process given by a ‘true’ RDE model and can be defined as

$$Z_j = h(t_j; \mathbf{Q}_0) + \mathcal{E}_j = f_d(t_j; \mathbf{q}_0) + \mathcal{E}_j, \quad j = 1, \dots, N \quad (5)$$

where \mathcal{E}_j is a normally distributed random variable with mean $\mathbb{E}(\mathcal{E}_j) = 0$ and known variance $Var(\mathcal{E}_j) = \sigma_0^2$, \mathbf{Q}_0 is the true random variable parameter in the RDE system, and \mathbf{q}_0 is a realization of the random variable \mathbf{Q}_0 .

Therefore, to obtain a parameter estimate, we first define

$$J_{RDE_1}^N(\mathbf{q}; \mathbf{Z}) = \frac{1}{N} \sum_{k=1}^N (Z_k - f_d(t_k; \mathbf{q}))^2 \quad (6)$$

to be the cost function. We then seek an estimator

$$\mathbf{q}_{RDE_1}^N = \operatorname{argmin}_{\mathbf{q} \in \Omega_q} J_{RDE_1}^N(\mathbf{q}; \mathbf{Z})$$

with realization

$$\hat{\mathbf{q}}_{RDE_1}^N = \operatorname{argmin}_{\mathbf{q} \in \Omega_q} J_{RDE_1}^N(\mathbf{q}; \mathbf{z}). \quad (7)$$

This estimator could be approximated by an estimate for the realization \mathbf{q} of \mathbf{Q} given a specific data set \mathbf{z} . However, as discussed above, we need to estimate all statistical parameters which completely define \mathbf{Q} , not simply one realization of \mathbf{Q} . Given multiple data sets \mathbf{z}_k for the given physical system, we can estimate the statistical parameters for the distribution. For example, in Equation (3), we need to estimate both the mean μ_R and the standard deviation σ_R . If we have M data sets for the physical system, then we can obtain estimates $\hat{\mathbf{q}}_k$, $k = 1, \dots, M$ using Equation (7) for each data set \mathbf{z}_k , $k = 1, \dots, M$. For an arbitrary random variable parameter $Q_i \sim \mathcal{N}(\mu_i, \sigma_i)$, then the mean μ_i can be approximated by

$$\hat{\mu}_i \approx \frac{1}{M} \sum_{k=1}^M (\hat{q}_i)_k. \quad (8)$$

Similarly, the standard deviation can be approximated by

$$\sigma_i \approx \sqrt{\frac{1}{M} \sum_{k=1}^M ((\hat{q}_i)_k - \mu_i)^2}. \quad (9)$$

We note that given only one data set, $M = 1$, Equation (8) yields $\mu_i \approx (\hat{q}_i)_1$; substituting μ_i into Equation (9), we have $\sigma_i = 0$. Therefore, it is necessary to have multiple data sets in order to approximate the standard deviation using this method. If one assumes the random variable parameters behave according to a distribution different from the normal distribution or the distribution is unknown, one can use tools in a computational software such as Matlab or Minitab to estimate the statistical distribution parameters and/or to determine the distribution which best fits the set of estimates $\{(\hat{q}_i)_k\}_{k=1}^M$ for each parameter Q_i .

4 Method 2: Parameter Estimation Method Using a Pointwise Equivalent SDE Model

An alternate method is based on the pointwise equivalence between a RDE and a SDE model. Established in [4] and summarized in [2], it was shown that there are classes of RDEs which are pointwise equivalent to corresponding Itô SDEs. If one assumes the solution to a RDE is a stochastic sample solution, i.e., the collection of solution trajectories of the sample deterministic equations, then there are classes of RDEs for which their solutions have the same probability density function at each time t as the solutions to corresponding Itô SDEs. Two such classes are given in [2, 4].

The first class are RDEs of the form

$$\frac{dx(t; X_0, \mathbf{Q})}{dt} = \alpha(t)x(t; X_0, \mathbf{Q}) + \gamma(t) + \mathbf{Q} \cdot \varrho(t), \quad x(0; X_0, \mathbf{Q}) = X_0, \quad (10)$$

where $\mathbf{Q} = (Q_0, Q_1, \dots, Q_{m-1})^T$, $Q_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 0, 1, 2, \dots, m-1$ and $\varrho = (\varrho_0, \varrho_1, \dots, \varrho_{m-1})^T$, γ and α are non-random functions of t . It is shown in [2] that a random differential equation of this form is pointwise equivalent to the SDE given by

$$dX(t) = (\alpha(t)X(t) + \gamma(t) + \boldsymbol{\mu} \cdot \boldsymbol{\varrho}(t)) dt + \sqrt{2h(t)}dW(t), \quad X(0) = X_0, \quad (11)$$

if $\boldsymbol{\varrho}$ has the property that the function

$$h(t) = \sum_{j=0}^{m-1} \sigma_j^2 \varrho_j(t) \left(\int_0^t \varrho_j(s) \exp \left(\int_s^t \alpha(\tau) d\tau \right) ds \right)$$

is non-negative for any $t \geq 0$. In Equation (11), $\boldsymbol{\mu} = (\mu_0, \mu_1, \dots, \mu_{m-1})^T$ is the vector of mean values μ_j , $j = 1, \dots, m-1$, for each random parameter Q_j in \mathbf{Q} .

The second case considered in [2, 4] are RDEs of the form

$$\frac{dx(t; X_0, \mathbf{Q})}{dt} = (\mathbf{Q} \cdot \varrho(t) + \gamma(t)) (x(t; X_0, \mathbf{Q}) + c), \quad x(0; X_0, \mathbf{Q}) = X_0, \quad (12)$$

where again $\mathbf{Q} = (Q_0, Q_1, \dots, Q_{m-1})^T$, $Q_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 0, 1, 2, \dots, m-1$, $\varrho = (\varrho_0, \varrho_1, \dots, \varrho_{m-1})^T$ is a non-random vector function of t , γ is a non-random function of t and c is a constant. It is shown in [2] that a random differential equation of this form is pointwise equivalent to the SDE given by

$$dX(t) = \left(\boldsymbol{\mu} \cdot \boldsymbol{\varrho}(t) + \gamma(t) + \tilde{h}(t) \right) (X(t) + c) dt + \sqrt{2\tilde{h}(t)} (X(t) + c) dW(t), \quad X(0) = X_0, \quad (13)$$

if $\boldsymbol{\varrho}$ has the property that the function

$$\tilde{h}(t) = \sum_{j=0}^{m-1} \sigma_j^2 \varrho_j(t) \left(\int_0^t \varrho_j(s) ds \right)$$

is non-negative for any $t \geq 0$.

If the RDE model is of the form in Equation (10) or (12) or can be transformed such that it has one of the given forms, then as described above, a pointwise equivalence can be established between the RDE model and a corresponding SDE model. The general form of an Itô SDE as discussed in [1] is given by

$$dX(t) = \mu(t, X(t))dt + B(t, X(t))dW(t), \quad t \geq 0, \quad (14)$$

where

$$\mu(t, X) = \frac{E(\Delta X)}{\Delta t}, \quad B(t, X) = V^{1/2} \text{ with } V = \frac{E(\Delta X \Delta X^T)}{\Delta t},$$

and W is a Wiener process such that $W(0) = 0$ and

$$W(t) - W(s) \approx \mathcal{N}(0, t - s).$$

For large population sizes [10, 6], the dynamics in the SDE are similar to the dynamics of a deterministic system in which the stochastic effects become less important in the overall dynamics of the solution. To approximate the SDE with a deterministic system, we take the expectation of the SDE,

$$\mathbb{E}(dX(t, \mathbf{q})) = \mathbb{E}(\mu(t, X(t, \mathbf{q}))dt) + \mathbb{E}(B(t, X(t, \mathbf{q}))dW(t)) = \mu(t, X(t, \mathbf{q}))dt$$

or

$$\frac{\mathbb{E}(dX(t, \mathbf{q}))}{dt} = \mu(t, X(t, \mathbf{q}))$$

since $\mathbb{E}(dW) = 0$. Therefore, the expected trend for an SDE is given by the expected deterministic system. Hence, given a RDE model of the form in Equation (10) or (12) or which can be transformed into one of these forms, we can first develop the pointwise equivalent SDE which can then be approximated with a deterministic system if the population size is ‘sufficiently large’.

In this method, we again assume the data, $\mathbf{z} = (z_1, z_2, \dots, z_N)^T$ is a realization of random variable \mathbf{Z} which is generated from a stochastic process given by a ‘true’ RDE model but the model can be approximated by a deterministic system which is estimated from a pointwise equivalent SDE for the purpose of parameter estimation. Let $\mathbf{h}(t, \mathbf{q})$ represent the observation process in the RDE model and $\mathbf{f}_s(t, \mathbf{q})$ the corresponding observation process in the approximate deterministic system for the pointwise equivalent SDE, then we assume

$$Z_j = h(t_j; \mathbf{q}_0) + \mathcal{E}_j \approx f_s(t_j; \mathbf{q}_0) + \mathcal{E}_j, \quad j = 1, \dots, N, \quad (15)$$

where \mathcal{E}_j is a normally distributed random variable with mean $\mathbb{E}(\mathcal{E}_j) = 0$ and known variance $Var(\mathcal{E}_j) = \sigma_0^2$. In the parameter estimation problem, define

$$J_{RDE_2}^N(\mathbf{q}; \mathbf{Z}) = \frac{1}{N} \sum_{k=1}^N (Z_k - f_s(t_k; \mathbf{q}))^2 \quad (16)$$

to be the cost function with which we seek an estimator

$$\mathbf{q}_{RDE_2}^N = \arg \min_{\mathbf{q} \in \Omega_q} J_{RDE_2}^N(\mathbf{q}; \mathbf{Z})$$

with realization

$$\hat{\mathbf{q}}_{RDE_2}^N = \arg \min_{\mathbf{q} \in \Omega_q} J_{RDE_2}^N(\mathbf{q}; \mathbf{z}). \quad (17)$$

In the implementation of this method, the approximate deterministic system $f_s(t, \mathbf{q})$ contains all values which must be estimated to fully define the normally distributed random variable parameter. Therefore, we do not need to estimate the parameters of the statistical distribution separately. Two examples are given below.

5 Results

We test this methodology for two nested models, the logistic growth RDE model given by Equation (3) and the Bernoulli growth RDE model given by

$$\frac{dx(t; Q)}{dt} = Rx(t; Q) \left(1 - \left(\frac{x(t; Q)}{\kappa} \right)^\beta \right) \quad (18)$$

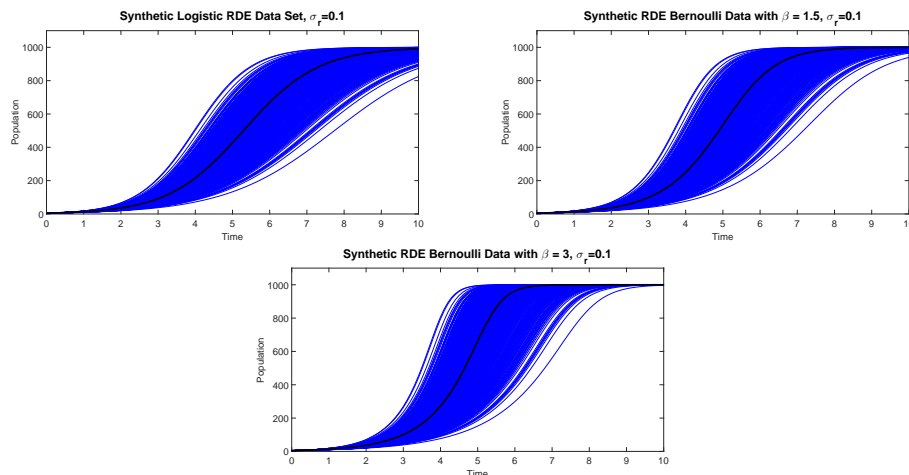


Figure 1: One thousand synthetic data sets *without noise added* using $\mu_r = 1$, $\kappa = 10000$ with $\sigma_r = 0.1$ in the logistic RDE model (Eq. (3)) (first plot), the Bernoulli RDE model (Eq. (18)) with $\beta = 1.5$ (second plot) and $\beta = 3$ (third plot). The deterministic model with both $r = 1$ and $\kappa = 1000$ constant is given in black in each of the figures.

assuming the growth parameter R is a random variable, $R \sim \mathcal{N}(\mu_R, \sigma_R^2)$, while both β and the limiting capacity κ are held constant. These are nested models, because if we set $\beta = 1$ in Equation (18), we obtain the logistic RDE model given in Equation (3). We consider trials using three different values of β in Equation (18), $\beta = 1$ (equivalent to the logistic RDE model, Eq. (3)), $\beta = 1.5$ and $\beta = 3$. In each trial, we let $\kappa = 1000$, $\mu_r = 1$ and $\sigma_r = 0.1$. For each trial, we generated 1000 different data sets (Figure 1) by taking 1000 realizations r_j , $j = 1, \dots, 1000$, of $R \sim \mathcal{N}(\mu_R, \sigma_R^2)$ and solving the sample deterministic system

$$\frac{dx(t)}{dt} = r_j x(t) \left(1 - \left(\frac{x(t)}{\kappa} \right)^\beta \right), \quad j = 1, \dots, 1000.$$

We then add noise using Equation (5) with $\sigma_0 = 10$ for \mathcal{E}_j ; the data with noise added is shown in Figure 2.

5.1 Method 1 Results

In implementing Method 1, we note that for the logistic model, the sample deterministic differential equation is given by Equation (2),

$$\frac{dx}{dt} = rx(t) \left(1 - \frac{x(t)}{\kappa} \right);$$

for the Bernoulli RDE model, the sample deterministic differential equation is given by

$$\frac{dx(t)}{dt} = rx(t) \left(1 - \left(\frac{x(t)}{\kappa} \right)^\beta \right). \quad (19)$$

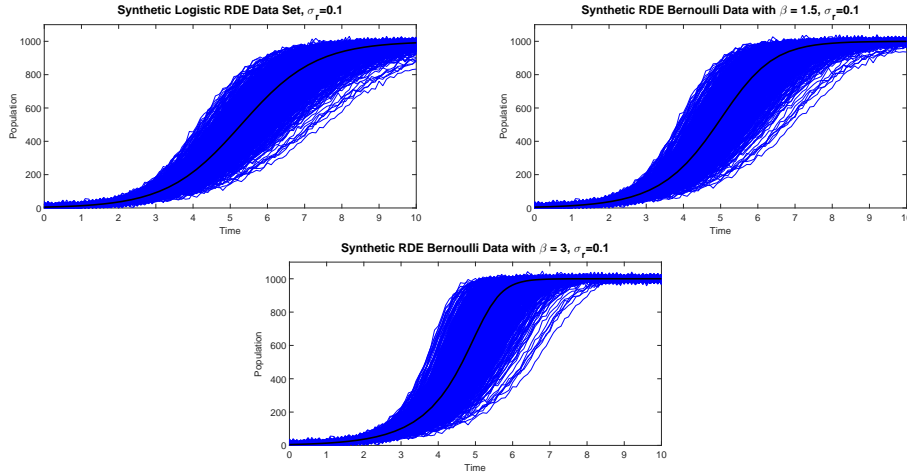


Figure 2: One thousand synthetic data sets *with noise* $\mathcal{E} \sim \mathcal{N}(0, 100)$ using $\mu_r = 1$, $\kappa = 10000$ with $\sigma_r = 0.1$ in the logistic RDE model (Eq. (3)) (first plot), the Bernoulli RDE model (Eq. (18)) with $\beta = 1.5$ (second plot) and $\beta = 3$ (third plot). The deterministic model with both $r = 1$ and $\kappa = 1000$ constant is given in black in each of the figures.

We use the built-in Matlab program *fminsearch* to minimize the cost function given in Equation (6) where $f_d(t, \mathbf{q})$ is the solution to the sample deterministic differential system given by either Equation (2) or Equation (19) depending on the data set (synthetic data from the Logistic or Bernoulli RDE model, respectively). The initial guess is set as $1.10\mathbf{q}_0$ where \mathbf{q}_0 is the exact value. We first test the method using data generated from the Logistic RDE model with no noise added (first plot in Figure 1). We test the method using a random draw of M data sets from the 1000 initial data sets generated (this is to simulate having M data sets taken from the physical system). We consider $M = 1, 2, 3, 4, 5, 10, 15, 20, \dots, 95, 100$. Recall, using $M = 1$, we have $\sigma_R = 0$; therefore, M must be chosen to be greater than 1 in order to estimate $\sigma_R \neq 0$. We consider different values of M to investigate the effect that the choice of M has on the accuracy of the estimate. Each estimate \hat{r}_k , $k = 1, \dots, M$ will each be different. Depending on *which* M data sets are randomly chosen from the set of 1000, this will result in a different estimate for μ_R (as well as the mean value for κ). If we examine Figure 1, we note that the realization r of $R \sim \mathcal{N}(\mu_R, \sigma_R)$ which generated the ‘lowest’ curve (or the one with the lowest rate of growth) will be smaller than the realization which generated the ‘highest’ curve (or one with the fastest rate of growth). If $M = 2$, the value of μ_R will be quite different using these extreme data sets than if we have two data sets which behave similarly. Therefore, we use different draws of M data sets to simulate for the variation which might be seen depending on *which* M data sets from the physical system are collected. The results are given in Figure 3.

In Figure 3, the red line indicates the exact values of $\mu_R = 1$, $\sigma_R = 0.1$ and $\kappa = 1000$. Using 100 different draws of M data sets, we have plotted the largest, smallest and median estimated values for each parameter reflecting the extent *which group of* M has on the resulting estimate. We note that when M is small ($M < 30$), the range of possible estimates vary more than when M is large, especially for μ_R and σ_R . However,

even when using only 1 data set, the median estimate is quite accurate for both μ_R and κ . In order to obtain close to the same level of accuracy for the median value of σ_R , we need $M \geq 5$ (also see Table 1). We do notice that the estimate for κ is always overestimated with a median value of approximately 1001; however, this only represents a 0.10% error in the estimated value. In fact, the maximum percent relative error in κ does not vary much as a function of the number of data sets. We also observe that although we are assuming κ is constant, we do have a small standard deviation of about 0.25 in the estimate of κ simply due to the variation in the estimated value of κ given different data sets. However, Table 1 shows much more variation in the range of estimates for both μ_R and σ_R as a function of the number of data sets available. Ten or more data sets are required to obtain estimates with less than 10% relative error in μ_R ; however, even with 100 data sets from the physical system, there is still the potential of having more than 15% relative in the estimate of σ_R . Recall that the accuracy of the estimate depends on *which* M data sets are used in the estimation. Even though the potential is to have over 10% relative error using $M = 100$ data sets; only 16% of the random draws of 100 data sets resulted in percent relative error more than 10%. If $M \geq 70$, then less than 25% of the time there is an estimate with a percent relative error of greater than 10% and when $M \geq 25$, then less than 40% of the time, there is an estimate for σ_R with more than 10% relative error. In summary, using Method 1 for estimating the parameters in the RDE logistic model, depending on the number of data sets available and which data sets are in this group, we were able to obtain good estimates for both κ and μ_R with as little as 1 data set. However, to ensure relative error less than 10% in μ_R , it was best to use 10 or more data sets if possible. The estimate for σ_R was much more difficult. Since σ_R is smaller with an exact value of only 0.10, it is much harder to estimate with the same degree of percent relative error. Nonetheless, with 15 or more data sets, we could obtain estimates within 0.05 of the actual value.

Note that even though the synthetic data was generated using the Logistic RDE model, this is identical to generating the data using the Bernoulli RDE model with $\beta = 1$. Therefore, the parameter estimation problem could be considered for the Bernoulli model in which we need to estimate not only μ_R , σ_R , and κ but β as well. Therefore, we also attempted to estimate these parameters using only $M = 15$ data sets and obtained the results in Table 2. The estimates are comparable to those obtained when using the sample deterministic logistic model and estimating one less parameter.

We repeat the parameter estimation problem for the Logistic RDE model using the noisy synthetic data from the Logistic RDE Model (first plot in Figure 2). We only consider $M = 15$ in this trial. The results are given in Table 3 which are comparable to the results we saw using synthetic data without noise added. Adding the noise to the data did, however, result in a slightly larger standard deviation of 2.49 for κ even though κ is assumed constant. However, when noise is added to the data, it is expected that there will be some variation simply due to the variation in the estimates. We note, though, the calculated standard deviation in κ is small compared to the magnitude of κ .

Again, using only $M = 15$ data sets, we use method 1 to estimate the parameters in the Bernoulli growth model with $\beta = 1.5$ and $\beta = 3$ using data with and without

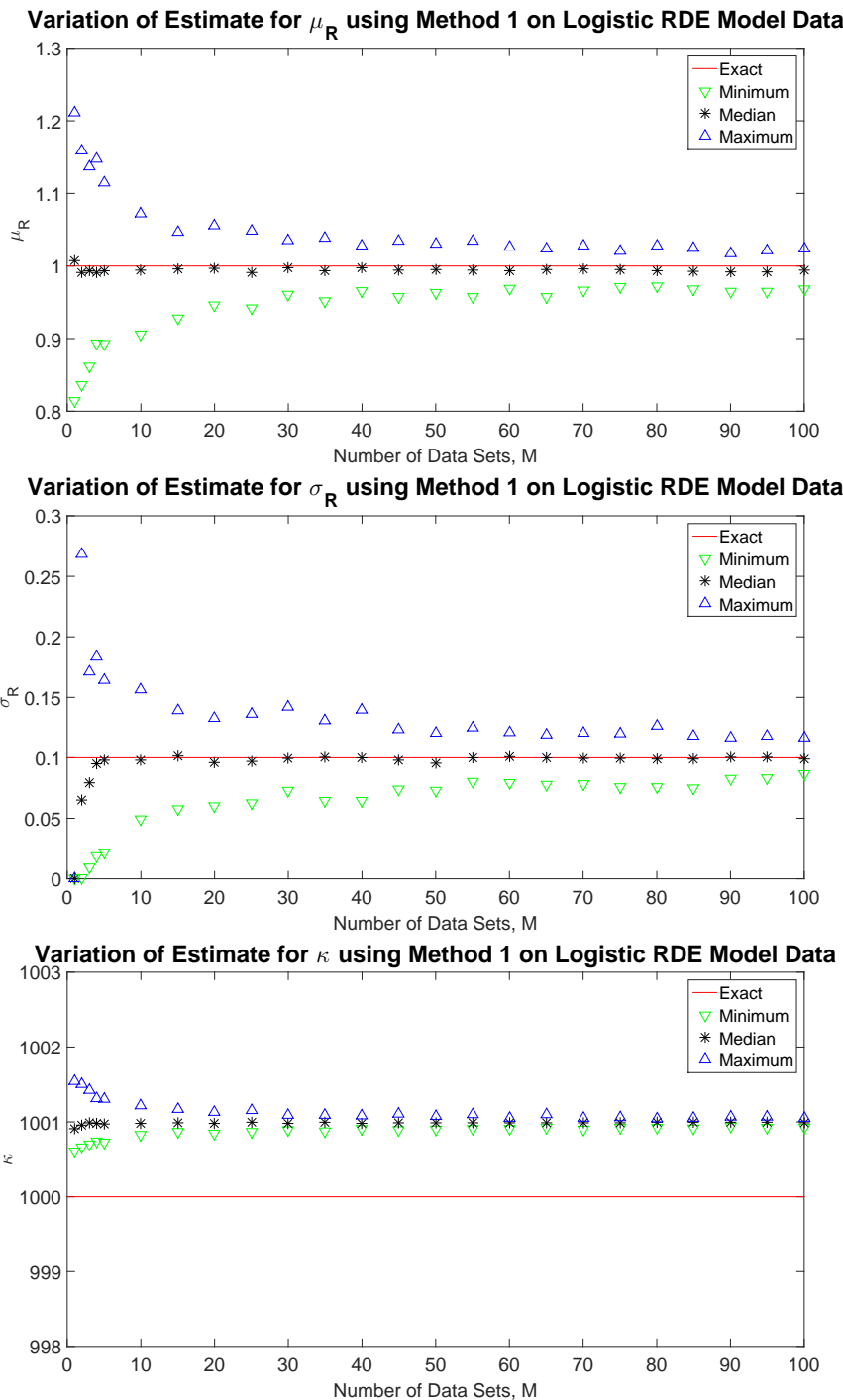


Figure 3: Variation in the estimated values for μ_R , σ_R , and κ when the number of data sets M are varied when using synthetic data from the Logistic RDE Model in Equation (3). The red line indicates the exact parameter value. We consider 100 different random draws of M data sets from the 1000 generated. The black star represents the median estimated value across the different groups of M data sets; the blue upward triangle is the maximum estimated parameter value, and the green downward triangle is the minimum estimated value.

Table 1: Variation of Percent Relative Error in Parameter Estimate for Logistic RDE Model using Synthetic Data with No Noise Added Using Method 1 with 100 Different Random Draws of M Data Sets.

μ_R		
M	Max Perc. Rel. Error	Med Perc. Rel. Error
1	21.13%	0.77%
2	16.35%	0.88%
3	13.78 %	0.67%
4	14.74%	0.91%
5	11.47%	0.65%
10	9.39%	0.54%
15	7.19%	0.44%
20	5.56%	0.32%
σ_R		
M	Max Perc. Rel. Error	Med Perc. Rel. Error
1	100.00%	100.00%
2	168.57%	35.05%
3	90.27%	20.86%
4	83.46%	5.04%
5	77.92%	1.89%
10	56.25%	2.01%
15	42.55%	1.50%
20	40.04%	4.11%
40	39.81%	0.15%
60	21.25%	0.82%
80	26.42%	1.04%
100	16.75%	1.09%

Table 2: Parameter Estimates for Bernoulli RDE Model using Synthetic Logistic RDE Model Data with No Noise Added Using Method 1 with 100 Different Random Draws of $M = 15$ Data Sets.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9906	0.94%	7.66%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.1010	1.04%	42.89%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.14	0.01%	0.06%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	1.0084	0.84%	0.94%

Table 3: Parameter Estimates for Logistic RDE Model using Synthetic Data with Noise Added Using Method 1 with 100 Different Random Draws of $M = 15$ Data Sets.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9958	0.42%	7.49%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.0971	2.90%	47.46%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.17	0.02%	0.17%

Table 4: Parameter Estimates for Bernoulli RDE Model using Synthetic Data with $\beta = 1.5$, No Noise Added, Using Method 1 with 100 Different Random Draws of $M = 15$ Data Sets.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9896	1.04%	8.09%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.0956	4.34%	51.52%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.03	0.003%	0.34%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1.5	1.5131	0.87%	20.81%

noise added. The results are given in Tables 4 - 7. Using data without noise (Tables 4 and 6), we obtain estimates similar to those obtained using synthetic data from the Logistic model for μ_R , σ_R and κ ; however, the maximum relative error in the estimates for β are slightly worse than expected. Some of the difficulty in the estimation may lie in the identifiability of the parameters in the sample deterministic model since κ and β are found jointly in the denominator in the term $\frac{(x(t))^\beta}{\kappa^\beta}$. To address this issue, we let $\tilde{\kappa} = \kappa^\beta$ in the sample deterministic differential equation and seek an estimate $\hat{\mathbf{q}}$ for $\mathbf{q} = [\mu_R, \sigma_R, \tilde{\kappa}, \beta]$ in the system

$$\frac{dx(t)}{dt} = rx(t) \left(1 - \frac{(x(t))^\beta}{\tilde{\kappa}} \right) \quad (20)$$

where the estimate for κ can be calculated as $\hat{\kappa} = \hat{\tilde{\kappa}}^{(1/\hat{\beta})}$.

We compare the estimation of r , κ , and β in the original sample deterministic system to the estimates obtained in the re-parameterized sample deterministic system for one of the data sets (with no noise added) in which the estimate for β is poor. Figure 4 shows the results in the parameter estimation. Using the re-parameterized

Table 5: Parameter Estimates for Bernoulli RDE Model using Synthetic Data with $\beta = 1.5$, With Noise Added, Using Method 1 with 100 Different Random Draws of $M = 15$ Data Sets.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	1.0048	0.48%	6.48%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.1001	0.06%	40.26%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	999.98	0.02%	0.27%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1.5	1.5002	0.01%	18.83%

Table 6: Parameter Estimates for Bernoulli RDE Model using Synthetic Data with $\beta = 3$, No Noise Added, Using Method 1 with 100 Different Random Draws of $M = 15$ Data Sets.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9923	0.77%	7.56%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.1004	0.36%	44.76%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	998.98	0.10%	2.04%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
3	3.3928	13.09%	130.78%

Table 7: Parameter Estimates for Bernoulli RDE Model using Synthetic Data with $\beta = 3$, With Noise Added, Using Method 1 with 100 Different Random Draws of $M = 15$ Data Sets.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	1.0068	0.68%	6.96%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.0899	10.08%	44.72%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	999.24	0.07%	1.33%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
3	3.3301	11.00%	101.83%

deterministic model produced more accurate estimates for not only β but all three parameters. The original estimates are given by $\hat{r} = 0.95$, $\hat{\kappa} = 987.34$ and $\hat{\beta} = 7.75$. The estimates using the re-parameterized system are given by $\hat{r} = 0.98$, $\hat{\kappa} = 1000.03$ and $\hat{\beta} = 3.04$.

Using the same type of re-parameterization to estimate the parameters for the RDE model, we estimated the parameters using synthetic data from the RDE model for $\beta = 1.5$ and $\beta = 3$ with noise added. If we compare the results in Table 5 of the parameter estimation for the Bernoulli model with $\beta = 1.5$ using the original sample deterministic system (Eq. (19)) and those results in Table 8 using the re-parameterized sample deterministic system (Eq. (20)), we see that although the results for μ_R and σ_R are slightly worse, the estimate for κ is slightly better and the maximum error in β is significantly better with a maximum percent relative error of only 1.7% instead of 18.83% when using the original model. Similar improvement can be found when comparing Table 7 using the original sample deterministic system and Table 9 using the re-parameterized sample deterministic system when $\beta = 3$. Although the maximum relative error in β is still about 19%, there are only 9 trials out of the 100 which resulted in an error of over 10%.

To summarize, using Method 1 to estimate parameters in a RDE model provided good estimates for both the mean of the random variable parameter and all constant parameters in the model in the examples presented for the Logistic RDE Model and the Bernoulli RDE Model with either $\beta = 1.5$ or $\beta = 3$ when synthetic noisy data was used in the estimation problem. When using $M = 15$ data sets, the accuracy of σ_R depended on *which* data sets were used in the estimation problem. Figure 5 shows the difference in the RDE solution or collection of solution trajectories to the sample deterministic equation when using the exact value of σ_R compared to the estimate which resulted in the worse percent relative error when using $\beta = 3$ and the re-parameterized sample deterministic system. In this case, σ_R was underestimated at approximately 0.06 (compared to the exact value of 0.1). We can see that the inaccuracy in the estimated value of σ_R resulted in less variability in the collection

Table 8: Parameter Estimates for Bernoulli RDE Model using Synthetic Data with $\beta = 1.5$, With Noise Added, Using Method 1 with Reparameterized Sample Deterministic System and 100 Different Random Draws of $M = 15$ Data Sets.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9958	0.42%	7.45%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.0981	1.86%	49.24%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.07	0.008%	0.15%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1.5	1.5010	0.07%	1.70%

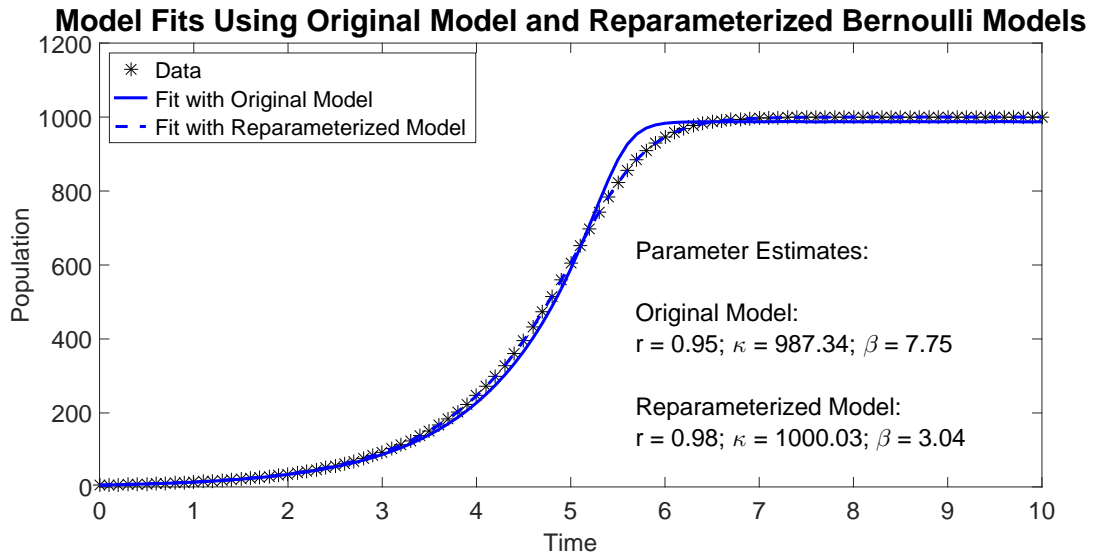


Figure 4: Comparison of the estimated values of r , κ and β using synthetic Bernoulli model data (black stars) with $\beta = 3$ in the sample deterministic Bernoulli models given by Equation (19) (original model fit given by the solid blue line) and Equation 20 (reparameterized model fit given by the dashed blue line).

Table 9: Parameter Estimates for Bernoulli RDE Model using Synthetic Data with $\beta = 3$, With Noise Added, Using Method 1 with Reparameterized Sample Deterministic System and 100 Different Random Draws of $M = 15$ Data Sets.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9986	0.14%	7.44%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.0949	5.08%	39.28%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	999.82	0.02%	0.62%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
3	3.0234	0.78%	19.26%

of solution trajectories. We expect this type of underestimation when the data sets used in the estimation problem are very similar with little variability in their growth rate. Nonetheless, the more data sets we are able to use, the more accurate we expect the estimate for σ_R (see Table 1). We do note, however, that when implementing this method, one must still address the identifiability of parameters in the sample deterministic system. We found that a re-parameterization of the sample deterministic system provided much better estimates of the RDE parameters than when using the original system in which there was an issue with identifiability.

Using this method, we also tested the ability to estimate the random parameter R if R had a Weibull distribution with scale parameter $A = 1$ and shape parameter $B = 10$, $R \sim \mathcal{W}(A, B)$. As before, we generated 1000 different solution trajectories without noise, first plot in Figure 6, and with noise, second plot in Figure 6. We then estimated the parameters A and B of the Weibull statistical distribution for R and the constant limiting capacity κ . As we did in the previous trials, we estimate the realization \mathbf{q} of \mathbf{Q} . Using the realizations, we estimate the parameters for the statistical distribution of the random variable parameters. In this case, R is the only random variable. Therefore, if we have M data sets, then we use $\{r_k\}_{k=1}^M$ to estimate the statistical parameters A and B of the Weibull distribution. To do this, we use the built-in Matlab function *fitdist*. The results using $M = 15$ and $M = 100$ synthetic data sets containing no noise are given in Table 10 and in Table 11 respectively. Even though the maximum percent relative error in the shape parameter B is 113% when using $M = 15$ data sets, the percent relative error was less than 10% in 98 of the 100 trials. Therefore, with only $M = 15$ data sets, we can use method 1 to fairly accurately estimate both the constant and random parameter values even when the random parameter satisfies a distribution other than the normal distribution. We obtain only slightly better estimates using $M = 100$ data sets in the estimation process (Table 11). Moreover, the method was still robust when noise was added to the data as shown in Tables 12 and in Table 13 using $M = 15$ and $M = 100$ randomly chosen synthetic data sets.

Comparison of RDE Solution Trajectories with Exact and Worst Estimated σ_R

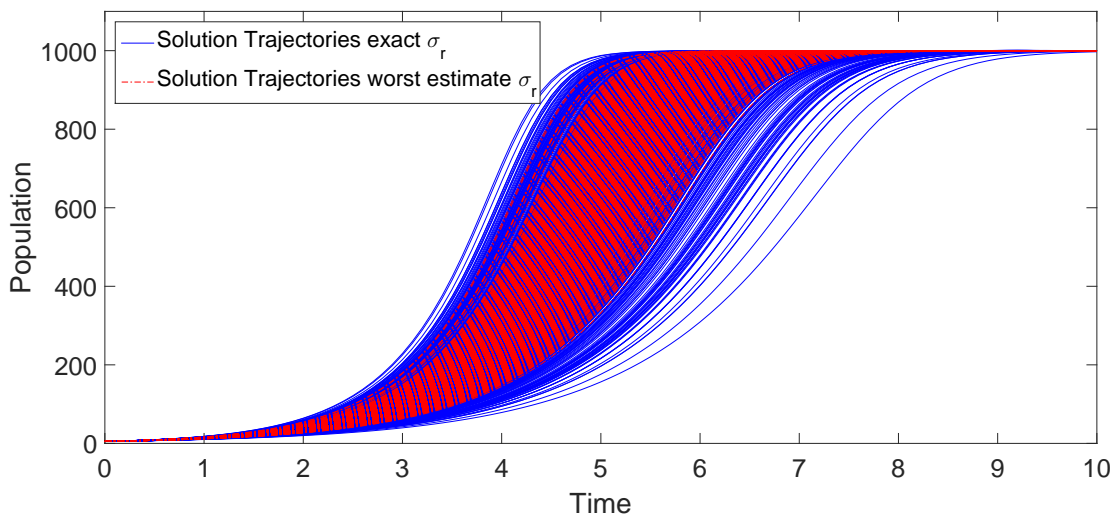


Figure 5: Comparison of the RDE solution or collection of solution trajectories to the sample deterministic system for the RDE Model using the exact value for σ_R and the estimated value for σ_R , $R \sim \mathcal{N}(\mu_R, \sigma_R)$, when $\beta = 3$ and $\kappa = 1000$ in Bernoulli RDE model given by Equation (18).

Table 10: Parameter Estimates for Logistic RDE Model using Synthetic Data with $R \sim \mathcal{W}(1, 10)$, No Noise Added, Using Method 1 with 100 Different Random Draws of $M = 15$ Data Sets.

Scale Parameter A			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9961	0.3881%	9.17%
Shape Parameter B			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
10	10.6128	6.13%	113.82%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	999.99	0.0007%	0.003%

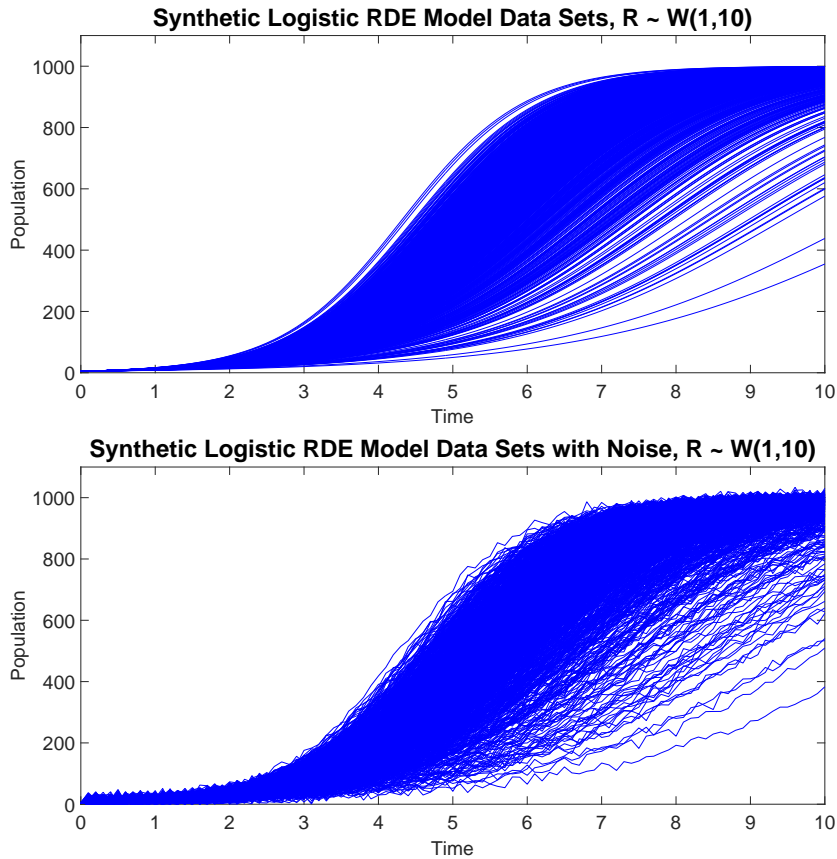


Figure 6: One thousand synthetic data sets without noise added (first plot) and with noise added (second plot) for the Logistic RDE Model (Eq. (3)) assuming $R \sim \mathcal{W}(1, 10)$ and $\kappa = 10000$.

Table 11: Parameter Estimates for Logistic RDE Model using Synthetic Data with $R \sim \mathcal{W}(1, 10)$, No Noise Added, Using Method 1 with 100 Different Random Draws of $M = 100$ Data Sets.

Scale Parameter A			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	1.0010	0.10%	2.98%
Shape Parameter B			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
10	10.1268	1.27%	27.18%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	999.99	2.56e-05%	0.004%

Table 12: Parameter Estimates for Logistic RDE Model using Synthetic Data with $R \sim \mathcal{W}(1, 10)$, With Noise Added, Using Method 1 with 100 Different Random Draws of $M = 15$ Data Sets.

Scale Parameter A			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9981	0.19%	7.25%
Shape Parameter B			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
10	10.1252	1.25%	82.11%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	999.76	0.02%	0.44%

Table 13: Parameter Estimates for Logistic RDE Model using Synthetic Data with $R \sim \mathcal{W}(1, 10)$, With Noise Added, Using Method 1 with 100 Different Random Draws of $M = 100$ Data Sets.

Scale Parameter A			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	1.0053	0.53%	2.99%
Shape Parameter B			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
10	10.0549	0.55%	22.01%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.75	0.07%	3.23%

5.2 Method 2 Results

In order to use Method 2, we must also first derive a pointwise equivalent SDE for each model and then approximate the SDE with an appropriate deterministic system. In order to use a pointwise equivalent SDE for model selection, the RDE model has to be in the form of either Equation (10) or (12). We first transform both the logistic RDE model and Bernoulli RDE model into the appropriate form and derive the pointwise equivalent SDE for each in Section 5.2.1 and 5.2.2 respectively.

5.2.1 Derivation of Pointwise Equivalent SDE for Logistic Model

The deterministic logistic growth model is given by

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{\kappa}\right).$$

We first transform the deterministic model by letting $y = \frac{1}{x}$. Then

$$\begin{aligned} \frac{dy}{dt} &= -\frac{1}{x^2} \frac{dx}{dt} \\ &= -\frac{1}{x^2} \left(rx \left(1 - \frac{x}{\kappa}\right) \right) \\ &= -r \left(\frac{1}{x} - \frac{1}{\kappa} \right) \\ &= -r \left(y - \frac{1}{\kappa} \right) \end{aligned}$$

with $y(0) = \frac{1}{x_0}$. The transformed RDE is then given by

$$\frac{dy(t; Q)}{dt} = -R \left(y(t; Q) - \frac{1}{\kappa} \right), \quad y(0; Q) = y_0$$

which has the form of Equation (12) with $Q = R$, $\varrho = -1$, $\gamma(t) = 0$, and $c = -\frac{1}{\kappa}$. The pointwise equivalent RDE in $Y(t)$ can be found using the form in Equation (13). Given $\mu = \mu_R$ and

$$\tilde{h}(t) = \sigma_0^2 \varrho_0(t) \int_0^t \varrho_0(s) ds = -\sigma_R^2 \int_0^t -1 ds = \sigma_R^2 t,$$

the equivalent SDE is given by

$$dY(t) = \left(-\mu_R + \sigma_R^2 t \right) \left(Y(t) - \frac{1}{\kappa} \right) dt + \sqrt{2t} \sigma_R \left(Y(t) - \frac{1}{\kappa} \right) dW(t), \quad Y(0) = y_0.$$

To determine the pointwise SDE in $X(t)$, we let $X(t) = \frac{1}{Y(t)}$ and use Itô's formula, the chain rule for Itô calculus ([2]), to find $dX(t)$. Let

$$dY(t) = g(t, Y(t))dt + \sigma(t, Y(t))dW(t),$$

and assume h is a function of t and y that is continuously differentiable in t and twice differentiable in y . Then Itô's formula gives the chain rule for differentiation as

$$dh(t, Y(t)) = \frac{\partial h(t, Y(t))}{\partial t} dt + \frac{\partial h(t, Y(t))}{\partial y} dY(t) + \frac{1}{2} \sigma^2(t, Y(t)) \frac{\partial^2 h(t, Y(t))}{\partial y^2} dt.$$

Applying this to $X(t) = \frac{1}{Y(t)}$, we have

$$\begin{aligned} dX(t) &= -\frac{1}{Y^2(t)} \left[(-\mu_R + \sigma_R^2 t) \left(Y(t) - \frac{1}{\kappa} \right) dt + \sqrt{2t} \sigma_R \left(Y(t) - \frac{1}{\kappa} \right) dW(t) \right] \\ &\quad + \frac{1}{Y^3(t)} \left[\sqrt{2t} \sigma_R \left(Y(t) - \frac{1}{\kappa} \right) \right]^2 dt \\ &= \frac{1}{Y(t)} \left[(\mu_R - \sigma_R^2 t) \left(1 - \frac{1}{\kappa Y(t)} \right) + 2t \sigma_R^2 \left(1 - \frac{1}{\kappa Y(t)} \right)^2 \right] dt \\ &\quad - \sqrt{2t} \sigma_R \frac{1}{Y(t)} \left(1 - \frac{1}{\kappa Y(t)} \right) dW(t) \\ &= X(t) \left[(\mu_R - \sigma_R^2 t) \left(1 - \frac{X(t)}{\kappa} \right) + 2t \sigma_R^2 \left(1 - \frac{X(t)}{\kappa} \right)^2 \right] dt \\ &\quad - \sqrt{2t} \sigma_R X(t) \left(1 - \frac{X(t)}{\kappa} \right) dW(t) \end{aligned}$$

Therefore, the pointwise equivalent SDE to the logistic RDE model given in Equation (3) is given by

$$\begin{aligned} dX(t) &= X(t) \left[(\mu_R - \sigma_R^2 t) \left(1 - \frac{X(t)}{\kappa} \right) + 2t \sigma_R^2 \left(1 - \frac{X(t)}{\kappa} \right)^2 \right] dt \\ &\quad - \sqrt{2t} \sigma_R X(t) \left(1 - \frac{X(t)}{\kappa} \right) dW(t), \quad X(0) = X_0. \end{aligned}$$

We note that the *SDE deterministic approximation* is given by

$$\frac{dx}{dt} = x \left[(\mu_R - \sigma_R^2 t) \left(1 - \frac{x(t)}{\kappa} \right) + 2t \sigma_R^2 \left(1 - \frac{x(t)}{\kappa} \right)^2 \right], \quad x(0) = x_0. \quad (21)$$

Note, that the deterministic system contains all three parameters which must be estimated to fully define the RDE model, μ_R , σ_R , and κ .

5.2.2 Derivation of Pointwise Equivalent SDE for Bernoulli Model

In a similar manner, we can derive the pointwise equivalent RDE for the Bernoulli model. The deterministic Bernoulli model is given by

$$\frac{dx}{dt} = rx \left(1 - \left(\frac{x}{\kappa} \right)^\beta \right).$$

We transform this deterministic model by letting $y = \frac{1}{x^\beta}$. Then

$$\begin{aligned} \frac{dy}{dt} &= -\frac{\beta}{x^{\beta+1}} \left(rx \left(1 - \left(\frac{x}{\kappa} \right)^\beta \right) \right) \\ &= -r\beta \left(\frac{1}{x^\beta} - \frac{1}{\kappa^\beta} \right) \\ &= -r\beta \left(y - \frac{1}{\kappa^\beta} \right) \end{aligned}$$

with $y(0) = \frac{1}{x_0^\beta}$. The transformed RDE is then given by

$$\frac{dy(t; Q)}{dt} = -R\beta \left(y(t; Q) - \frac{1}{\kappa^\beta} \right), \quad y(0; Q) = y_0$$

which has the form of Equation (12) with $Q = R$, $\varrho = -\beta$, $\gamma(t) = 0$, and $c = -\frac{1}{\kappa^\beta}$. The pointwise equivalent RDE in $Y(t)$ can be found using the form in Equation (13). Given $\mu = \mu_R$ and

$$\tilde{h}(t) = \sigma_0^2 \varrho_0(t) \int_0^t \varrho_0(s) ds = -\beta \sigma_R^2 \int_0^t -\beta ds = \beta^2 \sigma_R^2 t,$$

the pointwise equivalent SDE in $Y(t)$ is given by

$$dY(t) = \left(-\beta \mu_R + \beta^2 \sigma_R^2 t \right) \left(Y(t) - \frac{1}{\kappa^\beta} \right) dt + \sqrt{2t} \beta \sigma_R \left(Y(t) - \frac{1}{\kappa^\beta} \right) dW(t), \quad Y(0) = y_0.$$

As in the previous section, we need to transform the SDE into an SDE in $X(t)$. We let $X(t) = \frac{1}{Y^{1/\beta}(t)}$. Then applying Itô's formula we have

$$\begin{aligned}
dX(t) &= -\frac{1}{\beta Y^{1/\beta+1}(t)} \left[(-\beta\mu_R + \beta^2\sigma_R^2 t) \left(Y(t) - \frac{1}{\kappa^\beta} \right) dt + \sqrt{2t}\beta\sigma_R \left(Y(t) - \frac{1}{\kappa^\beta} \right) dW(t) \right] \\
&\quad + \frac{1}{\beta Y^{1/\beta+2}(t)} \left(\frac{1}{\beta} + 1 \right) \left[\sqrt{2t}\beta\sigma_R \left(Y(t) - \frac{1}{\kappa^\beta} \right) \right]^2 dt \\
&= \frac{1}{Y^{1/\beta}(t)} \left[(\mu_R - \beta\sigma_R^2 t) \left(1 - \frac{1}{\kappa^\beta Y(t)} \right) + 2t\sigma_R^2(1 + \beta) \left(1 - \frac{1}{\kappa^\beta Y(t)} \right)^2 \right] dt \\
&\quad - \frac{\sqrt{2t}\sigma_R}{Y^{1/\beta}(t)} \left(1 - \frac{1}{\kappa^\beta Y(t)} \right) dW(t) \\
&= X(t) \left[(\mu_R - \beta\sigma_R^2 t) \left(1 - \left(\frac{X(t)}{\kappa} \right)^\beta \right) + 2t\sigma_R^2(1 + \beta) \left(1 - \left(\frac{X(t)}{\kappa} \right)^\beta \right)^2 \right] dt \\
&\quad - \sqrt{2t}\sigma_R X(t) \left(1 - \left(\frac{X(t)}{\kappa} \right)^\beta \right) dW(t)
\end{aligned}$$

Therefore, the pointwise equivalent SDE to the Bernoulli RDE model given in Equation (18) is given by

$$\begin{aligned}
dX(t) &= X(t) \left[(\mu_R - \beta\sigma_R^2 t) \left(1 - \left(\frac{X(t)}{\kappa} \right)^\beta \right) + 2t\sigma_R^2(1 + \beta) \left(1 - \left(\frac{X(t)}{\kappa} \right)^\beta \right)^2 \right] dt \\
&\quad - \sqrt{2t}\sigma_R X(t) \left(1 - \left(\frac{X(t)}{\kappa} \right)^\beta \right) dW(t), \quad X(0) = X_0.
\end{aligned}$$

We note that the deterministic approximation for this SDE is given by

$$\frac{dx}{dt} = x \left[(\mu_R - \beta\sigma_R^2 t) \left(1 - \left(\frac{x}{\kappa} \right)^\beta \right) + 2t\sigma_R^2(1 + \beta) \left(1 - \left(\frac{x}{\kappa} \right)^\beta \right)^2 \right], \quad x(0) = x_0 \quad (22)$$

which again contains all the necessary parameters μ_R , σ_R , κ and β which must be estimated.

5.2.3 Results

We use the built-in Matlab program *fminsearch* to minimize the cost function given in Equation (16) where $f_s(t, \mathbf{q})$ is the solution to the deterministic approximation for the pointwise equivalent SDE given by either Equation (21) or Equation (22), depending on the data set (synthetic data from the Logistic or Bernoulli RDE model, respectively). As done in testing the first method, the initial guess is set as $1.10\mathbf{q}_0$ where \mathbf{q}_0 is the

exact value. We consider one random data set drawn from the 1000 possible synthetic data sets generated without noise (see Figure 1)) and estimate μ_R , σ_R , κ , and β (for the Bernoulli model only). We do this 100 times. The estimated values are given in Tables 14, 15 and 16 for synthetic data from the Logistic model, Bernoulli model with $\beta = 1.5$, and Bernoulli model with $\beta = 3$ respectively.

Analyzing all the tables, we first note that we could estimate κ and β (for the Bernoulli models) quite well; however, although the median estimate for μ_R had less than a 2.1% relative error in each case, the maximum relative error across 100 different individual data sets was approximately 23 - 26%. We also implement the parameter estimation problem using the re-parameterization $\tilde{\kappa} = \kappa^\beta$ as above and also $\gamma = \beta\sigma_R^2$ where the estimate for κ can be calculated as $\hat{\kappa} = \hat{\tilde{\kappa}}^{(1/\hat{\beta})}$ and the estimate for σ_R is calculated as $\hat{\sigma}_R = \sqrt{\hat{\gamma}/\hat{\beta}}$. Results in Table 17 and 18 indicate that the re-parameterization has little effect on the resulting estimates. We do note that there was one result when $\beta = 3$ which resulted in a relative error of about 12% in the estimate for β ; however, this was one case in the 100 trials. The main problem in the estimation results are in the estimation of σ_R . Similar to what we obtained using the first method, it was impossible to estimate σ_R using only one data set. In the first method, σ_R was calculated to be 0 when only one data set was available; in this second method, σ_R is an estimated parameter in the deterministic approximation, but the estimation is still approximately 0. However, this is understandable, since there is no variance with only one data set. Therefore, we reformulate the cost function to incorporate multiple data sets and seek a realization

$$\hat{\mathbf{q}}_{RDE_2}^N = \arg \min_{\mathbf{q} \in \Omega_q} \tilde{J}_{RDE_2}^N(\mathbf{q}; \mathbf{z}). \quad (23)$$

of

$$\tilde{\mathbf{q}}_{RDE_2}^N = \arg \min_{\mathbf{q} \in \Omega_q} \tilde{J}_{RDE_2}^N(\mathbf{q}; \mathbf{Z})$$

where

$$\tilde{J}_{RDE_2}^N(\mathbf{q}; \mathbf{Z}) = \sum_{m=1}^M \left(\frac{1}{N} \sum_{k=1}^N (Z_k^m - f_s(t_k; \mathbf{q}))^2 \right) \quad (24)$$

with M the number of data sets $\{Z^m\}$ available.

Figure 7 shows the results using synthetic data from the Logistic RDE model with no noise added when varying the number of data sets available and minimizing Equation (24). We note that the median estimate for μ_R and κ are best when only one data set is used, and both approach a limiting value when $M > 20$. The value for μ_R approaches an underestimate with approximately 4-4.5% relative error, while κ approaches an overestimate with approximately 0.8% relative error. We note that as M is increased, the maximum relative error in the estimate for μ_R is decreased as we also saw using the first method. For $M \geq 5$, the maximum percent relative error in κ is between about 1.5-2%. We see a different trend in the estimate for σ_R , however. Even though the median estimate for σ_R also approaches a limit, the minimum estimated value of σ_R is 0 when considering 100 different random choices of M data sets. The limiting value of the median is approximately 0.16 with about a 60% relative error. However, the estimated value of approximately 0 indicates that regardless of how many data sets

Table 14: Parameter Estimates for Logistic RDE Model using Synthetic Data, With No Noise Added, Using Method 2 with 100 Different Random Draws of *One* Data Set.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9927	0.73%	22.81%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	2.95e-08	100%	100%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.98	0.10%	0.21%

Table 15: Parameter Estimates for Bernoulli RDE Model with $\beta = 1.5$ using Synthetic Data, With No Noise Added, and Method 2 with 100 Different Random Draws of *One* Data Set.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9952	0.48%	25.76%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.0034	96.59%	100%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.04	0.004%	0.29%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1.5	1.5131	0.87%	1.93%

Table 16: Parameter Estimates for Bernoulli RDE Model with $\beta = 3$ using Synthetic Data, With No Noise Added, and Method 2 with 100 Different Random Draws of *One* Data Set.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9794	2.05%	26.01%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.0030	96.95%	100%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.04	0.004%	0.14%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
3	3.0356	1.19%	3.16%

Table 17: Parameter Estimates for Bernoulli RDE Model with $\beta = 1.5$ using Synthetic Data and the reparameterization $\tilde{\kappa} = \kappa^\beta$ and $\gamma = \beta\sigma_R^2$, With No Noise Added, and Method 2 with 100 Different Random Draws of *One* Data Set.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9872	1.28%	23.02%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.02	79.81%	100%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.02	0.002%	0.02%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1.5	1.5132	0.88%	1.12%

Table 18: Parameter Estimates for Bernoulli RDE Model with $\beta = 3$ using Synthetic Data and the reparameterization $\tilde{\kappa} = \kappa^\beta$ and $\gamma = \beta\sigma_R^2$, With No Noise Added, and Method 2 with 100 Different Random Draws of *One* Data Set.

μ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1	0.9979	0.21%	26.01%
σ_R			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
0.1	0.0127	87.31%	100%
κ			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
1000	1000.03	0.003%	0.31%
β			
Exact	Median Estimate	Med Perc. Rel. Error	Max Perc. Rel. Error
3	3.0384	1.28%	11.73%

are available, there is a potential to estimate that the random variable R is actually a constant. This result would indicate that the entire model should be a deterministic differential equation and not a random differential equation as assumed.

To summarize, the initial advantage of method 2 over method 1 was that the approximate deterministic equation, J_s , used in the cost function contained *all* the parameters which must be estimated as opposed to the first method in which the standard deviation of R must be calculated after estimating each realization r of $\mathcal{N}(\mu_R, \sigma_R)$ for each of the available data sets. However, using only one data set, it was practically impossible to estimate a nonzero value for σ_R . Reformulating the cost function to include multiple data sets allowed us to estimate a nonzero value for σ_R ; however, regardless of the number of data sets (M) available, in each of the 100 trials, at least one combination of potential data sets resulted in an estimate of σ_R as approximately 0.

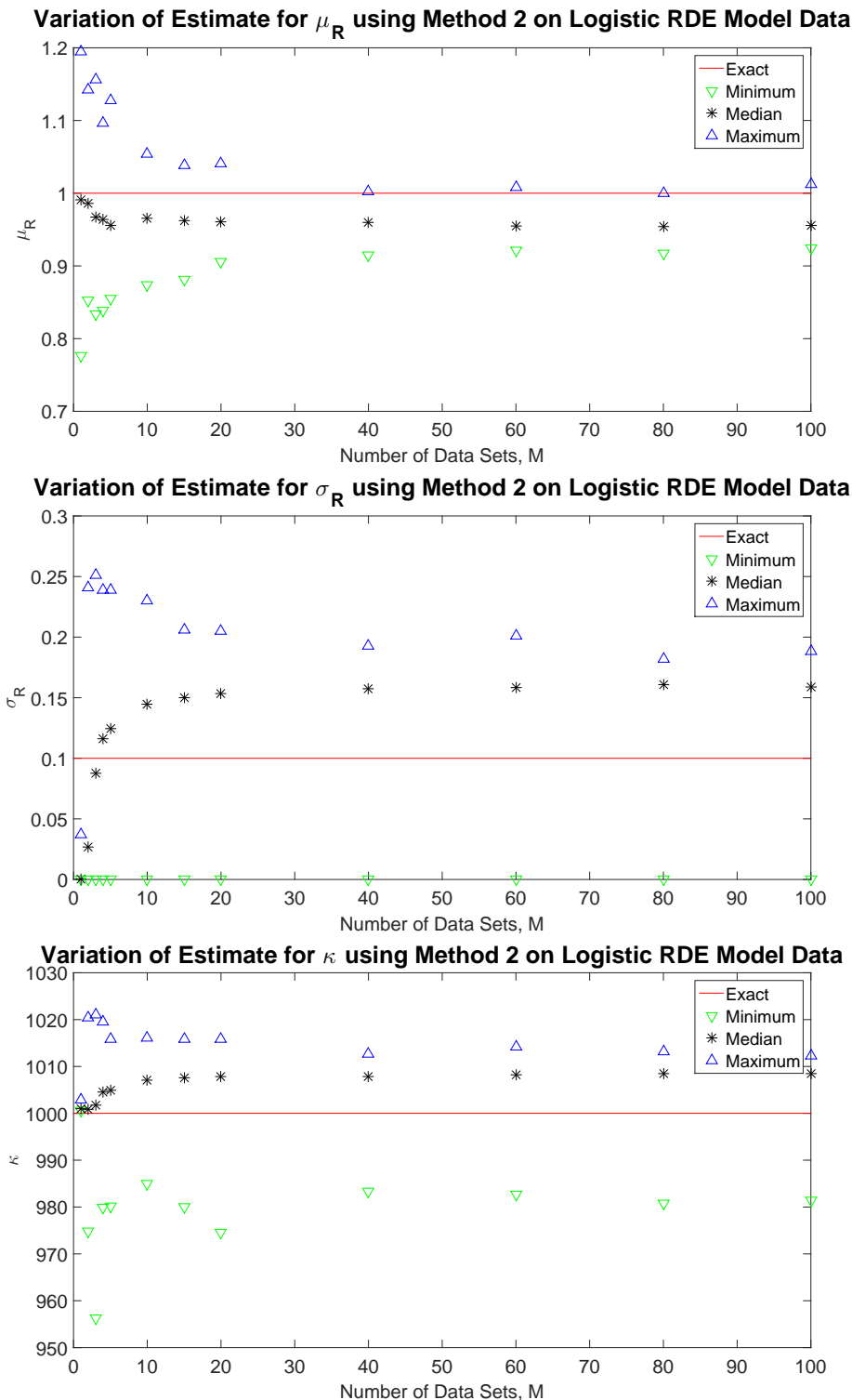


Figure 7: Variation in the estimated values for μ_R , σ_R , and κ by minimizing Equation (24) when the number of data sets M are varied when using synthetic data from the Logistic RDE Model in Equation (3). The red line indicates the exact parameter value. We consider 100 different random draws of M data sets from the 1000 generated. The black star represents the median estimated value across the different groups of M data sets; the blue upward triangle is the maximum estimated parameter value, and the green downward triangle is the minimum estimated value.

6 Parameter Estimation with Experimental Data using Method 1

In this section, we use method 1 for estimating the parameters in a RDE logistic and RDE Bernoulli growth model using longitudinal data collected from algae growth. In a paper by Banks et. al. [3], longitudinal data was collected from four replicate population experiments with green algae, formally known as *Raphidocelis subcapitata*. Four beakers were initially seeded with 1L of Bold’s Basal Media (BBM) and then conditions were set to maintain a chemostat steady-state equilibrium system, constant volume, sufficient oxygen supply, and homogeneous state; details on the experimental collection process can be found in [3]. Two measurements for each of the four replicates were taken twice a day at 9 am and 5 pm daily; the data is shown in Figure 8. Two measurements were taken from each beaker initially so the data could be averaged to minimize human measurement error; however, we consider each data set as an individual data set regardless of whether the data came from the same beaker or a different beaker.

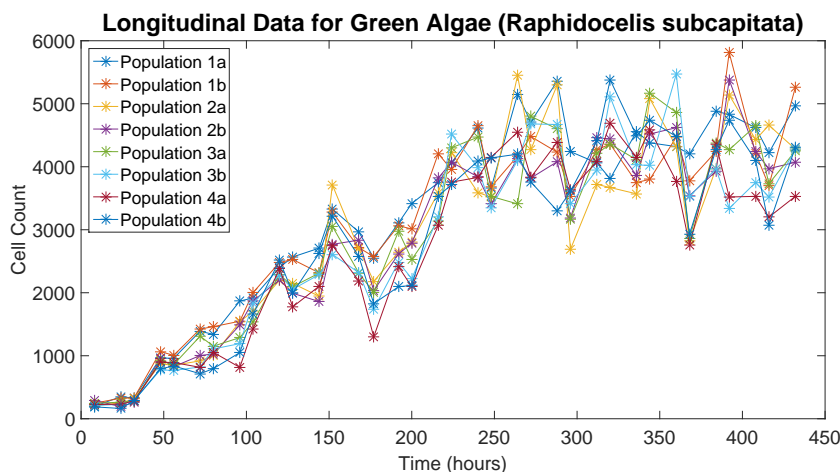


Figure 8: This figure shows the longitudinal data for the two measurements of each of the four replicates of *Raphidocelis subcapitata*.

We first assume the data satisfies the logistic RDE growth model with normally distributed random variable parameters for the growth rate $R \sim \mathcal{N}(\mu_R, \sigma_R)$ and limiting capacity $K \sim \mathcal{N}(\mu_K, \sigma_K)$. In addition, we assume, as done in the paper by Banks et. al. [3], that the initial population is also an unknown parameter which must be estimated; in this case, we assume it is a random variable parameter, $X_0 \sim \mathcal{N}(\mu_{X_0}, \sigma_{X_0})$. We note that we initially assume all parameters are random variable parameters; however, for a parameter Q , if the estimated value of σ_Q is small relative to the value μ_Q , we can instead declare this parameter a constant parameter in the model. To estimate the individual data parameters for the sample deterministic system required in method 1, we use an initial guess of $r = 0.02$, $\kappa = 4100$, and $x_0 = 414$. The estimates for the eight individual data sets are given in Table 19. We notice the variability in both the estimates across the four beakers but also the variation in estimates even using data from the same beaker.

Table 19: Parameter Estimates in the Logistic Sample Deterministic Model for Individual Algae Data Sets.

Data Set	r	κ	x_0
1a	0.0187	4611.42	399.17
1b	0.0179	4398.34	430.31
2a	0.0189	4281.04	320.84
2b	0.0171	4322.88	360.53
3a	0.0174	4348.04	358.31
3b	0.0184	4210.42	291.76
4a	0.0192	3988.83	248.42
4b	0.0180	4262.62	308.07

In the estimate of the statistical parameters for the random variable model parameters, we analyze the experimental data in the same manner as we did with the synthetic data. We examine how the choice of the *number* of data sets and *which* data sets were used affect the estimated values for both the mean and standard deviation of each random variable parameter. Since we only have eight total data sets, we consider each combination of M data sets, $M = 2, \dots, 8$ in the calculation of μ_Q (Equation (8)) and σ_Q (Equation (9)) for each parameter Q . The results are shown in Figure 9. As expected, there is a larger spread in possible estimates when only a few data sets are used; however the median estimate is fairly constant across M . The estimated values for all statistical parameters when using all eight data sets are given in Table 20. Figure 10 illustrates the RDE model solution together with the data under the assumptions that different model parameters are random variable parameters or constant parameters. We note that $\sigma_R \approx 0.04\mu_R$, $\sigma_K \approx 0.04\mu_K$ and $\sigma_{X_0} \approx 0.17\mu_{X_0}$. The RDE solution or collection of solutions trajectories to the sample deterministic equation appears to ‘cover’ the data best when all model variables are assumed to be random model parameters (first plot) or when the limiting capacity K and initial cell count X_0 are assumed random (second row, second column plot). This comparison between the data and model simulations is similar to what is illustrated in Figure 5 with synthetic data when the value of σ_R was underestimated. However, if the noise level \mathcal{E} is increased to approximately 10% of the limiting capacity, the model simulations with noise added encompasses the data near the limiting capacity however is too large for the initial cell count and can cause negative or zero cell count at the lower cell count population; see Figure 11. Therefore, it may be more appropriate to assume a nonconstant variance statistical model.

In fact, in the paper by Banks et. al. [3], they demonstrated that the algae data exhibited proportional noise, or nonconstant variance noise, of the form

$$Z_j = h(t_j; \mathbf{q}_0) + h^\gamma(t_j; \mathbf{q}_0)\mathcal{E}_j, \quad \text{Var}(Z_j) = \sigma_0^2 h^{2\gamma}(t_j; \mathbf{q}_0) \quad (25)$$

where the best choice for γ was shown to be $\gamma = 1$. This is different from the assumptions made when we formulated method 1 above. We had initially made the assumption that the statistical model had constant variance noise, i.e.,

$$Z_j = h(t_j; \mathbf{q}_0) + \mathcal{E}_j, \quad \text{Var}(Z_j) = \sigma_0^2.$$

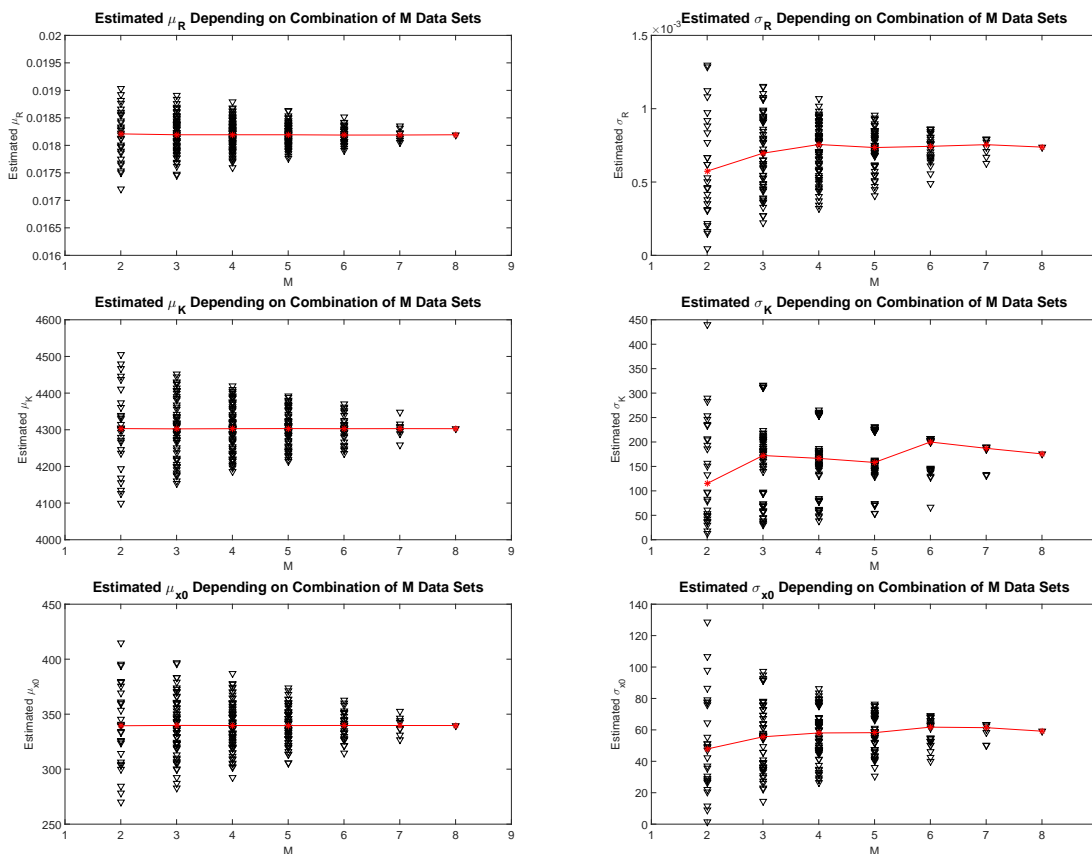


Figure 9: Variation in the estimated values for μ_Q (first column) and σ_Q (second column) in the RDE logistic growth model when the number of data sets M are varied when using experimental longitudinal data of algae cell count along with Equations (3), (8), and (9). The solid red line indicates the median trend across the number of data sets. The first row shows the trend in estimated statistical parameters for $Q = R$; second row is for $Q = K$, and third row is for X_0 .

To incorporate the nonconstant variance, for each individual data set, we seek a generalized least squares estimate

$$\mathbf{q}_{GLS}^N = \arg \min_{\mathbf{q} \in \Omega_q} J_{GLS}^N(\mathbf{q}; \mathbf{Z})$$

with realization

$$\hat{\mathbf{q}}_{GLS}^N = \arg \min_{\mathbf{q} \in \Omega_q} J_{GLS}^N(\mathbf{q}; \mathbf{z})$$

where the cost function J_{GLS}^N is defined by

$$J_{GLS}^N(\mathbf{q}; \mathbf{Z}) = \frac{1}{N} \sum_{k=1}^N w_k (Z_k - f(t_k; \mathbf{q}))^2 \quad (26)$$

with $w_k = f^{-2\gamma}(t_k, \mathbf{q})$ and f the output from the sample deterministic system. Minimizing $J_{GLS}^N(\mathbf{q}; \mathbf{z})$ using the built-in Matlab function *fminsearch* with the same initial guesses as above, we obtain the individual data parameter estimates for the sample

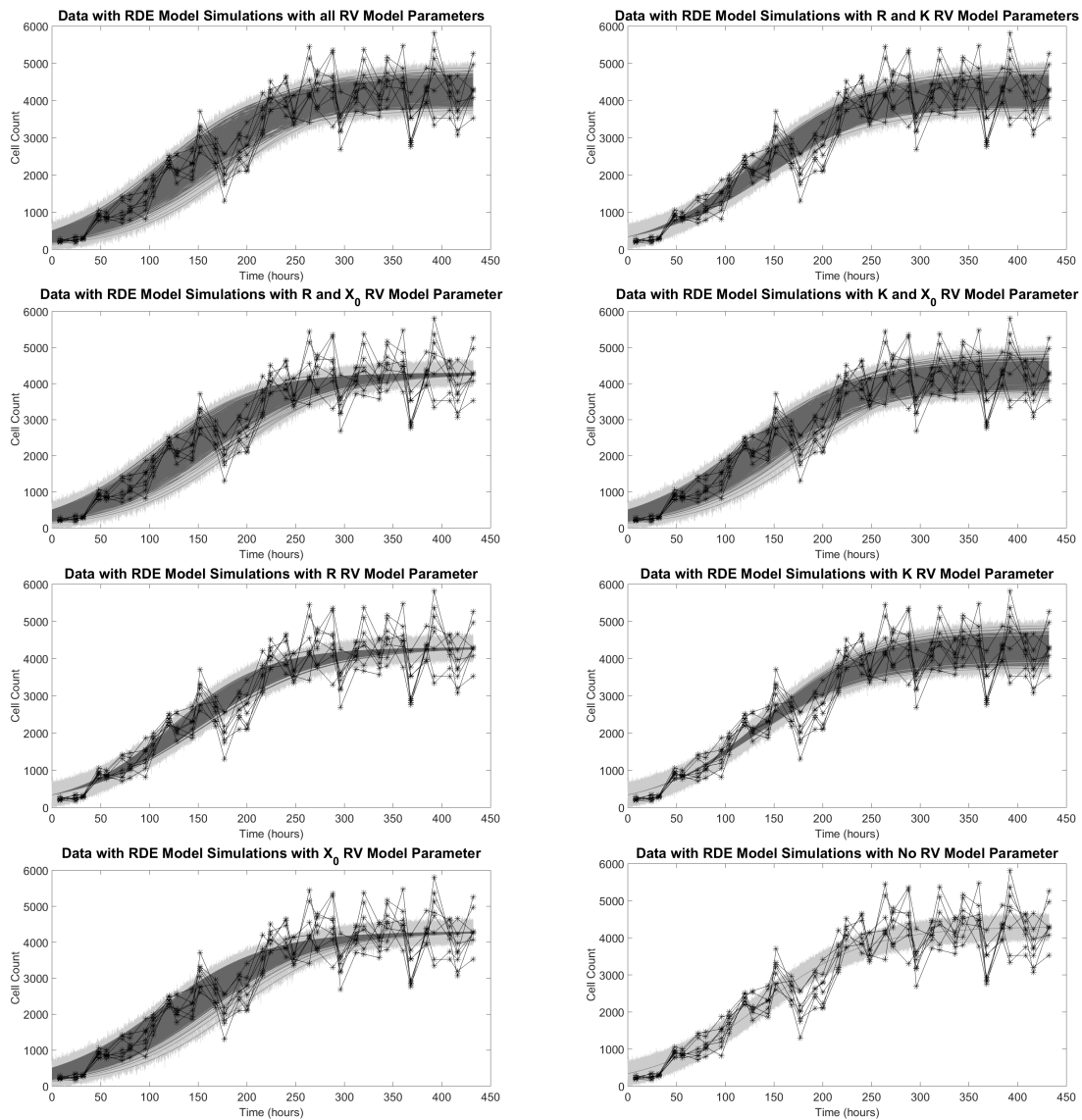


Figure 10: Each plot shows 1000 simulations of the logistic RDE model (dark grey) with the data (black) and with constant variance noise $\mathcal{E} \sim \mathcal{N}(0, 100^2)$ (light grey). The first plots assumes $R \sim \mathcal{N}(\mu_R, \sigma_R)$, $K \sim \mathcal{N}(\mu_K, \sigma_K)$, and $X_0 \sim \mathcal{N}(\mu_{X_0}, \sigma_{X_0})$ with estimated values given in Table 20. The second plot assumes both R and K are random variable model parameters but x_0 is a constant parameter. The third plot assumes both R and X_0 are random variable model parameters but κ is constant. The fourth plot assumes both K and X_0 are random variable model parameters but r is constant. The fifth plot shows the simulations assuming only R is a random variable model parameter. The sixth plot shows the simulations if only K is a random variable parameter. The seventh plot shows the simulations if only X_0 is a random variable parameter. Finally, the last plots shows the deterministic solutions assuming all model parameters are constant.

Table 20: Statistical Parameter Estimates for Random Variable Model Parameters in the Logistic RDE Model using Eight Algae Cell Count Data Sets.

Parameter	Estimate
μ_R	$1.82 \cdot 10^{-2}$
σ_R	$7.38 \cdot 10^{-4}$
μ_K	4302.95
σ_K	175.67
μ_{X_0}	339.68
σ_{X_0}	59.18

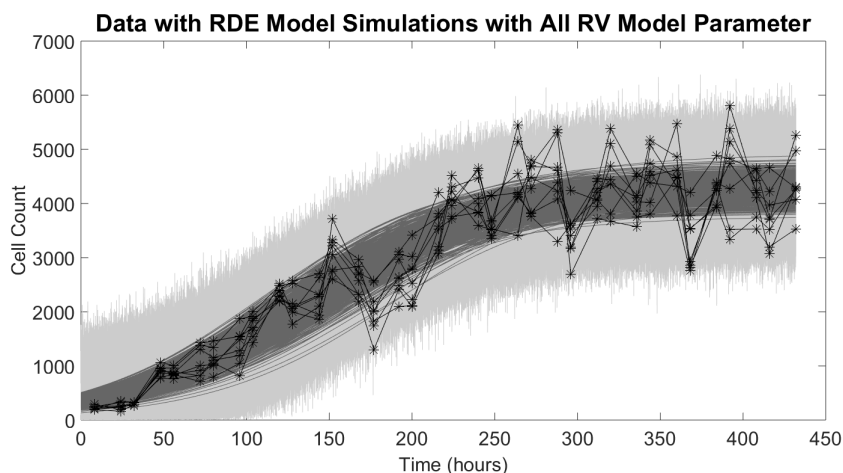


Figure 11: This plot shows 1000 simulations of the logistic RDE model (dark grey) with the data (black) and with increased constant variance noise $\mathcal{E} \sim \mathcal{N}(0, 400^2)$ (light grey) with random variable parameters $R \sim \mathcal{N}(\mu_R, \sigma_R)$, $K \sim \mathcal{N}(\mu_K, \sigma_K)$, and $X_0 \sim \mathcal{N}(\mu_{X_0}, \sigma_{X_0})$.

deterministic system in Table 21. We calculate the statistical parameters in the same way as before and obtain similar trends in the results as in the ordinary least squares method. The estimated values when using all eight data sets is given in Table 22. Figure 12 illustrates the RDE model solution with noise added using the estimated random variable parameters defined by the statistical parameters given in Table 22. We note that the RDE solution is similar to the one found when assuming constant variance; however, there appears to be a slightly better estimate of X_0 , a smaller relative variance in the limiting capacity, σ_K^2 , and a larger relative variance in the growth parameter σ_R^2 .

We also fit the RDE Bernoulli growth model assuming nonconstant variance in the data (i.e., the statistical model is given by Equation (25)). We minimize J_{GLS}^N in Equation (26) for the sample deterministic Bernoulli model for each data set and obtain the estimates in Table 23. Comparing parameter estimates for the logistic sample deterministic system in Table 21 and for the Bernoulli sample deterministic system in Table 23, we note that the estimates in the Bernoulli model for r and κ are smaller than those in the logistic model while the initial data x_0 is larger in the

Table 21: Parameter Estimates in the Logistic Sample Deterministic Model for Individual Algae Data Sets When Minimizing J_{GLS}^N (Eq. (26)).

Data Set	r	κ	x_0
1a	0.0266	4394.85	224.49
1b	0.0253	4224.49	258.68
2a	0.0245	4237.32	209.42
2b	0.0211	4177.31	271.39
3a	0.0222	4217.35	257.20
3b	0.0238	4104.45	200.28
4a	0.0214	3957.20	243.74
4b	0.0248	4124.85	185.06

Table 22: Statistical Parameter Estimates for Random Variable Model Parameters in the Logistic RDE Model using Eight Algae Cell Count Data Sets when Assuming Nonconstant Variance in the Statistical Model (i.e., when minimizing J_{GLS}^N).

Parameter	Estimate
μ_R	$2.37 \cdot 10^{-2}$
σ_R	$1.97 \cdot 10^{-3}$
μ_K	4179.73
σ_K	126.05
μ_{X_0}	231.28
σ_{X_0}	31.18

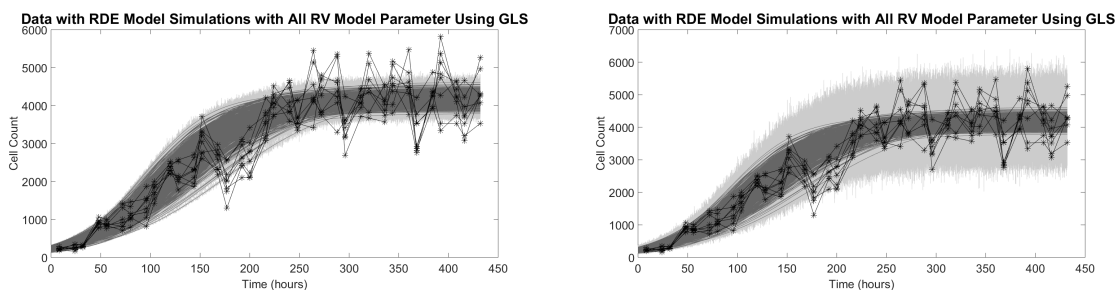


Figure 12: Each plot shows 1000 simulations of the logistic RDE model (dark grey) with the data (black) and with 2.5% relative noise or nonconstant variance noise (first plot) and 10% relative noise (second plot). In both plots, we assume $R \sim \mathcal{N}(\mu_R, \sigma_R)$, $K \sim \mathcal{N}(\mu_K, \sigma_K)$, and $X_0 \sim \mathcal{N}(\mu_{X_0}, \sigma_{X_0})$ with estimated values given in Table 22.

Table 23: Parameter Estimates in the Bernoulli Sample Deterministic Model for Individual Algae Data Sets When Minimizing J_{GLS}^N (Eq. (26)).

Data Set	r	κ	β	x_0
1a	0.0247	4386.71	1.1553	236.56
1b	0.0232	4215.55	1.1721	273.21
2a	0.0228	4222.27	1.1642	219.26
2b	0.0195	4158.98	1.1647	282.61
3a	0.0205	4206.92	1.1536	269.69
3b	0.0221	4088.01	1.1691	209.64
4a	0.0198	3946.38	1.1538	255.36
4b	0.0231	4108.87	1.1664	193.61

Table 24: Statistical Parameter Estimates for Random Variable Model Parameters in the Bernoulli RDE Model using Eight Algae Cell Count Data Sets when Assuming Nonconstant Variance in the Statistical Model.

Parameter	Estimate
μ_R	$2.20 \cdot 10^{-2}$
σ_R	$1.84 \cdot 10^{-3}$
μ_K	4166.71
σ_K	127.41
μ_β	1.1624
σ_β	$7.20 \cdot 10^{-3}$
μ_{X_0}	242.49
σ_{X_0}	32.74

Bernoulli model. Moreover, β is estimated to be slightly larger than 1 where $\beta = 1$ indicates the logistic model. Using these estimates with Equations (8) and (9), we obtain the estimates for the statistical parameters given in Table 24 when using all eight data sets. The trend when using a subset of M data sets is shown in Figure 13. In these results, we note that $\sigma_R \approx 0.08\mu_R$, $\sigma_K \approx 0.03\mu_K$, $\sigma_{X_0} \approx 0.13\mu_{X_0}$ and $\sigma_\beta \approx 0.006\mu_\beta$. Therefore, the relative value of σ_β is only approximately 0.6% the value of μ_β , and hence, we assume β is a constant parameter in the Bernoulli RDE growth model. The plot in Figure 14 shows 1000 simulations of the Bernoulli RDE growth model assuming random variable parameters R , K and X_0 with a constant parameter β . We also include the model with 10% relative noise added in light grey. This plot does not appear visually different that the plot for the logistic RDE model in Figure 12.

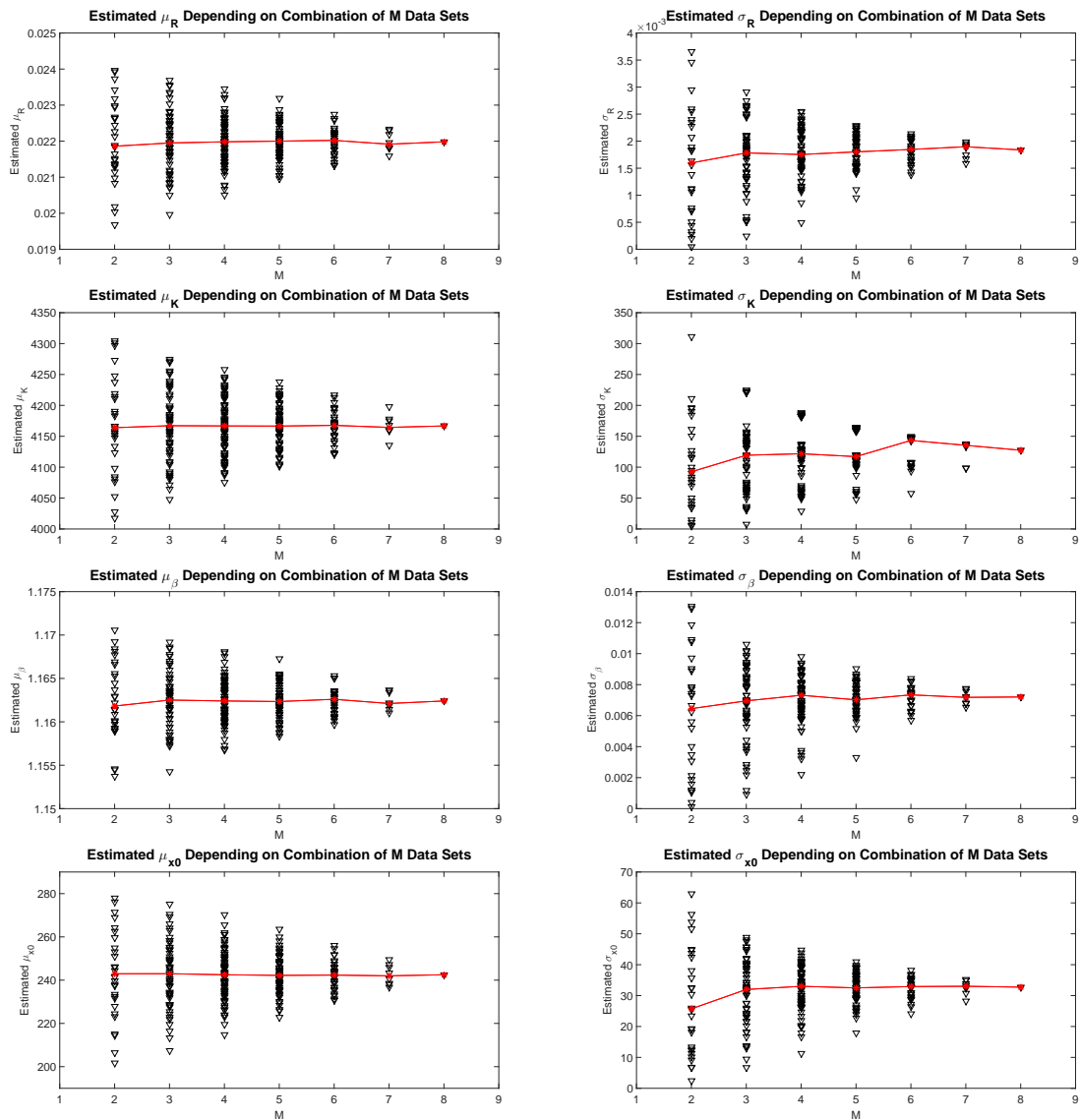


Figure 13: Variation in the estimated values for μ_Q (first column) and σ_Q (second column) in the RDE Bernoulli growth model when the number of data sets M are varied when using experimental longitudinal data of algae cell count along with Equations (3), (8), and (9). The solid red line indicates the median trend across the number of data sets. The first row shows the trend in estimated statistical parameters for $Q = R$; second row is for $Q = K$, and third row is for X_0 .

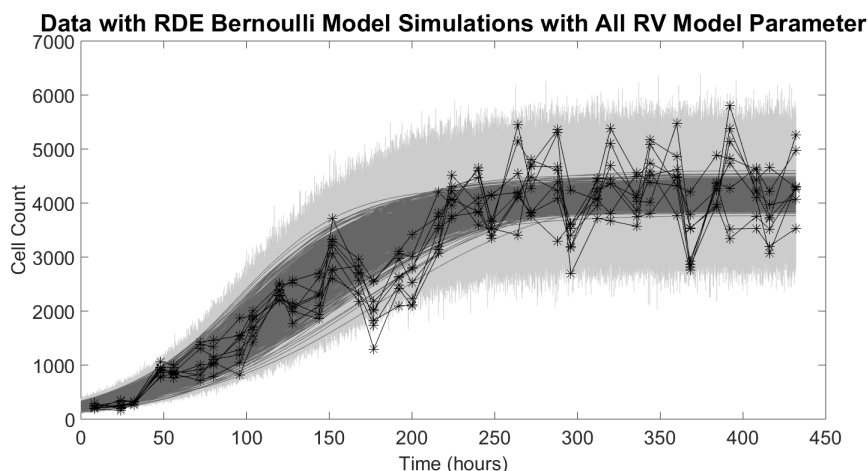


Figure 14: This plot shows 1000 simulations of the Bernoulli RDE model (dark grey) with the data (black) and with 10% relative noise or nonconstant variance noise under the assumption that $R \sim \mathcal{N}(\mu_R, \sigma_R)$, $K \sim \mathcal{N}(\mu_K, \sigma_K)$, $X_0 \sim \mathcal{N}(\mu_{X_0}, \sigma_{X_0})$ and $\beta = \mu_\beta$ with estimated values given in Table 24.

7 Conclusions

In this paper, we investigated two methods for parameter estimation in a RDE model. The first method was based on approximating the RDE with its sample deterministic equation while the second method was based on first deriving the pointwise equivalent SDE model for the RDE model and then approximating the SDE model with a deterministic model. In both parameter estimation techniques, deterministic parameter estimation techniques were used once the appropriate deterministic system was identified.

In the first method, constant parameters for the deterministic system are estimated. These constants are then used to estimate the statistical parameters for the random variable distributions. We tested this method on two nested RDE models in which there was one random variable parameter. Depending on the number of data sets available, some of the statistical parameters were harder to estimate than others. We tested the method assuming either a normally distributed random variable parameter or a random variable parameter with a 2-parameter Weibull distribution. When testing the method with a normally distributed random variable parameter, depending on the random choice of data sets and how many data sets were used in the estimation, it could be difficult to obtain an extremely accurate estimate for the standard deviation. However, this makes intuitive sense since if the data sets behave similarly, then the variance (and hence the standard deviation) may be estimated smaller than the actual standard deviation. Figure 5 gives an illustration of the collection of solution trajectories if the standard deviation is estimated smaller than the actual standard deviation. When estimating the random variable parameter assumed to have a Weibull distribution, the shape parameter B was slightly more difficult to estimate; however, even though there is the potential for a poor estimate, in most cases, the relative error in both statistical parameters was less than 10% when $M = 15$ or $M = 100$. In summary, the first

parameter estimation technique was a viable method for the models analyzed in this paper.

The second method for parameter estimation was a technique which was derived for a select type of RDE models. It is assumed the random variable parameters in the RDE model behave according to a normal distribution and the RDE model has an appropriate form or can be transformed into an appropriate form. Next a pointwise equivalent SDE is derived which is then approximated by a deterministic model. In this technique, all statistical parameters for the random variable and all constant model parameters are contained in the approximate deterministic system and can be estimated simultaneously. In order to estimate a nonzero value for the standard deviation, it was shown that we must use more than one data set. Although the median estimate approaches a limiting value as M increases, in each of the 100 trials, there was at least one trial which resulted in an estimate of approximately 0 for the standard deviation regardless of how many data sets were used. If one were ‘unlucky’ in the choice of *which* M data sets were available, one could obtain a result which would indicate the random variable should be a constant and the model was in fact a deterministic differential equation instead of a random differential equation. Therefore, *this technique (method 2) is not a viable method for the example models presented in this paper.*

Finally, we applied the first method in estimating the random variable model parameters in both the logistic and Bernoulli RDE models using experimental data from algae growth. The resulting RDE solution using the random variable parameter estimates obtained from method 1 appear plausible when compared to the eight data sets. We also illustrated the differences in the RDE solution assuming one or multiple random variable parameters in the model and compared each scenario with the eight data sets. Visually, the RDE solution in which all model parameters were assumed to be random variables appears to be the best model for the data; however, unless the noise has a large enough variance, the resulting model with noise did not encompass the entire range in the cell count data. If we assumed a statistical model with constant variance, when the constant variance was large enough to encompass the large cell count, it was too large for the smaller cell count. Therefore, we reformulated method 1 to incorporate a nonconstant variance. We illustrated the slight differences in the estimated values and resulting solution curve with noise. In this case, assuming a 10% *relative* noise in the model simulation resulted in simulations which covered the range in the data while accounting for the size of the cell count. Furthermore, we approximated the random variable model parameters in the Bernoulli RDE model as well using the experimental data under the assumption of nonconstant variance in the statistical model. Similar results were found as with the logistic model. In general, we have shown that method 1 was a viable approach for estimating parameters in a random differential equation model not only using synthetic data but with experimental data as well.

Acknowledgements

This research was supported in part by the National Institute on Alcohol Abuse and Alcoholism under grant number 1R01AA022714-01A1, and in part by the Air Force

Office of Scientific Research under grant number AFOSR FA9550-15-1-0298.

References

- [1] Linda Allen. *An Introduction to Stochastic Processes with Applications to Biology*. Taylor and Francis Group, LLC, 2011.
- [2] H. T. Banks, Shuhua Hu, and W. Clayton Thompson. *Modeling and Inverse Problems in the Presence of Uncertainty*. CRC Press, 2014.
- [3] H.T. Banks, Elizabeth Collins, Kevin Flores, Prayag Pershad, Michael Stemkovski, and Lyric Stephenson. Standard and proportional error model comparison for logistic growth of green algae (*Raphidocelis subcapiala*). *Applied Mathematical Letters*, 64:213–222, 2017.
- [4] H.T. Banks and S. Hu. Nonlinear stochastic Markov processes and modeling uncertainty in populations. *Mathematical Biosciences and Engineering*, 9(1):1–25, 2012.
- [5] H.T. Banks and Michele L. Joyner. Deterministic methodology for comparison of nested stochastic models. *CRSC-TR16-13, NCSU, Raleigh, NC, November*, 2016.
- [6] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*, volume 282. John Wiley & Sons, 2009.
- [7] Avner Friedman. *Stochastic Differential Equations and Applications*. Dover Publications, 2006.
- [8] Thomas C. Gard. *Introduction to Stochastic Differential Equations*. Marcel Dekker, 1988.
- [9] Mircea Grigoriu. *Stochastic Calculus: Applications in Science and Engineering*. Springer Science & Business Media, 2002.
- [10] Thomas G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7(1):49–58, 1970.
- [11] Zeev Schuss. *Theory and Applications of Stochastic Processes: An Analytical Approach*, volume 170. Springer Science & Business Media, 2009.
- [12] Tsu T. Soong. *Random Differential Equations in Science and Engineering*. Elsevier, 1973.
- [13] Howard M. Taylor and Samuel Karlin. *An Introduction to Stochastic Modeling*. Academic Press, 2014.
- [14] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, Oxford, 2010.

H. T. Banks
Center for Research in Scientific Computation (CRSC)
North Carolina State University
Raleigh, NC, United States
htbanks@ncsu.edu

M.L. Joyner
Dept of Mathematics and Statistics
East Tennessee State University
Johnson City, TN 37614
joynerm@etsu.edu

Received 14.12.2016, Accepted 13.02.2017