8-2012

Latin Hypercube Sampling and Partial Rank Correlation Coefficient Analysis Applied to an Optimal Control Problem

Boloye Gomero
bgomero@utk.edu

Recommended Citation

To the Graduate Council:

I am submitting herewith a thesis written by Boloye Gomero entitled "Latin Hypercube Sampling and Partial Rank Correlation Coefficient Analysis Applied to an Optimal Control Problem." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Mathematics.

<div align="right">Dr. Suzanne Lenhart, Major Professor</div>

We have read this thesis and recommend its acceptance:

Dr. Judy Day, Dr. Yulong Xing

<div align="right">Accepted for the Council:</div>
<div align="right">Carolyn R. Hodges</div>

<div align="right">Vice Provost and Dean of the Graduate School</div>

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Boloye Gomero entitled "Latin Hypercube Sampling and Partial Rank Correlation Coefficient Analysis Applied to an Optimal Control Problem." I have examined the final paper copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Mathematics.

_____
Suzanne Lenhart, Major Professor

We have read this thesis
and recommend its acceptance:

_____
Yulong Xing

_____
Judy Day

Accepted for the Council:

_____
Carolyn R. Hodges
Vice Provost and Dean of the Graduate School

To the Graduate Council:

I am submitting herewith a thesis written by Boloye Gomero entitled "Latin Hypercube Sampling and Partial Rank Correlation Coefficient Analysis Applied to an Optimal Control Problem." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Mathematics.

Suzanne Lenhart, Major Professor

We have read this thesis
and recommend its acceptance:

Yulong Xing
_____

Judy Day
_____

Accepted for the Council:

Carolyn R. Hodges
_____

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Latin Hypercube Sampling and Partial Rank Correlation Coefficient Analysis Applied to an Optimal Control Problem

A Thesis Presented for

The Master of Science

Degree

The University of Tennessee, Knoxville

Boloye Gomero

August 2012

To mother and father, Mr. Patrick Gomero and Mrs. Gift Gomero.

Acknowledgements

Isaiah 46:3-4 (The Message Bible)

"Listen to me, family of Jacob,

everyone that's left of the family of Israel.

I've been carrying you on my back

from the day you were born,

And I'll keep on carrying you when you're old.

I'll be there, bearing you when you're old and gray.

I've done it and will keep on doing it,

carrying you on my back, saving you."

Abstract

Latin Hypercube Sampling/Partial Rank Correlation Coefficient (LHS/PRCC) sensitivity analysis is an efficient tool often employed in uncertainty analysis to explore the entire parameter space of a model. Despite the usefulness of LHS/PRCC sensitivity analysis in studying the sensitivity of a model to the parameter values used in the model, no study has been done that fully integrates Latin Hypercube sampling with optimal control analysis.

In this thesis, we couple the optimal control numerical procedure to the LHS/PRCC procedure and perform a simultaneous examination of the effects of all the LHS parameter on the objective functional value. To test the effectiveness of our procedure, we examine the sensitive parameters in a deterministic ordinary differential equations cholera model having seven human compartments and two bacterial compartments. Our procedure cuts down on simulation time and helps us perform a more comprehensive analysis of the influential parameters in the cholera model, than would be possible otherwise.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A mathematical model is an abstraction of a real-world problem into a mathematical problem. In particular, we create disease models as tools for discovering underlying patterns in epidemiology. The model, thus, becomes a tool for studying disease epidemiology along with the effect and targeting of various interventions. When we create models, we must make simplifications and assumptions both about the way the model is built and about the values of the parameters required. Moreover, due to the uncertainty that may accompany choices for parameter values, it is often important to understand the effects of model parameter values on specific outcome measures (output). Uncertainty in the parameter values chosen introduces variability to the model's prediction of resulting dynamics. The more uncertain parameters there are, the more significant the variability introduced. Thus, a sensitivity analysis is often performed to assess this variability in the model predictions.

Latin Hypercube Sampling/Partial Rank Correlation Coefficient (LHS/PRCC) sensitivity analysis is an efficient tool often employed in uncertainty analysis to explore the entire parameter space of a model with a minimum number of computer simulations [10]. It involves the combination of two statistical techniques, Latin Hypercube Sampling (LHS), which was first introduced by McKay et al. in 1979 [2] and further developed by Iman et al. [10], and Partial Rank Correlation Coefficient

(PRCC) analysis. The goal of LHS/PRCC sensitivity analysis is to identify key parameters whose uncertainties contribute to prediction imprecision and to rank these parameters by their importance in contributing to this imprecision.

The LHS procedure is implemented by dividing the range of values for a given parameter into equally probable intervals. The LHS scheme is a so-called stratified scheme whereby probability distributions are assigned to parameters, the intervals in the distribution are divided into equiprobable regions, and these intervals are then each sampled without replacement. One advantage of the LHS procedure is that the parameters are sampled independently of one another.

According to McKay et al. [2], the LHS method performs an un-biased estimate of the average model output. The LHS sampling procedure is different from random sampling because in LHS sampling, random samples can be taken one at a time, while taking into account the row and column of the previously generated sample points. Thus, compared to simple random sampling schemes, the LHS sampling procedure requires fewer samples to achieve the same level of accuracy.

Correlation is a statistical technique used to measure the strength of the relationship between the outcome measures and the parameters in a model. Using the residuals obtained from the regression procedure, Partial Correlation characterizes the linear relationship between the LHS parameters and the outcome measure after discounting the linear effects of the LHS parameters (inputs), x_j, on the outcome measure (outputs), y [18]. PRCC is a robust sensitivity measure for nonlinear but monotonic relationships between x_j and y, as long as little to no correlation exists between the inputs [18]. Compared to ordinary Partial Correlation Coefficient procedure, PRCC is considered to be more powerful at determining the sensitivity of a parameter that is strongly monotonic yet highly nonlinear [4]. If we use regression coefficients obtained from the raw sample values of each parameter along with the corresponding raw outcome measure, we obtain a Pearson correlation. If, on the other hand, we use regression coefficients from rank-transformed values instead, we obtain a Spearman or rank correlation coefficient. Note that the estimated PRCC and the

standardized regression coefficient are considered to be essentially the same when using ranks, and both measures exhibit the same pattern of sensitivity ranking [7]. Rank-transformation is done to reduce the effect of non-linear data, and it works best when there is a monotonic relationship between the outcome measure and the parameter of interest [8]. Specifically, when rank transformed data is used, the rank correlation coefficient will indicate the level of monotonicity between the LHS parameters and the outcome measures [5].

Several previous studies have used the LHS/PRCC procedure on ordinary differential equations (ODE) systems without controls and then varied those 'most sensitive' parameters in an investigation of optimal control problems [16, 15]. However, we are not aware of any work using the LHS/PRCC procedure with optimal control of ODE systems wherein the outcome measure used is the value of objective functional (at the optimal control). In our study, we couple the optimal control numerical procedure to the LHS/PRCC procedure and perform a simultaneous examination of the effects of all the LHS parameters on the objective functional value evaluated at the optimal control and corresponding states. Our procedure cuts down on simulation time and helps us perform a more holistic, comprehensive and accurate analysis than would be possible if we studied the parameters one at a time.

Our study of LHS/PRCC parameter sensitivity analysis will be presented in the next three chapters. In Chapter 2, we will outline the procedure for using a combination of methods developed by Sally Blower and Denise Kirschner [18] to perform the LHS/PRCC sensitivity analysis. In Chapter 3, we will implement LHS/PRCC parameter sensitivity analysis in a deterministic cholera epidemic model. Finally, in the Chapter 4, we will add a control to our model. We will then couple the LHS/PRCC sensitivity analysis procedure to the optimal control analysis procedure with the goal of studying the effects of the LHS parameters on the objective functional value of our model.

Chapter 2

Explaining the LHS/PRCC Procedure

The following sections outline the steps in the Latin Hypercube Sampling/Pearson Partial Rank Correlation Coefficient (LHS/PRCC) procedure. The examples and illustrations in this chapter are done for a hypothetical model to illustrate the procedure.

2.1 Steps for Sampling the LHS Parameters

To perform the Latin Hypercube Sampling, the following steps are included:

1. Start out with a mathematical model of interest. The LHS/PRCC procedure can be applied to various types of mathematical models, including deterministic or stochastic models, with continuous or discrete features.

2. List the parameters for the model and their corresponding values. Some of the parameter values will be known with certainty and others will not.

3. Identify the uncertain parameters in your parameter list. For some of these, we might know a possible range where the exact values might fall. In particular, we

define Baseline Values as the values of the parameters we know with certainty as well as the middle (or near the middle) of the range of values for the parameters whose exact values we are unsure of.

4. Next, we decide on the sample size for our analysis. The sample size will be determined by the number of simulations we intend to run. Suppose we decide to do N model simulations (or runs) for our analysis. Also suppose there are K uncertain parameters, v_i, $1 \leq i \leq K$. Then the parameter space for the uncertain parameters would be defined by K dimensions. Note that the choice of N is not arbitrary. If N is the number of simulations, the following inequality has to be satisfied: $N > (4/3)K$ [11, 2].

5. Each of the K dimensions will correspond to an uncertain parameter and the length of each dimension is determined by the number of runs, N, chosen. For each uncertain parameter, each of the N input values would be selected/determined by the LHS sampling scheme.

6. To implement this LHS sampling scheme, we begin by specifying a probability density or distribution function (pdf) for each uncertain parameter. This way, the variability in the pdf becomes a direct measure of the variability of the uncertain parameter. Each specified pdf describes a range of possible values and the probability of occurrence of any specific value for the parameter. An example of this could be the case where we specify a minimum and maximum for a parameter and use those values to compute a uniform distribution for the variable.

   Note that the chosen probability density functions are determined by an observed distribution of a plot of available data. For example, upon observation, we may notice that our data are represented by one of the following:

   (a) Left skewed or right skewed: This is observed when one part of the interval of values has a higher probability of occurrence than the other.

(b) Multi-modal Weibull distribution: In this case, more than one region in the interval has a definite probability of occurrence and we wish to study those regions simultaneously.

(c) Triangular: We use a triangular distribution function when we wish to reflect the expectation that values close to the peak of the triangle distribution pattern are those considered to be more likely to occur.

(d) Uniform distribution: This occurs when the probability of occurrence of any part of the interval is relatively even. In the case where data is not available, the uniform distribution is most appropriate to use as the default [18]. In this case, each interval in the pdf has an equal probability of being sampled. Once we have determined the probability density function for all of the uncertain parameters, we proceed with selecting the input values for each of the N numerical simulations.

7. To sample the values for each parameter, each probability density function is divided into N non-overlapping equiprobable intervals. As a result, the sampling distribution of the values for each parameter reflects the shape of the particular pdf.

8. Every equiprobable interval of each parameter is then randomly sampled once. The frequency of the selection of possible values of each parameter is determined by the probability of occurrence in the pdf. Each parameter is sampled independently; hence, the parameters are uncorrelated.

9. Once this step is complete, each of the K uncertain parameters, v_i, $1 \leq i \leq K$, will have N values. Hence, we store the sampled values in an $N \times K$ table/matrix. Note that the values for each column are random. They are not arranged in any particular order according to magnitude. A sample LHS matrix/table is illustrated in Figure 2.1. Note that for this table, each column has entry, $(r_j, \ v_i)$, or the jth sampled random value of the ith uncertain

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	...
r_1	0.002146	0.00155	0.000858	0.64906	0.799178	4.90E-05	0.042043	0.520353	...
r_2	0.009791	0.001116	0.000861	0.814074	0.364338	0.000205	0.026748	0.659609	...
r_3	0.004744	0.001033	0.000347	0.445106	0.016567	9.93E-05	0.025842	0.915153	...
r_4	0.00495	0.002887	0.000749	0.576221	0.867036	0.000465	0.034892	0.728892	...
r_5	0.024555	0.002407	0.000631	0.406264	0.672348	7.46E-05	0.011587	0.680772	...
r_6	0.024106	0.001336	0.000513	0.31633	0.315899	0.000498	0.038492	0.87865	...
.	.	.	.	.	.	.	.	.	...
.	.	.	.	.	.	.	.	.	...
.	.	.	.	.	.	.	.	.	...

Figure 2.1: A Sample LHS Matrix/Table.

parameter, where $1 \leq i \leq K$, $1 \leq j \leq N$. Thus, each row in the matrix comprises K random values, each corresponding to a specific LHS parameter, respectively.

2.2 Interpreting the Monotonicity Plots

As we discussed briefly in Chapter 1, a partial rank correlation measures the strength of a relationship between two variables while controlling the effect of the other variables, and it does this by indicating the degree of monotonicity between a specific input and corresponding output variable. Thus, only outcome measures having a monotonic relationship with input variables should be chosen for this type of sensitivity analysis. Consequently, we first verify that a monotonic relationship exists between each outcome variable chosen and the LHS parameters.
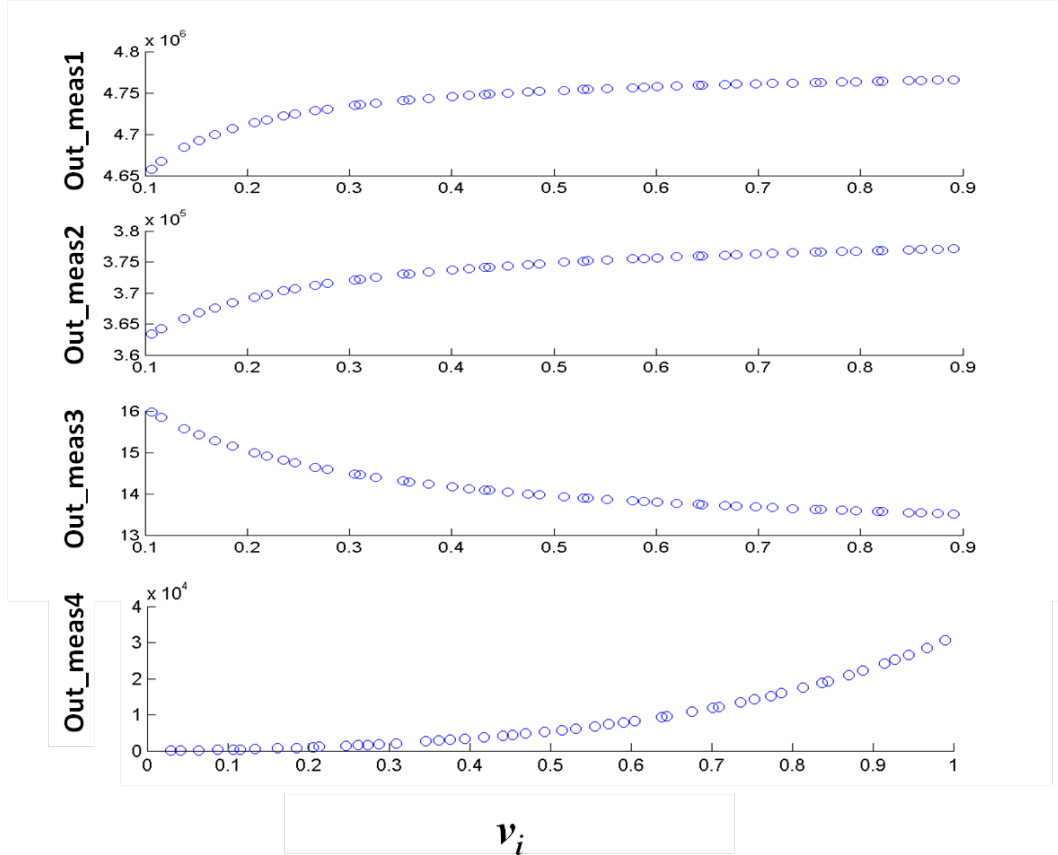
Figure 2.2: A Sample Monotonicity Plot of v_i for four outcome measures, Out_meas1, Out_meas2, Out_meas3 and Out_meas4.

To investigate the level of monotonicity for an LHS test parameter, v_i, for example, we use baseline values for all parameters except this particular v_i. (Recall that the baseline has been set to a value at or near the middle of the range between the minimum and the maximum values for v_i.) We then pick column i, $1 \leq i \leq K$ and run simulations using parameter values for v_i from that column. This will result in N simulations for this monotonicity test.

Suppose we have a test model with four outcome measures we will call, Out_meas1, Out_meas2, Out_meas3 and Out_meas4. Hypothetical monotonicity plots are shown for the LHS test parameter, v_i and these outcome measures in Figure 2.2. In the example shown in Figure 2.2 above, we see that monotonicity exists between the uncertain parameter, v_i, and the four outcome measures analyzed. However,
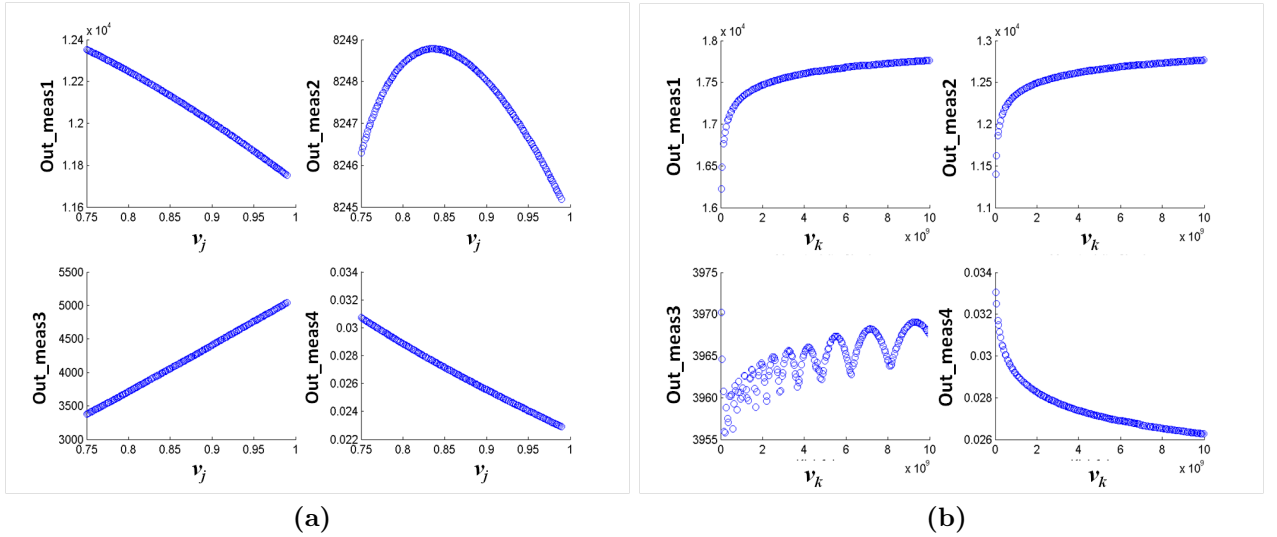
Figure 2.3: Sample graphs illustrating monotonicity and non-monotonicity properties for LHS test variables, v_j and v_k

it is not always the case that we see such monotone relationships as the ones portrayed. For example, observe in Figure 2.3, hypothetical monotonicity plots for two independent LHS test variables, say v_j and v_k, where some monotonic and non-monotonic relationships are seen. In the next two sections, we explore how to address the sensitivity analysis in cases where monotonicity fails, as in Figure 2.3. In the two sections that follow thereafter, we discuss other inferences one can make from monotonicity plots.

2.2.1 Excluding outcome measures for LHS parameters

In the monotonicity plots for our hypothetical v_j (Figure 2.3a), we observe that as v_j changes, there is a non-monotonic relationship with outcome measure, Out_meas2. However, the effect is minimal since the observed range of values for Out_meas2 is small (i.e., 8,246 to 8,249) for numbers of order 10^3. Therefore, we would consider v_j in runs for the other three outcomes, but not for Out_meas2. Doing so will ensure that that the lack of monotonicity for Out_meas2 does not give us an erroneous reading for the other outcome measures. In Figure 2.3b, there is a lack of monotonicity with the

increasing size of v_k for outcome measure, Out_meas3. Here again, due to the small range of observed values, we can see that it is not reasonable to conclude that the observed trend has much to do with Out_meas3. So for these results, we would re-run the LHS sampling procedure, this time excluding v_k for Out_meas3, and leaving v_k in the analysis for the other three measures.

2.2.2   Truncating the LHS parameter range when non-monotonicity exists

Examine the monotonicity plot for Out_meas2 in Figure 2.3a once more and notice that the graph could be broken up into two monotonic regions. If instead of the small range of outcome measures observed for Out_meas2, the range had been several hundred or thousand units, we would have considered truncating the range and looking at each truncated half separately. In general, if non-monotonic regions exist over large intervals for any of the uncertain parameters, and if the interval can be split into two or more intervals over large ranges for both the LHS parameter and the outcome measure, we may choose to adjust the chosen intervals (by selecting only monotonic regions) for that parameter and re-generate Table 2.1 before proceeding with the next steps of running statistics on our simulation results or proceeding with PRCC analysis. Note that the range chosen for analysis is determined by feasible parameter values. Moreover, exploring the PRCC analysis after excluding non-monotonic parameters and subsequently within monotonic sub-ranges for the parameters that remain can provide a clearer and more accurate picture of the remaining parameters' sensitivity.

2.2.3   Predicting strength of PRCC values

Monotonicity plots are useful for making conjectures about which parameters we would expect to have strong PRCC values. This is a way of verifying that our results are reasonable. In Figure 2.3a, for example, where monotonicity is observed for

*Out_meas*1, *Out_meas*3 and *Out_meas*4, we notice that the range of values spanned by the outcome measures is not negligible. Since the range is larger for *Out_meas*1 and *Out_meas*3, we can infer that that parameter v_j has a sizeable effect on these two outcome measures. On the other hand, since the range is very small for *Out_meas*4, we will most likely observe a much smaller effect for that outcome measure.

2.2.4 Truncating the LHS parameter range due to variation in the output values on certain subintervals

For *Out_meas*1, *Out_meas*2, and *Out_meas*4, the hypothetical parameter, v_k (Figure 2.3b) might actually take on any value on the interval, $[1 \times 10^9, 10 \times 10^9]$. However, the PRCC result for LHS in this range might suggest that the parameter is not important, while indeed the parameter is extremely sensitive within the smaller range, $[1 \times 10^9, 4 \times 10^9]$. If we run the LHS analysis for this smaller range, the parameter may show up as an important predictor of change in the outcome measure. On the other hand, we will only see negligible change to the outcome measure in the range, $[4 \times 10^9, 10 \times 10^9]$.

2.3 Handling the PRCC Steps

Once we have adjusted our LHS parameter ranges and generated a final version of the table in Figure 2.1, we may then proceed with running our simulations to use with PRCC analysis. For each of our N simulations, all the K values in each row in the table in Figure 2.1 are used as input values for the numerical simulation of the model. (See Figure 2.4.)

Upon completion of the N simulations, frequency histogram and descriptive statistics (minimum, maximum, mean, variance, 95% confidence interval, etc.) could be calculated for the outcome measures (e.g., total infected on the last day of an infection, cumulative number infected, etc.). The minimum and maximum of these

v_1	v_2	v_3	v_4	...
v_{11}	v_{12}	v_{13}	v_{14}	...
v_{21}	v_{22}	v_{23}	v_{24}	...
v_{31}	v_{32}	v_{33}	v_{34}	...
v_{41}	v_{42}	v_{43}	v_{44}	...
.	.	.	.	...
.	.	.	.	...
.	.	.	.	...
v_{N1}	v_{N2}	v_{N3}	v_{N4}	...

The N x K LHS matrix for K uncertain parameters and N simulations

The N x 1 matrix produced after the N simulations. It contains the values for the outcome measure of interest.

Outcome measure

$f(v_{11}, v_{12}, v_{13}, v_{14}, ...)$
$f(v_{21}, v_{22}, v_{23}, v_{24}, ...)$
$f(v_{31}, v_{32}, v_{33}, v_{34}, ...)$
$f(v_{41}, v_{42}, v_{43}, v_{44}, ...)$
.
.
.
$f(v_{N1}, v_{N2}, v_{N3}, v_{N4}, ...)$
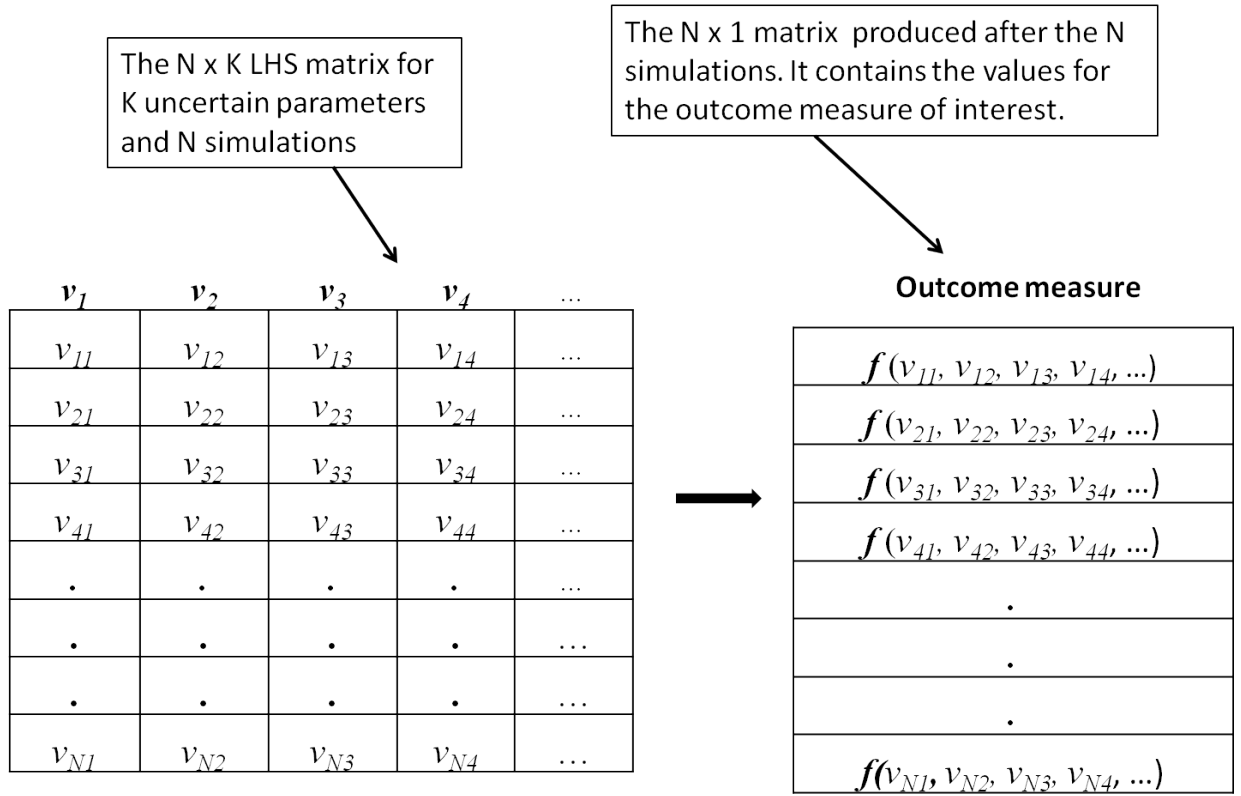
Figure 2.4: How Parameter Values Are Selected Per Run.

outcome measures, for example, would reflect the likely ranges of possible outcomes. The frequency distributions of the outcome measures can also be used to assess the probability of specific outcomes. Nevertheless, these statistics do not provide precise information on the model's sensitivity to its parameters. Thus, we see that to answer the question of which parameters contribute the most uncertainty to our model prediction, we cannot use these statistics or frequency distributions. Instead, we will use the non-parametric partial rank correlation coefficients to determine the most sensitive parameters.

2.3.1 PRCC Methodology

Rank transformation procedures are procedures in which a parametric procedure is applied to the ranks of the data instead of the data themselves [6]. The ranking

is simply done by arranging the values of our data in rank order and then assigning the smallest a value of 1, the next smallest a value of 2 etc. The transformation usually results in uniform residuals for the transformed variables. It is useful in non-parametric analysis such as Partial Rank Correlation when working with non-uniform data.

Hence, to perform PRCC, we will first rank the LHS matrix and the matrix for the outcome measure using a sort routine. For each parameter and each outcome measure, two linear regression models are found, the first representing that ranked parameter in terms of the other ranked parameter values and the second representing the ranked outcome measures in terms of the other ranked parameter values. A Pearson correlation coefficient for the residuals from those two regression models gives the PRCC value for that specific parameter. (See Equations (1) and (2) in [18].) These steps are illustrated in Figure 2.5.

2.3.2 PRCC results

The Partial Rank Correlation analysis equips us with Partial Rank Correlation Coefficients (PRCC) and corresponding p-values with which to assess the level of uncertainty an LHS parameter contributes to the model.

The magnitude as well as the statistical significance of the PRCC value of a parameter indicates that parameter's contribution to the model's prediction imprecision. The parameters with large PRCC values (> 0.5 or < -0.5) as well as corresponding small p-values (< 0.05) are the most important [19]. The closer the PRCC value is to $+1$ or -1, the more strongly the LHS parameter influences the outcome measure. The sign indicates the qualitative relationship between the input variable and the output variable. A negative sign indicates that the LHS parameter is inversely proportional to the outcome measure.

Sample LHS Matrix				
v_{11}	v_{12}	v_{13}	v_{14}	...
v_{21}	v_{22}	v_{23}	v_{24}	...
v_{31}	v_{32}	v_{33}	v_{34}	...
v_{41}	v_{42}	v_{43}	v_{44}	...
.	.	.	.	...
.	.	.	.	...
.	.	.	.	...
v_{N1}	v_{N2}	v_{N3}	v_{N4}	...

$\mathbf{A} =$

Sample outcome measure
$b_1 = f(v_{11}, v_{12}, v_{13}, v_{14}, ...) = 20$
$b_2 = f(v_{21}, v_{22}, v_{23}, v_{24}, ...) = 10$
$b_3 = f(v_{31}, v_{32}, v_{33}v_{34}, ...) = 5$
$b_4 = f(v_{41}, v_{42}, v_{43}, v_{44}, ...) = 30$
.
.
.
$b_N = f(v_{N1}, v_{N2}, v_{N3}, v_{N4}, ...)$

$\mathbf{B} =$

II) After Ranking

$\mathbf{A_{ranked}} =$

Sample LHS Ranked Matrix			
3	2	4	5
5	4	2	1
2	1	3	4
1	5	6	3
.	.	.	.
.	.	.	.
.	.	.	.

$\mathbf{B_{ranked}} =$

Sample Ranked outcome measure
3
2
1
5
.
.
.

Multiple Linear Regression on each column of $\mathbf{A_{ranked}}$

Multiple Linear Regression on each column of $\mathbf{B_{ranked}}$

Obtain Residuals, $\mathbf{R_{A\text{-}Ranked}}$, from regression on $\mathbf{A_{ranked}}$

Obtain Residuals, $\mathbf{R_{B\text{-}Ranked}}$, from regression on $\mathbf{B_{ranked}}$

Plot($\mathbf{R_{A\text{-}Ranked}}$, $\mathbf{R_{B\text{-}Ranked}}$)

PRCC($\mathbf{R_{A\text{-}Ranked}}$, $\mathbf{R_{B\text{-}Ranked}}$)

Figure 2.5: An Illustration of the Partial Rank Correlation steps.

A sample PRCC output is shown in Figure 2.6 for six test parameters, v_1 through v_6. Also, three outcome measures (Out_meas1, Out_meas2, and Out_meas3) have been considered. The corresponding PRCC plot for Out_meas1 is shown in Figure 2.6. Note that in these plots, the x-axis corresponds to the regression coefficients for outcome measure while the y-axis represents regression coefficients for the LHS parameter we are studying.

For each plot in Figure 2.6, the y-axis represents the residuals for ranked LHS parameter values while the x-axis represents the residuals of the ranked outcome measure. On the top of each plot are two values $[x, y]$, with x representing the PRCC value and y representing the corresponding p-value. The data from this figure can be imported into a table for easier analysis. See Table 2.1.

In this table, important contributors to uncertainty have both their PRCC values (orange) and their p-values (grey) highlighted, not just one or the other. The symbol, (*) is used to indicate possible contributors (PRCC values: ∼0.5 to 0.69 or -0.5 to -0.69). We use (**) to indicate very likely contributors to uncertainty (PRCC values: ∼0.7 to 0.79 or -0.7 to -0.79). Finally, (***) is used to indicate highly likely contributors to uncertainty (PRCC values: ∼0.8 to 0.99 or -0.8 to -0.99).

From Table 2.1, we would conclude that in our hypothetical example, parameters v_3, v_5, and v_6 have an important influence on the model as a whole. Also notice that the small p-value associated with them indicate that they appear important for at least two outcome measures; so, we would conclude that they are important contributors to our model's prediction imprecision.
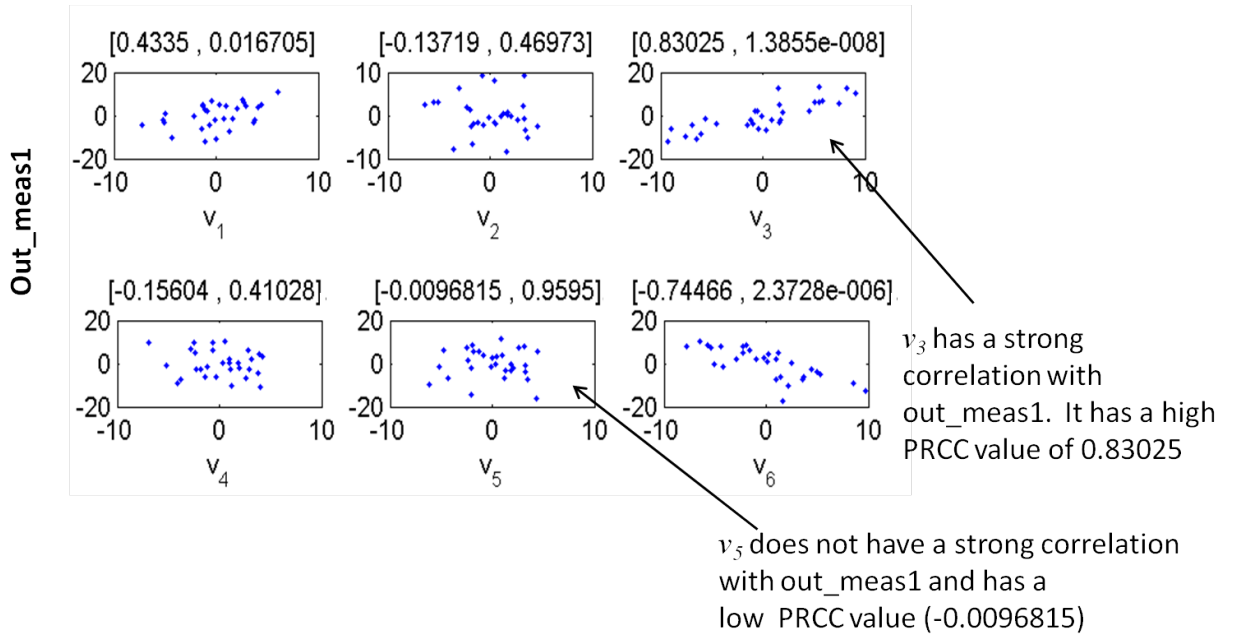
Figure 2.6: PRCC Plots for Out_meas1**.**

Table 2.1: Output from PRCC Analysis.

	Out_meas1		Out_meas2		Out_meas3	
	PRCC	*p-value*	*PRCC*	*p-value*	*PRCC*	*p-value*
v_1	0.4335	0.016705	0.063085	0.74051	-0.09318	0.62432
v_2	-0.13719	0.46973	-0.06742	0.72334	0.045885	0.80973
v_3	***0.83025	1.39E-08	*0.60836	0.000362	**0.78816	2.33E-07
v_4	-0.15604	0.41028	0.16662	0.37886	0.2302	0.22102
v_5	-0.00968	0.9595	***-0.88865	5.52E-11	***-0.88101	1.33E-10
v_6	***-0.74466	2.37E-06	*-0.53965	0.002086	-0.00793	0.96682

Chapter 3

Applying the LHS Procedure to a Cholera Epidemic Model

3.1   Background

Cholera is an acute water-borne diarrheal disease caused by infection of the human intestines by the bacterium Vibrio cholerea. The disease can be transmitted either directly by human-to-human contact (fecal-oral transmission) or indirectly via environment-to-human contact (food and water-borne transmission). The cause of death is mainly dehydration, and in severe cases, without treatment, death may occur within hours of infection. Preventive measures include improved sanitation and water supply and more recently, oral vaccines.

The spread of cholera is currently viewed as a global threat to public health and a key indicator of lack of social development. As of 2010, there were 317,534 cases and 7,543 deaths reported worldwide, with as much as 115,106 cases originating in Africa, $179,594$ in the Americas, and 13,819 cases in Asia [13]. Nevertheless, studies show that this statistic only reflects 5-10% of the actual number of deaths [14]. The ensuing enormous loss of life and economic burden caused by the disease underscore an urgent need for a better understanding of the disease dynamics.

3.2 Details of the Cholera Model

The model studied in this thesis is an extension of the a model developed by Neilan et al. [16] and found in the dissertation work of Peng Zhong [15]. It is a system of nine ordinary differential equations shown below. Figure 3.1 has been provided as a diagrammatic representation of these equations.

$$
\begin{aligned}
\frac{dS}{dt} &= -\Big[\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}\Big]S(t) \\
&\quad + b\big(S(t) + \hat{S}(t) + I_S(t) + I_A(t) + R_S(t) + R_A(t) + V(t)\big) \\
&\quad - dS(t) + \omega_3 \hat{S}(t) + \omega_4 V(t) - uS(t) &(3.1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{dI_S}{dt} &= p\Big[\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}\Big]S(t) \\
&\quad - dI_S(t) - \gamma_2 I_S(t) - e_2 I_S(t) &(3.2)
\end{aligned}
$$

$$
\frac{dR_S}{dt} = -dR_S(t) + \gamma_2 I_S(t) - \omega_2 R_S(t) \qquad (3.3)
$$

$$
\begin{aligned}
\frac{d\hat{S}}{dt} &= -\Big[\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}\Big]\hat{S}(t) - d\hat{S}(t) &(3.4) \\
&\quad - \omega_3 \hat{S}(t) + \omega_1 R_A(t) + w_2 R_S(t) - u\hat{S}(t)
\end{aligned}
$$

$$
\begin{aligned}
\frac{dI_A}{dt} &= \Big[\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}\Big]\hat{S}(t) - dI_A(t) - e_1 I_A(t) \\
&\quad - \gamma_1 I_A(t) + (1-p)\Big[\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}\Big]S(t) &(3.5)
\end{aligned}
$$

$$
\frac{dR_A}{dt} = -dR_A(t) + \gamma_1 I_A(t) - \omega_1 R_A(t) \qquad (3.6)
$$

$$
\frac{dV}{dt} = u(\hat{S}(t) + S(t)) - \omega_4 V(t) - dV(t) \qquad (3.7)
$$

$$
\frac{dB_H}{dt} = \eta_1 I_A(t) + \eta_2 I_S(t) - \chi B_H(t) \qquad (3.8)
$$

$$
\frac{dB_L}{dt} = \chi B_H(t) - \delta B_L(t) \qquad (3.9)
$$

Each differential equation describes the disease dynamics for either a human or bacterial subpopulation (or class). Of the nine classes, seven describe the disease dynamics for humans and the other two describe the disease dynamics for the bacterial
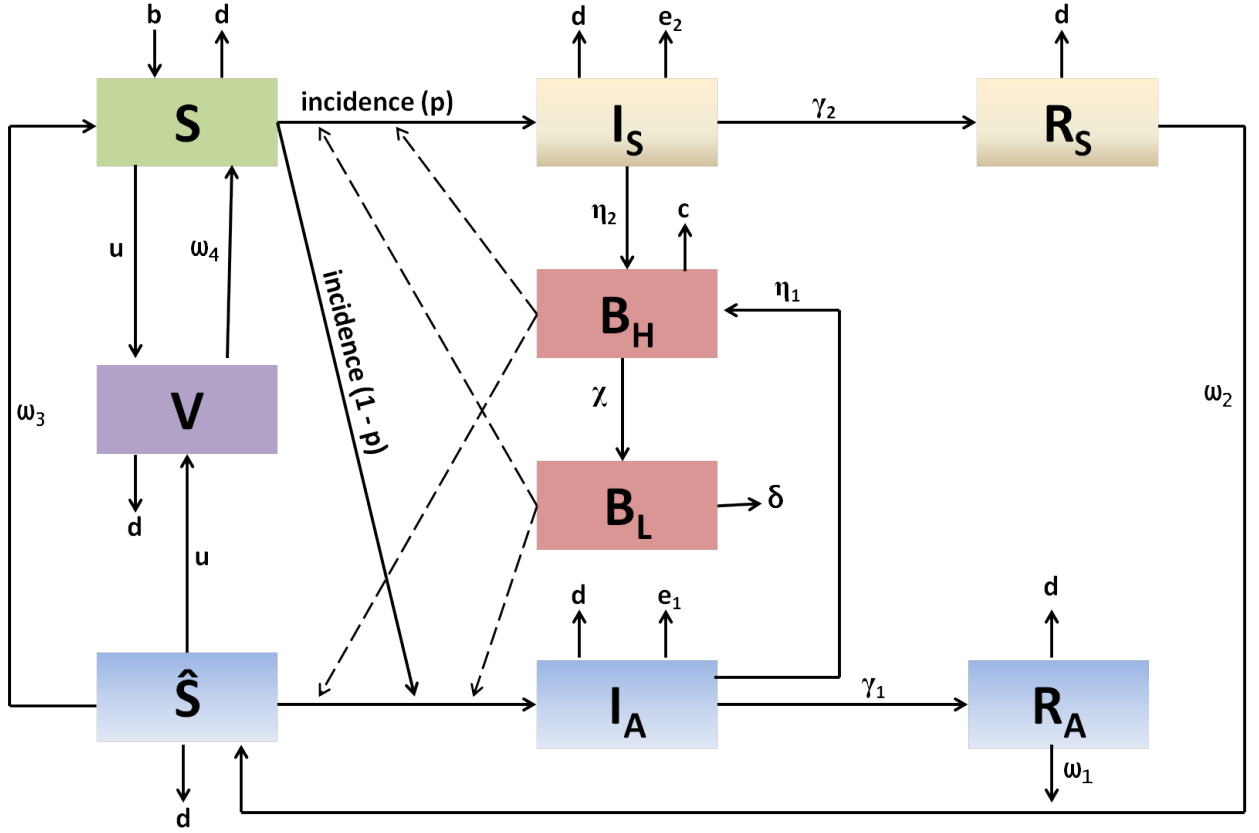
Figure 3.1: Cholera Epidemic Model.

population. There are two classes of susceptible humans, S and $\hat{S}$, respectively. When members of the S class fall sick, they can either proceed to the symptomatic infected class, I_S, or the infected asymptomatic class, I_A. On the other hand, members of the $\hat{S}$ class have a natural immunity that causes them to proceed to the infected asymptomatic class and subsequently the recovered asymptomatic class without ever showing symptoms of the disease. Members of both the $\hat{S}$ and the S class can be vaccinated and, thus, they move to the vaccinated compartment (V). The two bacterial compartments consist of one class characterized by low levels of infectivity (B_L) and another class with hyper-infectious transmission (B_H).

The model proposed is unique because it incorporates the following features:

1. A separate class for the mildly infectious or, asymptomatic humans ($\hat{S}$, I_A, and R_A), a feature first suggested by King et al. [1].

2. A hyperinfectious, short-lived bacterial state (B_H) as suggested by Merrell et al. [3] and Hartley et al. [12]. In the model, when symptomatic and asymptomatic susceptible humans drink contaminated water, they are infected at rate of β_H, for hyperinfectious bacteria, and at a rate of rate β_L, for bacteria with a low level of infectivity.

3. A vaccinated class (V).

4. Waning disease immunity rates for the Recovered classes (ω_1 and ω_2), for the asymptomatic susceptible class (ω_3) and for the vaccinated class (ω_4).

As mentioned earlier, once infected, a proportion (p) of the susceptible individuals will proceed to the symptomatic infected class while another proportion $(1-p)$ will move to the asymptomatic infected class. The latter will have partial immunity and, thus, will experience mild or inapparent symptoms. The asymptomatic infected class will also have fewer cholera-related deaths, e_1, smaller bacteria shedding rate, η_1, and a larger recovery rate, γ_1, than the symptomatic infected class. For our analysis, we have assumed that the vaccination rate, u, is zero so that this is the study of the model without any control. Table 3.1 outlines the parameters used in the model and their values in system (3.1) - (3.9).

3.3   Performing the Latin Hypercube Sampling

For the cholera model, 14 uncertain or Latin Hyerpcube Sampling (LHS) parameters are identified. These are the parameters deemed most significant in the disease dynamics. To determine their various roles in the model predictions, we begin by performing LHS analysis. In our analysis, 30 model simulations are performed and the model is run for 180 days per run. Thus, the parameter space (or LHS matrix) for the LHS parameters has dimension of length 14 with each dimension specifying an uncertain parameter vector of length 30.

Table 3.1: Parameter List for Cholera Epidemic Model.

Symbol	Description	Value
$\hat{S}_0$	Initial # susceptible humans with partial immunity	3000
S_0	Initial # susceptible humans without partial immunity	$10{,}000 - \hat{S}_0$
I_{A0}	Initial # asymptomatic infecteds	0
I_{S0}	Initial # symptomatic infecteds	0
R_{A0}	Initial # recovered humans (asymptomatic)	0
R_{S0}	Initial # recovered humans (symptomatic)	0
V_0	Initial # humans with vaccinated immunity	0
B_{H0}	Initial concentration of highly infectious (HI) vibrios in environment	0
B_{L0}	Initial concentration of non-highly infectious (non-HI) vibrios in environment	$\kappa_L/2$
p	Probability of infecteds moving from symptomatic class to infected class without partial immunity	0.6
r	Scaling factor used to compute β_H from β_L.	0.1
β_L	Ingestion rate of non-HI vibrio from environment	0.008 day^{-1}
β_H	Ingestion rate of HI vibrio from environment.	$r*\beta_L$ day^{-1}
κ_L	Half saturation constant of non-HI vibrios	103 cells/ml
κ_H	Half saturation constant of HI vibrios	$\kappa_L/700$ cells/ml
e_1	Cholera-related death rate for asymptomatic infecteds	$e_2/20$ day^{-1}
e_2	Cholera-related death rate for symptomatic infected	0.03 day^{-1}
γ_1	Cholera recovery rate (asymptomatic)	0.75 day^{-1}
γ_2	Cholera recovery rate (symptomatic)	0.1 day^{-1}
ω_1	Rate of waning cholera immunity from asymptomatic infecteds to susceptibles with partial immunity	1/180 day^{-1}
ω_2	Rate of waning cholera immunity from symptomatic infecteds to susceptible humans with partial immunity	$1/(365 * 2)$ day^{-1}
ω_3	Immunity waning rate: susceptibles without partial immunity $\rightarrow$ susceptibles with partial immunity	$1/(10 * 365)$ day^{-1}
ω_4	Immunity waning rate: humans with vaccinated immunity $\rightarrow$ susceptibles without partial immunity	0.001 day^{-1}
s	Scaling factor used to compute η_2 from η_1	100
η_1	Rate of contribution to HI vibrios in environment by asymptomatic infecteds	0.008 cells/ml-day-human
η_2	Rate of contribution to HI vibrios in environment by symptomatic Infected.	$s*\eta_1$ cells/ml-day-human
χ	Transaction rate of vibrios from HI to non-HI state	5 day^{-1}
d	Death rate of vibrios	1/30 day^{-1}
u	Rate at which susceptible and asymptomatic infecteds are vaccinated daily	0 day^{-1}
b	Natural birth rate of humans	0.03/365 day^{-1}
d	Natural death rate of humans	0.02/365 day^{-1}

Table 3.2: Baseline, Maximum and Minimum Values Used in the LHS Analysis.

Parameter	Min	Baseline	Max
ω_1	0.001	1/180	0.03
ω_2	0.001	1/(2*365)	0.003
ω_3	0.0003	1/(10*365)	0.001
p	0.3	0.6	0.9
r	0.01	0.1	1
e_2	0.01	0.03	0.05
$e_1 = e_2/20$	0.0005	0.0015	0.0025
γ_1	1/2	0.75	1
γ_2	1/14	0.1	1/7
η_1	0.0001	0.008	0.05
η_2	1	100	200
s	0.001	0.008	0.08
B_{L0}	$\kappa_L/500$	$\kappa_L/2$	κ_L
S_0	6000	10000 - $\hat{S}_0$	10000

Maximum and minimum values are determined for each of the 14 LHS parameters (Table 3.2). Note that the baseline value for each LHS parameter has been set to a value at or near the middle of the range between the minimum and maximum values for that parameter.

For each LHS parameter, each of the 30 input values are obtained by the sampling a uniform probability density distribution. The 30 input values are then used to populate the LHS matrix from which we output monotonicity plots for each variable. We calculate two outcome measures for each run: Total Infecteds (i.e., sum of Symptomatic and Asymptomatic infecteds at each time step) and Total Symptomatic Infecteds, respectively. For additional intuition, on the top of each monotonicity plot, we display percent change in population values for each outcome measure. In Total Infecteds, for example, the percent change (PC) is calculated as: (Maximum #Total Infecteds - Minimum #Total Infecteds) / Maximum #Total Infecteds. If percent change equals zero, then we know that the outcome measure is not affected by the chosen LHS parameter. This indicates that in our analysis, we can omit that outcome measure for that parameter.

3.3.1   Analyzing the Monotonicty Plots

Two outcome measures are evaluated in these plots: Total number of infecteds (from symptomatic and asymptomatic compartments in the model) and total number of symptomatic infecteds (from symptomatic compartment alone).

The plots in Figures 3.2 and 3.3 indicate that all the LHS parameters have a monotonic relationship with the outcome measures. Consequently, we do not adjust the ranges chosen for the analysis and we proceed with the LHS matrix generated after the LHS sampling (see Section 2.2). After this step, frequency histogram and descriptive statistics (minimum, maximum, mean, variance, 95% confidence interval, etc.) could be calculated for the outcome measures, but we omit that step here and proceed with the Partial Rank Correlation Coefficient (PRCC) analysis.

3.3.2   Analyzing the PRCC Results

Next, we perform a multilinear regression analysis on the ranks obtained for the outcome measures (i.e., Total Symptomatic Infecteds and Total infecteds) and for the LHS parameters. We then perform a regression analysis on these ranks to obtain the regression coefficents. Since these regression coefficents provide a measure of the sensitivity of the model to the LHS parameters, we proceed to determine the strength of the relationship between each LHS parameter and each outcome measure by obtaining the PRCC values.

In PRCC analysis in general, the parameters with large PRCC values (> 0.5 or < -0.5) and corresponding small p-values (< 0.05) are deemed the most influential in the model. In both Figures 3.4 and 3.5, we have plotted residuals for ranked LHS parameter values on the y-axis. On the x-axis of Figure 3.4, we have also plotted the residuals for the ranked Total Infecteds, while on the x-axis of Figure 3.5, we have plotted the residuals for the Total Symptomatic Infecteds.
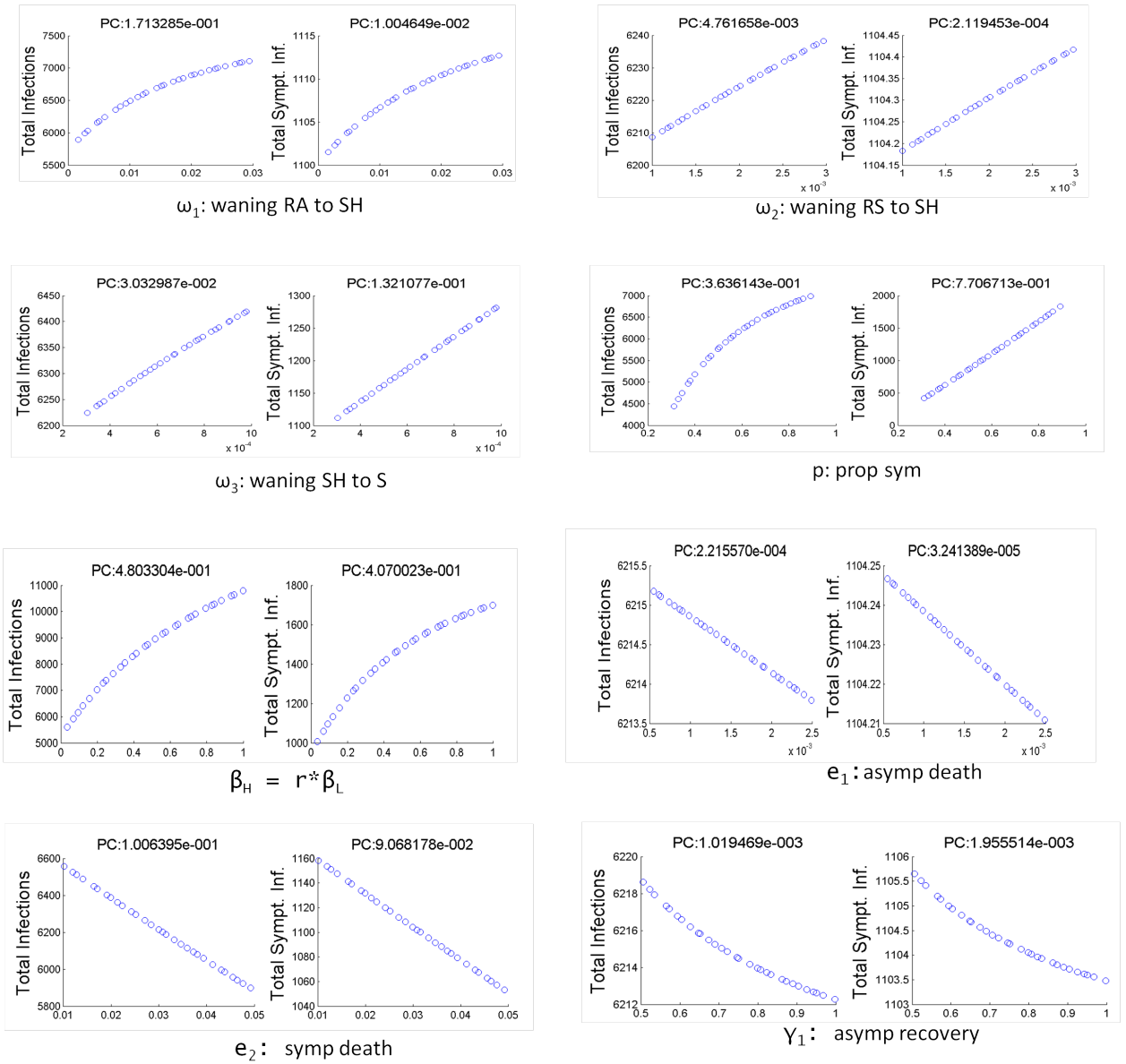
Figure 3.2: Monotonicity plots for ω_1, ω_2, ω_3, p, β_H, e_1, e_2, and γ_1

24
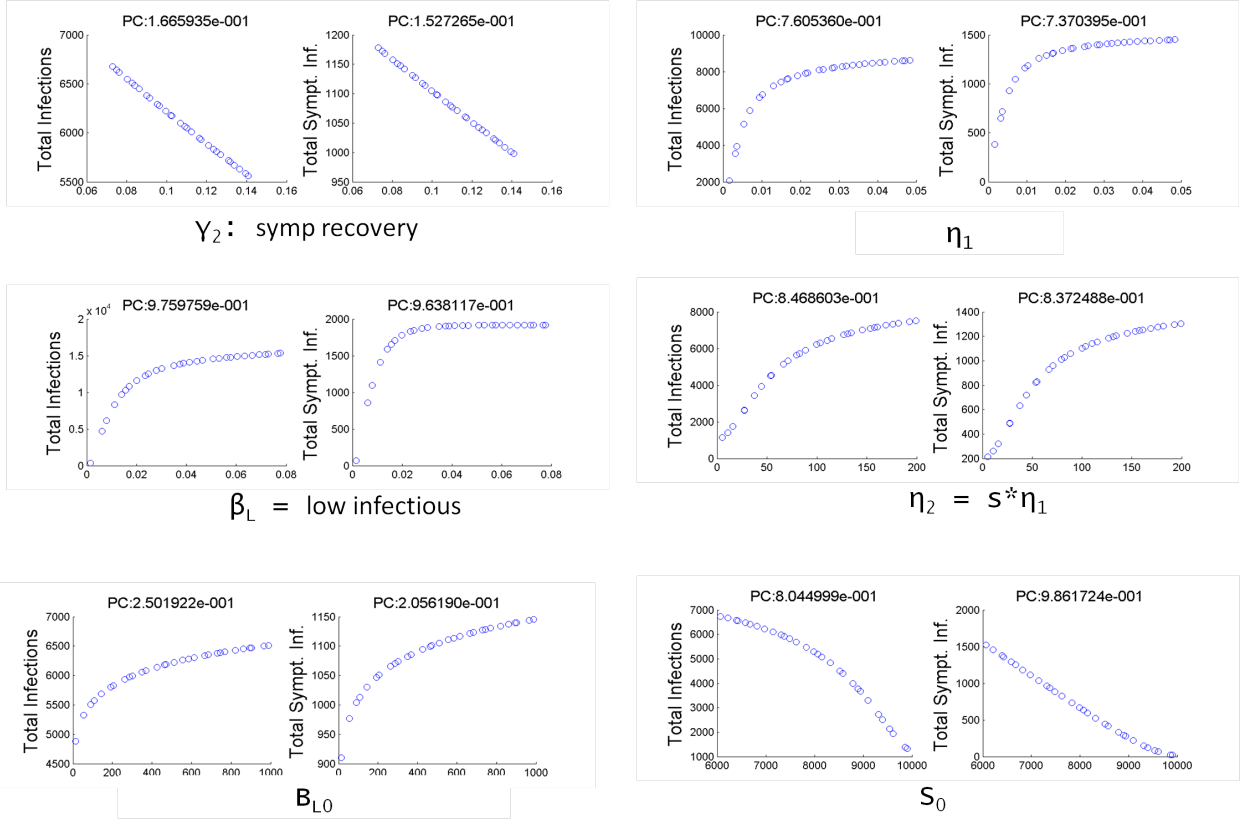
Figure 3.3: Monotonicity plots for γ_2, η_1, η_1, β_L, η_2, B_{L0}, and S_0

Note that on top of each PRCC plot are two values, $[x, y]$, with x representing the Pearson Partial Rank Correlation Coefficient (PRCC) value, and y representing the corresponding p-value. Also observe that the PRCC plots for the two outcome measures show that a strong correlation is observed for several LHS parameters.

The results from the PRCC plots are summarized in Table 3.3. In this table, important contributors to uncertainty have both their PRCC values (orange) and their p-values (grey) highlighted, not just one or the other. As before, (*) is used to indicate possible contributors (PRCC values: $\sim$0.5 to 0.69 or -0.5 to -0.69), (**) is used to indicate very likely contributors to uncertainty (PRCC values: $\sim$0.7 to 0.79 or -0.7 to -0.79) and (***) is used to indicate highly likely contributors to uncertainty (PRCC values: $\sim$0.8 to 0.99 or -0.8 to -0.99).

From Table 3.3, we observe that the most influential LHS parameters for the outcome measure, Total Infecteds (i.e., sum of symptomatic and asymptomatic infecteds) are the following:

1. The waning immunity rate, ω_1, when going from the recovered asymptomatic population, R_A, to the susceptible asymptomatic population S_H

2. The ingestion rate, β_H, from the environment of highly infectious vibrios

3. The ingestion rate, β_L, from the environment of non-highly infectious vibrios

For the outcome measure, Total Symptomatic Infecteds, on the other hand, we observe a strong correlation with the initial susceptible population, S_0.

The four parameters, ω_1, β_H, β_L and S_0 will be important contributors to uncertainty in the model. From Table 3.3, we notice that β_L and S_0 have higher PRCC values than the other two parameters and conclude that β_L and S_0 are the most influential parameters in the model.
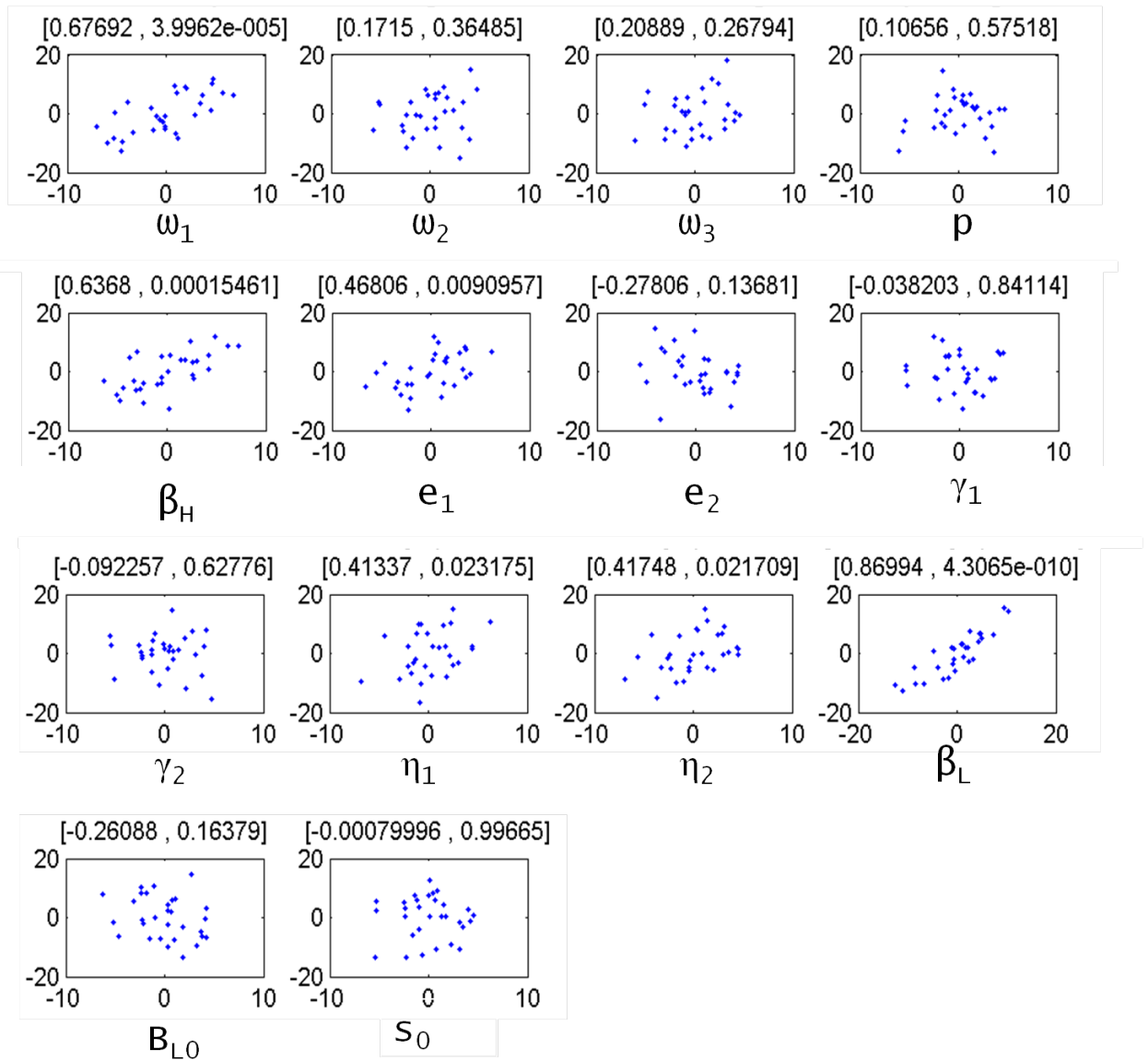
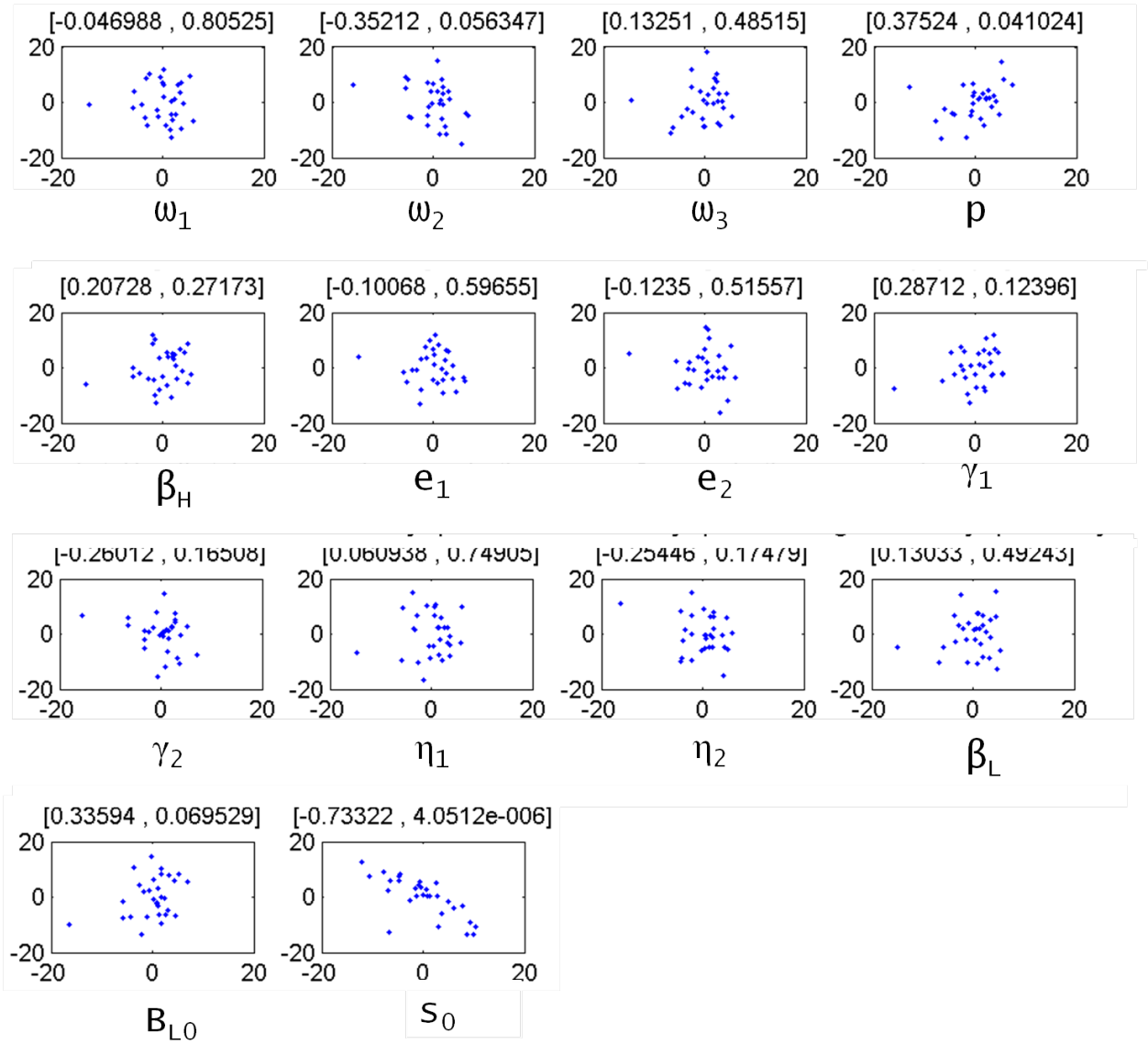Figure 3.4: PRCC Plots for Total Infecteds.

Figure 3.5: PRCC Plots for Total Symptomatic Infecteds.

Table 3.3: Output from PRCC Analysis

Parameters	Total Infectious		Total Symptomatic Infecteds	
	PRCC	*p-value*	*PRCC*	*p-value*
ω_1: Waning R_A to $\hat{S}$	*0.67692	4.00E-05	-0.04699	0.80525
ω_2: Waning R_S to $\hat{S}$	0.1715	0.36485	-0.35212	0.056347
ω_3: Waning $\hat{S}$ to S	0.20889	0.26794	0.13251	0.48515
p: Prop. Sympt.	0.10656	0.57518	0.37524	0.041024
$\beta_H = r * \beta_L$	*0.6368	0.000155	0.20728	0.27173
e_1:Asympt. death rate	0.46806	0.009096	-0.10068	0.59655
e_2:Sympt. death rate	-0.27806	0.13681	-0.1235	0.51557
γ_1	-0.0382	0.84114	0.28712	0.12396
γ_2	-0.09226	0.62776	-0.26012	0.16508
η_1	0.41337	0.023175	0.060938	0.74905
$\eta_2 = s * \eta_1$	0.41748	0.021709	-0.25446	0.17479
β_L: Low infectious	***0.86994	4.31E-10	0.13033	0.49243
B_{L0}	-0.26088	0.16379	0.33594	0.069529
S_0	-0.0008	0.99665	**-0.73322	4.05E-06

Chapter 4

LHS/PRCC Applied to Cholera Model with Control

4.1 Optimal Control formulation

The model we have studied so far has been implemented with no control (i.e., $u = 0$). In this chapter, we now introduce vaccination as a mitigation scheme with the goal of reducing cholera-related deaths while maintaining minimal vaccination cost. We apply optimal control analysis to our model and solve the system for the optimal control scheme proposed by the model.

A control scheme for the cholera epidemic model is considered optimal if it minimizes the objective functional:

$$J(u) = \int_0^T AI_S(t) + Bu(t)(S(t) + \hat{S}(t) + I_A(t) + R_A(t)) + C(S_0 + \hat{S}_0)u^2(t) \, \mathrm{d}t \quad (4.1)$$

where A, B and C are balancing coefficients which transform the integrand into units of dollars [15]. The goal of the objective functional is the minimize the number of infecteds (the first term in Equation (4.1)) as well as the cost of vaccination. In the equation, the cost is defined by both a linear term and a quadratic term. The linear

term measures the total population vaccinated while the quadratic term measures the non-linear costs that may result from high intervention levels.

Stated explicitly, the optimal control problem addressed by the model is the following:

Find $u^* \in U$ such that

$$J(u^*) = \min_{u \in U} \ J(u) \tag{4.2}$$

subject to the state system (3.1) - (3.9) and the initial conditions given in Table 3.1. Note that the control is the set $U = \{u \in L^\infty([0,T]) | 0 \leq u(t) \leq u_{max}, t \in [0,T]\}$ for $u_{max} < 1$.

We proceed to characterize our optimal control using Pontryagin's Maximum Principle [9]. Using this principle, we introduce nine adjoint functions that attach our system of nine differential equations (3.1) - (3.9) to our objective functional. According to Fleming and Rishel [20], we must begin our analysis by first showing that the optimal control exists. From [15], we know that the an optimal control exists and can apply Pontryagin's Maximum Principle. The optimal control characterization that results from applying Pontryagin's Maximum Principle has the optimal control expressed in terms of the state and the adjoint functions. The principle also converts the minimization problem (4.2) into a problem of minimizing the Hamiltonian with respect to the control at time t. The Hamiltonian is the following:

$$H =$$

$$AI_S(t) + Bu(t)(S(t) + \hat{S}(t) + I_A(t) + R_A(t)) + C(S + S_0)u^2(t)$$

$$+ \lambda_S \Bigg(- ([\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}] - b + d + u)S(t)$$

$$+ b(\hat{S}(t) + I_S(t) + I_A(t) + R_S(t) + R_A(t) + V(t)) + \omega_3 \hat{S}(t) + \omega_4 V(t) \Bigg)$$

$$+ \lambda_{\hat{S}} \Bigg(- ([\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}] + d + \omega_3 + u)\hat{S}(t)$$

$$+ \omega_1 R_A(t) + \omega_2 R_S(t) \Bigg)$$

$$+ \lambda_{I_S} \Bigg(p[\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}]S(t) - (d + \gamma_2 + e_2)I_S(t) \Bigg) \qquad (4.3)$$

$$+ \lambda_{I_A} \Bigg([\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}]\hat{S}(t) - (d + \gamma_1 + e_1)I_A(t)$$

$$+ (1 - p)[\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}]S(t) \Bigg)$$

$$+ \lambda_{R_S}(-(d + \omega_2)R_S(t) + \gamma_2 I_S(t))$$

$$+ \lambda_{R_A}(-(d + \omega_1)R_A(t) + \gamma_1 I_A(t))$$

$$+ \lambda_V(u(\hat{S}(t) + S(t)) - (\omega_4 + d)V(t))$$

$$+ \lambda_{B_H}(\eta_1 I_A(t) + \eta_2 I_S(t) - \chi B_H(t))$$

$$+ \lambda_{B_L}(\chi B_H(t) - \delta B_L(t))$$

where λ_S, $\lambda_{\hat{S}}$, etc. are the adjoint variables obtained for the states, S, $\hat{S}$, etc., respectively. These adjoint variables satisfy the following equations:

$$\frac{d\lambda_S}{dt} = -Bu + [\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}](\lambda_S - p\lambda_{I_S} \qquad (4.4)$$
$$-(1-p)\lambda_{I_A}) + (d - b + u)\lambda_S - \lambda_V u$$

$$\frac{d\lambda_{\hat{S}}}{dt} = -Bu + [\beta_L \frac{B_L(t)}{\kappa_L + B_L(t)} + \beta_H \frac{B_H(t)}{\kappa_H + B_H(t)}](\lambda_{\hat{S}} - \lambda_{I_A}) \qquad (4.5)$$
$$+(d + \omega_3 + u)\lambda_{\hat{S}} - (\omega_3 + b)\lambda_S - \lambda_V u$$

$$\frac{d\lambda_{I_S}}{dt} = -A + (d + \gamma_2 + e_2)\lambda_{I_S} - b\lambda_S - \gamma_2\lambda_{R_S} - \eta_2\lambda_{B_H} \qquad (4.6)$$

$$\frac{d\lambda_{I_A}}{dt} = -Bu + (d + \gamma_1 + e_1)\lambda_{I_A} - b\lambda_S - \gamma_1\lambda_{R_A} - \eta_1\lambda_{B_H} \qquad (4.7)$$

$$\frac{d\lambda_{R_S}}{dt} = -\omega_2\lambda_{\hat{S}} - b\lambda_S + (d + \omega_2)\lambda_{R_S} \qquad (4.8)$$

$$\frac{d\lambda_{R_A}}{dt} = -Bu - \omega_1\lambda_{\hat{S}} - b\lambda_S + (d + \omega_1)\lambda_{R_A} \qquad (4.9)$$

$$\frac{d\lambda_V}{dt} = -\omega_4\lambda_S - b\lambda_S + (d + \omega_4)\lambda_V \qquad (4.10)$$

$$\frac{d\lambda_{B_H}}{dt} = \beta_H \frac{\kappa_H(t)}{(\kappa_H + B_H)^2}(S(\lambda_S - p\lambda_{I_S} - (1-p)\lambda_{I_A}) + \hat{S}(\lambda_{\hat{S}} - \lambda_{I_A})) \qquad (4.11)$$
$$+\chi\lambda_{B_H} - \chi\lambda_{B_L}$$

$$\frac{d\lambda_{B_L}}{dt} = \beta_L \frac{\kappa_L(t)}{(\kappa_L + B_L)^2}(S(\lambda_S - p\lambda_{I_S} - (1-p)\lambda_{I_A}) + \hat{S}(\lambda_S - \lambda_{I_A})) \qquad (4.12)$$
$$+\delta\lambda_{B_L}$$

with transversality conditions for each of the adjoint equations having a value of zero at t = T. That is, $\lambda_S(T) = \lambda_{\hat{S}}(T) = \lambda_{I_S}(T) = \lambda_{I_A}(T) = \lambda_{R_S}(T) = \lambda_{R_A}(T) = \lambda_V(T) = 0$.

The optimal control is characterized by:

$$u^* = max\left(0, min\left(\frac{-B(S + \hat{S} + I_A + R_A) + S\lambda_S + \hat{S}\lambda_{\hat{S}} - (S + \hat{S})\lambda_V}{2C(S_0 + \hat{S}_0)}, u_{max}\right)\right).$$

The Hamiltonian, the adjoint equations, and their corresponding transversality conditions, together with the optimality condition, make up the necessary conditions for optimality when using the Pontryagin's Maximum Principle. Using the control

Table 4.1: Parameters in the Objective Functional.

Parameters	Value
Maximum value of u (u_{max})	0.04
A	1
B	0.5
C	2

characterization, the state system of differential equations and the adjoint systems of differential equations can be solved numerically using the forward-backward sweep method [17].

In the next section, we will couple our optimal control analysis procedure to the LHS/PRCC procedure. For each LHS/PRCC simulation, we will solve our problem numerically for the objective functional value at the optimal control. That objective functional value will be the outcome measure for that simulation. Our goal is to study the sensitivity of the value of the objective functional at the optimal control and state to the LHS parameters chosen.

4.2   Latin Hypercube Sampling Analysis

Following our optimal control characterization, we examine the same LHS parameters and use the same LHS parameter ranges as the ones in Tables 3.1 and 3.2 in Chapter 3. Note that the control parameter u in Table 3.1 is set to a maximum value of $u = 0.04$. Our outcome measure will be the objective functional value at the optimal control and state and our goal is to investigate the level of influence these LHS parameters have on the objective functional.

In Table 4.1, we see a list of the coefficients and other conditions used in the forward-backward sweep method. Note that often, we might be interested in including the optimal control balancing coefficients (i.e., A, B and C in the objective functional, Equation 4.1) in the list of LHS parameters to investigate. However, we do not include them in our LHS/PRCC parameter sensitivity analysis.

To begin the LHS procedure, we will first verify that a monotonic relationship exists between the objective functional value and the LHS parameters. For each parameter, we run 20 simulations at a tolerance level of 0.1 for 180 days. In running the simulations, we observe a lack of convergence of the forward-backward sweep method when computing the objective functional values at the optimal control and state for parameters, p, η_1, η_2, β_L, and β_H. Notice that the lack of convergence in β_H is anticipated since $\beta_H = r * \beta_L$. With the exception of β_L, we also observe a lack of monotonicity for these parameters (see Figures 4.1 and 4.2). Non-monotonicity is observed for other parameters (ω_1 , ω_2 and e_1) as well. Nevertheless, the range of objective functional values appears to be small for these parameters, hence, the lack of monotonicity is ignored and more attention is given to the parameters with a lack of monotonicity and/or lack of convergence.

Observe that if we had not encounered a lack of convergence, then we would not make the same exception for p and η_2 that we applied to ω_1 , ω_2 and e_1. This is because the lack of monotonicity observed occurs for a large range of objective functional values for p and η_2. As seen in section 2.2, we would proceed by truncating the range for these parameters and re-run the LHS analysis, making sure to re-set the baseline value to the middle of the new range chosen. Moreover, because $\eta_2 = s * \eta_1$., we would adjust the range for η_2 by changing the range for either s or η_1, depending on which one has the most contribution to the lack of monotonicity observed.

However, our goal is to only make predictions based on parameters for which the forward-backward sweep method converges. We note that further work could be done to understand difficulties with convergence. Thus, we remove parameters p, η_1, η_2, β_L, and β_H from the analysis. We also remove parameter s since η_2 is dependent on it (i.e., $\eta_2 = s * \eta_1$). This leaves us with 9 parameters out of the 14 we originally started with. We re-run the simulations and find that the forward-backward sweep method converges for all nine parameters when the corresponding objective functional values are computed.

From the monotonicity plots (Figures 4.3), we observe as before that the plots for ω_1, ω_2 and e_1 are not monotone. Nevertheless, since the range of objective functional values appears to be small, the non-monotonicity is ignored. All the other parameters studied appear monotone, thus, we proceed with the next phase of the analysis.

4.3   Partial Rank Correlation Coefficient Analysis

To verify the strength of the correlation between the LHS parameters and the objective functional for the monotonic parameters, we perform the PRCC analysis and create the PRCC plots (see Figure 4.4). These plots show residuals for ranked LHS parameter values on the y-axis and residuals for the objective functional on the the x-axis. As seen in previous sections, on top of each plot are two values, $[x, y]$, with x representing the Pearson Partial Rank Correlation Coefficient (PRCC) value, and y representing the corresponding p-value. The parameters with large PRCC values (> 0.5 or < -0.5) as well as corresponding small p-values (< 0.05) are the most influential in the model. The PRCC plots show that there exists a strong linear relationship between the objective functional and the LHS parameters, e_2, γ_2, B_{L0} and S_0.
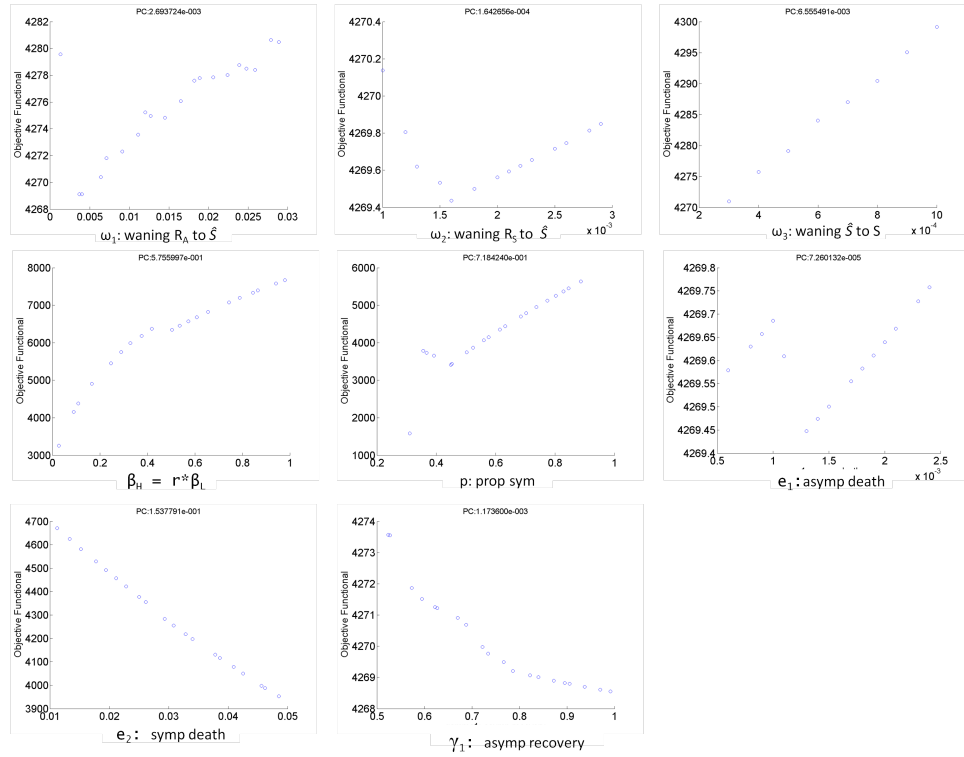
Figure 4.1: Monotonicity plots for ω_1, ω_2, ω_3, p, β_H, e_1, e_2, and γ_1
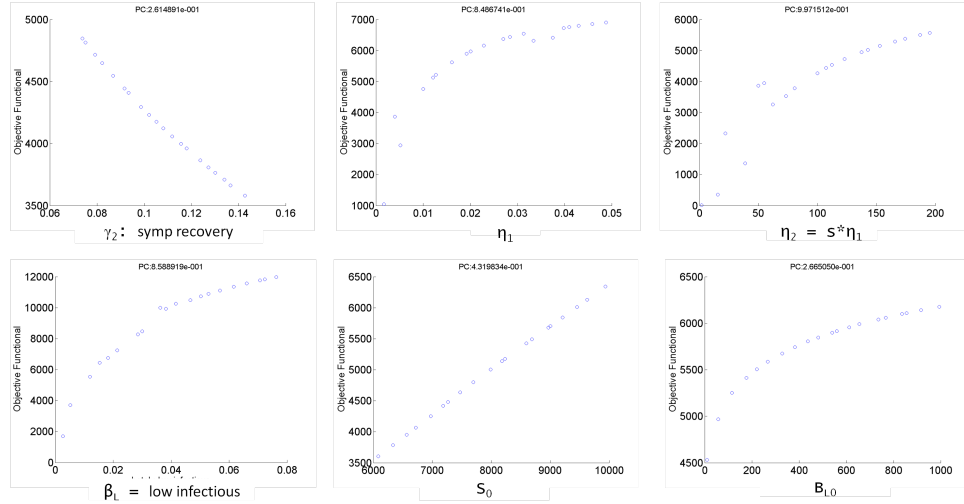
Figure 4.2: Monotonicity plots for γ_2, η_1, β_L, η_2, B_{L0}, and S_0

37
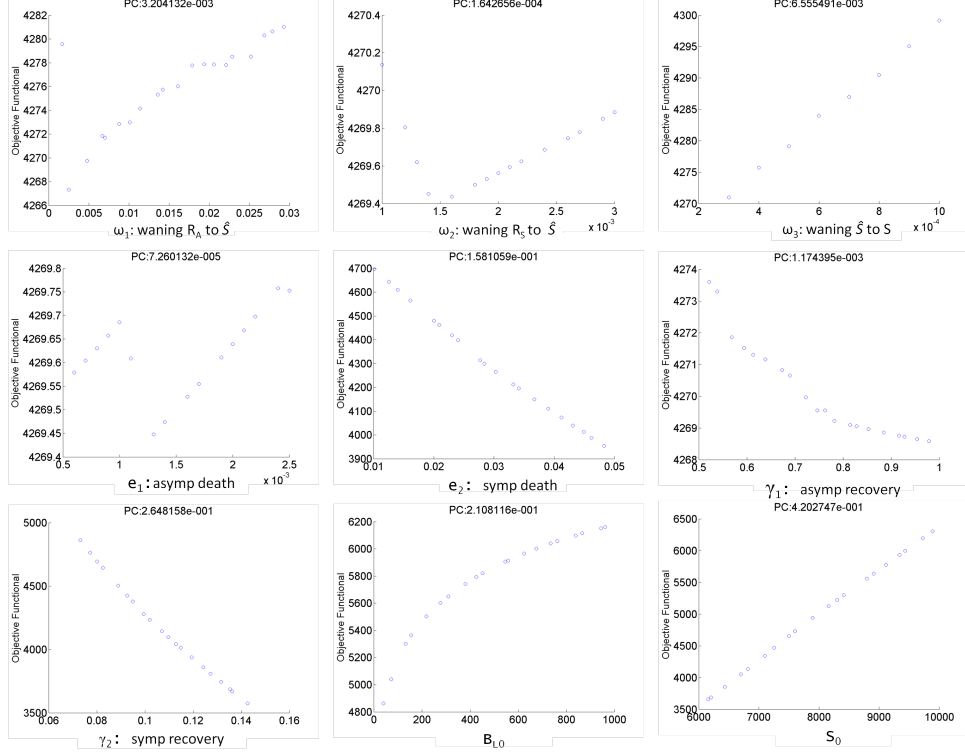
Figure 4.3: Monotonicity plots for ω_1, ω_2, ω_3, e_1, e_2, γ_1, γ_2, B_{L0}, and S_0

These results are tabulated in Table 4.2. In the table, (*) is used to indicate possible contributors (PRCC values: $\sim$0.5 to 0.69 or -0.5 to -0.69), (**) is used to indicate very likely contributors to uncertainty (PRCC values: $\sim$0.7 to 0.79 or -0.7 to -0.79), and (***) is used to indicate highly likely contributors to uncertainty (PRCC values: $\sim$0.8 to 0.99 or -0.8 to -0.99). Also as before, important contributors to uncertainty have both their PRCC values (orange) and their p-values (grey) highlighted, not just one or the other. Following this notation, we observe that the parameter e_2 is highly influential in the model and γ_2, B_{L0} and S_0 are even more so. Of the nine parameters considered in the LHS/PRCC process, we conclude that with respect to the objective functional, these four parameters are the most sensitive parameters in the model.
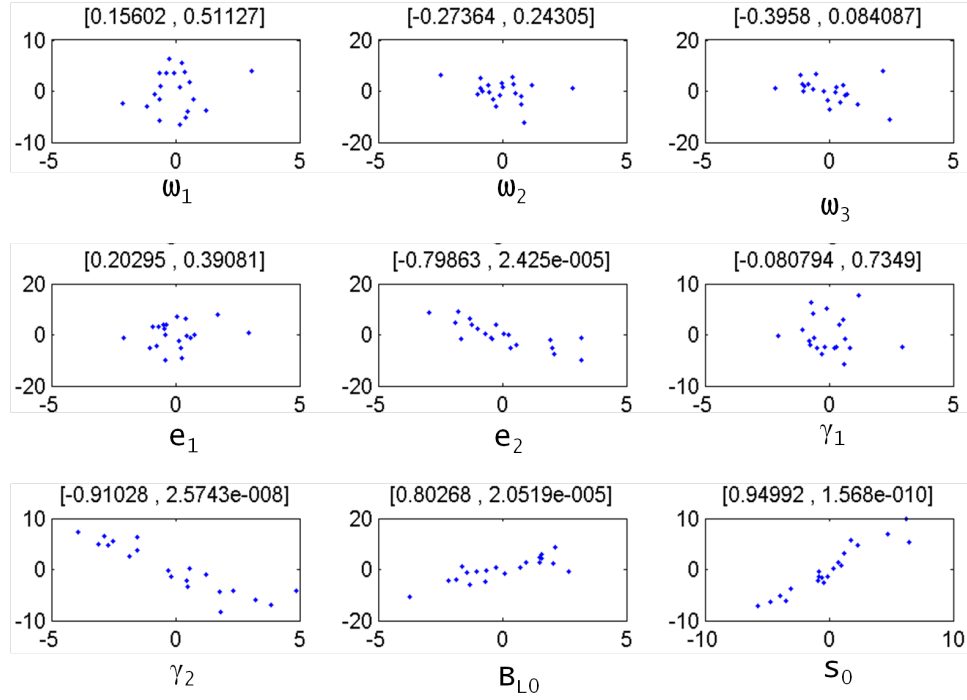
Figure 4.4: PRCC Plots for Objective Functional.

Table 4.2: Output from PRCC Analysis

Parameters	Objective Functional	
	$PRCC$	$p\text{-}value$
ω_1: Waning R_A to $\hat{S}$	0.15602	0.51127
ω_2: Waning R_S to $\hat{S}$	-0.27364	0.24305
ω_3: Waning $\hat{S}$ to S	-0.3958	0.084087
e_1: asymp death	0.20295	0.39081
e_2: symp death	**-0.79863	2.43E-05
γ_1: asymp recovery	-0.080794	0.7349
γ_2: symp recovery	***-0.91028	2.57E-08
B_{L0}	***0.80268	2.05E-05
S_0	***0.94992	1.57E-10

Conclusion

Our study of the influence of the LHS parameters on the objective functional is useful for decision makers interested in knowing the parameters in a model to which the objective functional value is most sensitive. The work in this thesis is novel due to using the objective functional value at the optimal control as the outcome measure in the LHS/PRCC technique. Using this technique, we can narrow down our parameter list to the most sensitive with respect to the objective functional value and then study their effect on the optimal control when we vary them.

It is of concern that when computing the objective functional values using the forward-backward sweep method, there were some parameters for which the method did not converge. For future utility of the code created for this analysis, the best outcome would be for us to be able to study the sensitivity of any number of parameters and have the simulations correctly predict the most influential parameters. In general, the influence of the parameters on a model is a composite effect. Consequently, when we remove parameters from the analysis, we lose information on them and are not able to tell if they would have been influential. Thus, for future study, further investigation needs to be done to obtain convergence of the method used for all uncertain parameters.

Bibliography

Bibliography

[1] King A. A., Ionides E. L., Pascual M., and Bouma M. J. Inapparent infections and cholera dynamics. *Nature*, 454(7206):877–880, 2008.

[2] McKay M. D., Beckman R. J., and Conover W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[3] Merrell S. D., Butler S. M., Qadri F., Dolganov N. A.and Alam A., Cohen M. B., Calderwood S. B., Schoolnik G. K., and Camilli1 A. Host-induced epidemic spread of the cholera bacterium. *Nature*, 417:642–645, 2002.

[4] Hamby D.M. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment*, 32(2):135–154, 1994.

[5] International Atomic Energy Agency (IAEA). Evaluating the reliability of predictions made using environmental transfer models, 1989. Vienna: Safety Series No. 100. Report No. STI/PUB/835.

[6] Conover W. J. and Iman R. L. Rank transformations as a bridge between parametric and non-parametric statistics. *The American Statistician*, 35(3):124–129, 1981.

[7] Iman R. L. and Helton J. C. An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Analysis*, 8(1):71–90, 1988.

[8] Iman R. L. and Conover W. J. The use of the rank transform in regression. *Technometrics*, 21(4):499–509, 1979.

[9] Pontryagin L.S., Boltyanskii V.G., Gamkrelize R.V., and Mishchenko E.F. *The Mathematical Theory of Optimal Processes.* Wiley, New York, 1967.

[10] Blower S. M., Hartel D., Dowlatabadi H., Anderson R. M., and May R. M. Drugs, sex and HIV: A mathematical model for New York city. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 331(1260):171–187, 1991.

[11] Blower S. M. and Dowlatabadi H. Sensitivity and uncertainty analysis of complex models of disease transmission: An HIV model, as an example. *International Statistical Review / Revue Internationale de Statistique*, 62(2):229–243, 1994.

[12] Hartley D. M., Morris Jr. J. G., and Smith D. L. Hyperinfectivity: A critical element in the ability of cholerae to cause epidemics? *PLoS Med*, 3(1):e7, 2005.

[13] World Health Organization. Cholera annual report 2010. *Weekly Epidemiological Record, 2011*, 86(31):325–340, 2010.

[14] World Health Organization. Cholera surveillance and number of cases. http://www.who.int/topics/cholera/surveillance/en/, 2012.

[15] Zhong P. *Optimal Theory Applied in Integrodifference Equation Models and in a Cholera Differential Equation Model.* PhD thesis, The University of Tennessee, Knoxville, August 2011.

[16] Miller N. R., Schaefer E., Gaff H., Fister K., and Lenhart S. Modeling optimal intervention strategies for cholera. *Bulletin of Mathematical Biology*, 72(8):2004–2018, 2010.

[17] Lenhart S. and Workman J. T. *Optimal control applied to biological models.* Chapman & Hall, London, 2007.

[18] Marino S., Hogue I. B., Ray C. J., and Kirschner D. E. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254(1):178–196, 2008.

[19] Richard Taylor. Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography*, 6(1):35–39, 1990.

[20] Fleming W.H. and Rishel R.W. *Deterministic and Stochastic Optimal Control.* Springer, Berlin, 1975.

Vita

Boloye Gomero was born in Sapele, Nigeria, to Patrick Gomero and Gift Gomero. She completed her high school education at International Secondary School (ISS) in 2001 in Port Harcourt, Nigeria. She also completed her undergraduate education at Marymount University in Arlington, Virginia, where she pursued a double major in Mathematics and Biology from 2004 to 2008. Thereafter, Miss Gomero spent a year working on mouse genetics research at the Oak Ridge National Laboratory and subsequently pursued a masters degree in Mathematics at the University of Tennessee from 2010 to 2012.

Miss Gomero's graduate education was partially funded by graduate teaching assistantship and while on this assistantship, she taught three semesters of "Introduction to Mathematical Reasoning." Miss Gomero's graduate studies were also supported for two semesters by funding from the National Institute for Mathematical and Biological Synthesis (NIMBioS), where she served as a graduate research assistant, and partially funded by NSF funding on Dr. Lenhart's subcontract with Murry State University on cholera modeling.

In August 2012, Miss Gomero graduated with her masters in Mathematics. She plans to teach Mathematics at the high school level. Starting in August 2011, she will be teaching Geometry and Algebra II full-time at Webb School of Knoxville in Knoxville, Tennessee.