



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2021

Bayesian Experimental Design For Bayesian Hierarchical Models With Differential Equations For Ecological Applications

Rebecca Atanga

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Statistical Methodology Commons](#)

© The Author

Downloaded from

<https://scholarscompass.vcu.edu/etd/6531>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

©Copyright 2021, Rebecca E. Atanga

BAYESIAN EXPERIMENTAL DESIGN FOR BAYESIAN HIERARCHICAL MODELS
WITH DIFFERENTIAL EQUATIONS FOR ECOLOGICAL APPLICATIONS

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor
of Philosophy at Virginia Commonwealth University.

by

REBECCA E. ATANGA

B.S. in Applied Mathematics, University of Pittsburgh at Greensburg, 2014

M.S. in Statistics, Virginia Commonwealth University, 2020

Director: Dr. Edward L. Boone,

Professor, Department of Statistical Sciences and Operations Research

Virginia Commonwealth University

Richmond, Virginia

May, 2021

Acknowledgements

I would like to thank my advisors Dr. Edward L. Boone and Dr. Ryad A. Ghanam, without whom I would not have been able to complete this dissertation. Your encouragement, guidance and continuous support over the past four years has been invaluable. I would like to thank my committee members Dr. Ben Stewart-Koster, Dr. David J. Edwards and Dr. Jenise L. Swall for the insightful feedback and appraisals that have contributed to this research. I would also like to thank Dr. D'Arcy Mays as a professor and the department chair for his kind and wise words of advice that have helped me in this endeavor. I am grateful to my mom and sisters for their love and encouragement as I have pursued my goals. I also greatly appreciate my husband, Prosper, for believing in me and supporting me with his enduring love. Finally, I thank God for all of the blessings he has bestowed upon me and for placing all of these wonderful people in my life. Without his steadfast love none of this would be possible.

Table of Contents

Chapter	Page
Acknowledgements	i
Table of Contents	ii
List of Tables	iii
List of Figures	iv
Abstract	xi
1 Introduction	1
1.1 Statistical Approach to Ecological Modeling	1
1.2 Population Dynamics	4
1.2.1 The Logistic Equation	4
1.2.2 The Lotka-Volterra Differential Equations	6
1.3 Simulation Techniques	9
1.3.1 Metropolis-Hastings Algorithm	9
1.3.2 Gibbs Sampler	10
1.3.3 Simulated Annealing	11
1.4 Optimal Experimental Design	12
2 Bayesian Experimental Design with the Logistic Equation for Population Growth Applications	15
2.1 Introduction	15
2.2 Logistic Growth Model	17
2.3 Methods	18
2.3.1 Statistical Model	18
2.3.2 Optimal Designs	20
2.3.3 Sequential Optimality	22
2.4 Simulation Study	34
2.4.1 Simulated Annealing	35
2.4.2 Sequential Design	45

2.4.3 Potential Difficulties	62
2.5 Discussion	64
2.6 Conclusions	66
3 Bayesian Experimental Design with the Lotka-Volterra Differential Equations for Predator-Prey Dynamics	68
3.1 Introduction	68
3.2 Predator-Prey Dynamics	69
3.3 Statistical Model	71
3.4 Experimental Design	73
3.5 Sequential Optimality	74
3.6 Simulation Study	75
3.7 Conclusions	82
4 Resource-based Sequential Bayesian Experimental Design for Dynamic Mod- eling	84
4.1 Introduction	84
4.2 Statistical Models	85
4.2.1 The Logistic Equation	86
4.2.2 The Lotka-Volterra Differential Equations	87
4.3 Sequential Optimality	88
4.4 Simulation Study	89
4.4.1 Logistic Growth	89
4.4.2 Predator-Prey Dynamics	100
4.5 Discussion	109
5 Closing Remarks and Future Work	111
VITA	125

List of Tables

Table	Page
1 Logistic Growth Scenarios	34
2 I-Optimal Design Steps for Normal Growth	90
3 Normal Logistic Growth Designs	92
4 Fast Logistic Growth Designs	94
5 Slow Logistic Growth Designs	96
6 Comparing Fit of Designs	98
7 I-Optimal Design Steps Exploring Window of Size 10	101
8 Lotka-Volterra Designs Exploring Windows of Size 10	104
9 Lotka-Volterra Designs Exploring Windows of Size 15	106
10 Lotka-Volterra Designs Exploring Windows of Size 5	108

List of Figures

Figure	Page
1 An initial design of three points is plotted in blue. The ground truth model of normal 10% growth is plotted as a black curve across one hundred time points with a carrying capacity of two thousand and initial population of two hundred.	24
2 A window of ten candidate points are selected from the last point sampled illustrated by the red region.	26
3 Each point is evaluated using A optimality criterion in the window from Figure 2 and the argument with the minimum trace in the window is selected.	27
4 The fifth design point is selected in panel (a) found in a window following the design from Figure 2: panel (b). Panels (b)-(f) represent each additional design point added until the budget of ten points is exhausted. Each panel plots the ground truth model in black, the design points in blue and fits the design with a red curve based on the parameter estimates. The red dotted lines represent the 2.5% and 97.5% prediction quantiles.	33
5 10,000 MCMC samples are used to simulate A_{Φ} - optimal designs for (a) normal, (b) fast and (c) slow growth models. All models are simulated across 100 time points with a carrying capacity of 2000. The ground truth models are plotted in black. The optimal design points are plotted in blue and fit by a solid red curve, where the 2.5% and 97.5% quantiles of the predicted values are plotted as red dotted lines.	38
6 10,000 MCMC samples are used to simulate D_{Φ} - optimal designs for (a) normal, (b) fast and (c) slow growth models. All models are simulated across 100 time points with a carrying capacity of 2000. The ground truth models are plotted in black. The optimal design points are plotted in blue and fit by a solid red curve, where the 2.5% and 97.5% quantiles of the predicted values are plotted as red dotted lines.	41

7	10,000 MCMC samples are used to simulate I-optimal designs for (a) normal, (b) fast and (c) slow growth models. All models are simulated across 100 time points with a carrying capacity of 2000. The ground truth models are plotted in black. The optimal design points are plotted in blue and fit by a solid red curve, where the 2.5% and 97.5% quantiles of the predicted values are plotted as red dotted lines.	44
8	The ground truth normal growth model is plotted as a black curve. I-optimality criterion guides the sequential optimality process. The design points are plotted in blue and are fit by a solid red curve. The red dotted lines represent the 2.5% and 97.5% predicted quantiles. Panel (a) plots the base design, panel (b) fits the design, and panels (c)-(i) plot the optimal points as they are fit.	54
9	10,000 MCMC samples are used in this simulation of a normal growth curve increasing at a 10% rate across 100 time points with a carrying capacity of 2000. Sequential optimality is used to find the ten best points for (a) I, (b) A_Φ and (c) D_Φ optimal designs. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.	55
10	10,000 MCMC samples are used in this simulation of a rapid growth curve increasing at a 100% rate across 100 time points with a carrying capacity of 2000. Sequential optimality is used to find the ten best points for (a) I, (b) A_Φ and (c) D_Φ optimal designs. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.	56
11	10,000 MCMC samples are used in this simulation of a slow growth curve increasing at a 5% rate across 100 time points with a carrying capacity of 2000. Sequential optimality is used to find the ten best points for (a) I, (b) A_Φ and (c) D_Φ optimal designs. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.	57

12	Bayesian optimality criteria are used to guide the sequential optimality process. The frequency tables plot the probabilities assigned to the first ten candidate points following the initial base design for a normal growth model. Panel (a) represents candidate probabilities under the \bar{U}_I criterion. Panel (b) plots probabilities of the candidates for \bar{U}_A criterion. Panel (c) provides the frequencies associated with \bar{U}_D criterion.	59
13	10,000 MCMC samples simulate a normal growth curve increasing at a 10% rate across 100 time points with a carrying capacity of 2000. Sequential optimality is used with the Bayesian design criteria to find the ten best points for (a) \bar{U}_I , (b) \bar{U}_A and (c) \bar{U}_D optimal designs. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.	61
14	10,000 MCMC samples are used to simulate normal 10% logistic growth across a 100 day period with a carrying capacity of 2000. The ground truth model is plotted as a black curve. I-optimality criterion guides the sequential optimality process. The optimal design points are plotted in blue and are fit by a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values. Panel (a) plots the base design of three points, panel (b) plots the fit of the base design, and panels (c)-(f) plot the fourth through seventh optimal points in the design as they are fit. This simulation demonstrates a lack of fit scenario.	64
15	The Lotka-Volterra differential equations are simulated using Runge-Kutta methods with parameter values $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The numerical solutions of the prey and predator populations are plotted as blue and orange curves respectively ranging from zero to ten thousand across one hundred time points.	70
16	I-optimality criterion is used to select the blue design points. The design is fit by a solid red curve and the true values are plotted as dashed grey curves. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Panel (a) plots the initial design, panel (b) fits the design, and panels (c)-(l) represent each optimal point added to the design selected from a window of size ten. The final design consists of fifteen points.	77

17	The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size ten to select a final design of size fifteen. The panels plot the final designs using (a) I-optimality, (b) A-optimality, and (c) D-optimality criteria.	79
18	The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size fifteen to select a final design of size ten. The panels plot the final designs using (a) I-optimality, (b) A-optimality, and (c) D-optimality criteria.	80
19	The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size five to select a final design of size twenty. The panels plot the final designs using (a) I-optimality, (b) A-optimality, and (c) D-optimality criteria.	81
20	Sequential optimality plots the base I-optimal design (a), the fit of the base design (b), and the fourth to tenth points as they are added to the design (c)-(i) for a normal growth model. The ground truth model is plotted as a black curve. The design points are plotted in blue and are fit by a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values.	91
21	Sequential optimality is used to find (a) I, (b) A and (c) D optimal designs across normal logistic growth models. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.	93

22	Sequential optimality is used to find (a) I, (b) A and (c) D optimal designs across fast logistic growth models. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.	95
23	Sequential optimality is used to find (a) I, (b) A and (c) D optimal designs across fast logistic growth models. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.	97
24	Sequential optimality is used to find (a) I, (b) A and (c) D optimal designs for a normal logistic growth model. Panel (d) compares convenience sampling by plotting a randomly generated design. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model. The designs are compared in Table 6.	99
25	I-optimality criterion is used to select the blue design points. The design is fit by a solid red curve and the true values are plotted as dashed grey curves. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Panel (a) fits the initial design, and panels (b)-(l) represent each optimal point added to the design selected from a window of size ten until the dynamics are captured.	103
26	The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size ten. The panels plot the final designs using (a) I, (b) A, and (c) D optimality criteria.	105
27	The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size fifteen. The panels plot the final designs using (a) I, (b) A, and (c) D optimality criteria.	107

28 The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size five. The panels plot the final designs using (a) I, (b) A, and (c) D optimality criteria. 109

Abstract

BAYESIAN EXPERIMENTAL DESIGN FOR BAYESIAN HIERARCHICAL MODELS WITH DIFFERENTIAL EQUATIONS FOR ECOLOGICAL APPLICATIONS

By Rebecca E. Atanga

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor
of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2021.

Director: Dr. Edward L. Boone,

Professor, Department of Statistical Sciences and Operations Research

Ecologists are interested in the composition of species in various ecosystems. Studying population dynamics can assist environmental managers in making better decisions for the environment. Traditionally, the sampling of species has been recorded on a regular time frequency. However, sampling can be an expensive process due to financial and physical constraints. In some cases the environments are threatening, and ecologists prefer to limit their time collecting data in the field. Rather than convenience sampling, a statistical approach is introduced to improve data collection methods for ecologists by studying the dynamics associated with populations of interest. Population models including the logistic equation and the Lotka-Volterra differential equations are employed to simulate species composition. This research focuses on sequentially learning about the behavior of dynamical systems to better inform ecologists of when to sample. The developed algorithm of sequential optimality designs sampling regimes to assist ecologists with resource allocation while providing maximum information from the data. This research in its entirety constructs a method for designing sampling schedules for ecologists based on the dynamics associated with temporal ecological models.

CHAPTER 1

INTRODUCTION

1.1 Statistical Approach to Ecological Modeling

Statistical inference refers to the use of statistics to draw conclusions about the unknown aspects of a population based on a sample. Sampling is an important stage in research given that studying whole populations is rarely practical, efficient or ethical according to Marshall (1996). Specifically in ecology, population and community studies rely heavily on sampling procedures (Green, 1979; Greig-Smith, 1983). Sampling is the first stage of ecological research and directly impacts the entire study. Ecologists are interested in improving sampling techniques given the nature of ecological data. Bowser (1986) separates historic ecological data into three categories as either planned, opportunistic or serendipitous. Planned data is defined as well-documented long-term records involving continuous data collection over an extended period of time. Opportunistic data are used for short-term goals typically limited by funding as seen in the literature. Whereas, serendipitous data are not collected for scientific intent and are usually used for census purposes. In practice, ecologists are most concerned with planned data and tend to follow the procedures of classical sampling theory (Cochran, 1977).

Issues arise when collecting ecological data due to inconsistencies between theory and practice (Albert et al., 2010). Classical sampling theory is impractical in ecology due to the fact that resources, accessibility and time do not permit excessive sampling. Thus, convenience sampling and simplified methods have become more commonplace but may disregard

many ecological questions. Population ecologists prefer to study the distribution of individuals in a population over time and space (Williams et al., 2002). Specifically, ecologists are interested in estimating population dynamics for spatial and temporal models. Techniques such as maximum likelihood and ordinary-least squares are commonly used. Bulmer (1974) uses maximum likelihood estimates to compare the log-normal and Poisson distributions when fitting species-abundance data. Hilborn (1990) studies the population dynamics and movement of tagged fish in spatial locations over time using maximum likelihood estimators. Whereas, ordinary least squares estimates are typically used for regression analysis of factors affecting a population. As seen in the literature, traditional methods are considered stronger and more robust based on their theoretical foundation. However, estimating ecological models is not limited to the frequentist framework.

Ecologists are shifting towards Bayesian modeling of natural systems based on the increasing resources for ecological Bayesian models (Hobbs and Hooten, 2015; Hooten and Hobbs, 2015). Bayesian analysis combines prior beliefs with sample information to make inferences about the *posterior* belief of the population parameter. The posterior is a conditional probability of an unknown parameter given the observance of data defined by Bayes' Theorem (Bayes, 1763). For more on Bayesian modeling see Anscombe (1961); Box and Tiao (1973); Ellison (2004). Using Bayesian inference, the probability of the model parameters can be predicted based on simulated data by employing Markov Chain Monte Carlo (MCMC) methods (Robert and Casella, 2013). Heavy MCMC sampling provides substantial representation of the model helping ecologists to learn about the system and identify sampling needs as well as where gaps in knowledge occur.

In practice, simulations are performed due to the limitations involved with collecting ecological data. Improving sampling techniques for ecologists can be done by using experimental design methods (Fisher et al., 1960). Specifically, optimal designs can be used to sample based on the dynamics associated with a population of interest. Though large sample sizes provide more information, sampling according to the population dynamics can

decrease the number of samples needed to accurately represent a model. According to the literature, designing experiments in practice can be quite difficult. McAllister and Peterman (1992) discuss issues with experimental design and possible ways to improve experimentation applied to fisheries management. In more recent years, Williams et al. (2018) demonstrate how to allocate resources for monitoring ecological processes using design-based inference for spatiotemporal models. In both cases, optimal designs improve estimation for abundance and population dynamics. This research specifically focuses on determining optimal sampling regimes that capture population dynamics of temporal models using various optimality criteria.

Not many disciplines have conducted research involving the prediction of future samples. Most fields study the prediction capabilities of a specified model. Houlahan et al. (2017) emphasizes that prediction is the core of all disciplines and demonstrates true scientific understanding. Similar to work by Pagendam and Pollett (2009), the goal of this research is to develop a methodology that can predict optimal design points by learning about dynamical systems in a sequential manner using specified design criteria. This dissertation is formatted with an introduction followed by three chapters each of which are individual manuscripts and concludes with chapter five. The first paper and second chapter of this dissertation introduces an approach that predicts optimal samples based on the dynamics associated with a simple univariate model. The second paper and third chapter further develops the methodology by applying the technique to a complex model involving a pair of first-order nonlinear differential equations. Finally, the third paper and fourth chapter focuses on determining a stopping criterion that can halt the process once the dynamics of the system are captured. Thus, repetition will occur throughout this dissertation, but the research as a whole completely develops an algorithm that ecologists can implement to improve statistical practices and optimize resource allocation.

1.2 Population Dynamics

This research focuses on temporal models that are commonly used in ecology. Population growth models are frequently studied by ecologists given the numerous applications and the simplistic nature of the models. Though easy to implement, simple models can be unrealistic. As a way to incorporate more complex dynamics, ecologists are also interested in studying predator-prey relationships as these models represent realistic environmental encounters. In this section, simple and complex population dynamics are introduced using well-known ecological models.

1.2.1 The Logistic Equation

Assessing population growth is essential when studying ecology and wildlife management. Growth models track the population dynamics of a single species across time, making them quite simple. Tsoularis and Wallace (2002) believe the simplest realistic model of population dynamics is exponential growth. Exponential growth follows the biological assumptions that in a population all individuals are genetically identical, no immigration or emigration occurs, the growth rate per capita is not time dependent and there are unlimited resources available. Given the aforementioned assumptions, exponential growth can be written in the mathematical form

$$\frac{dN}{dt} = rN \tag{1.1}$$

with solution

$$N(t) = N_0 e^{rt},$$

where r represents the growth rate per capita and N_0 is the population size at time $t = 0$. If the rate parameter r is negative, the solution equates to extinction making the model useless. On the other hand, a positive parameter r would result in uncontrolled increases in population making the model unrealistic and impractical in the context of ecological data. Real data has shown that growth rates eventually decline as the population begins to

approach an asymptotic level (Eberhardt, 1977; Eberhardt and Siniff, 1977). Ecologists thus need a better way to assess population growth other than using the exponential model.

The logistic equation was first developed by Verhulst (1838). The rediscovery of the equation by Pearl (1925) is popularly used in ecology and represents the population dynamic where the reproduction of a species is proportional to both the current population and resource availability. Tsoularis and Wallace (2002) explain logistic growth as an augmented version of exponential growth, where a factor of $1 - (N/K)$ represents the deficit of the current population from the carrying capacity

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \quad (1.2)$$

with solution

$$N(t) = \frac{KN_0}{(K - N_0)e^{-rt} + N_0}.$$

The parameter K represents the maximum level that the population can reach. All other variables and parameters are the same as those defined in Equation 1.2.1 for exponential growth. Population dynamics for the logistic equation show a low growth rate when the population size is close to zero or the carrying capacity. Alternatively, the highest growth rate for the logistic curve occurs at the inflection point when the population size is half of the carrying capacity, $N(t) = K/2$. Since this model is a dynamical system, steady-state conditions may exist. However, the two equilibrium states for logistic growth occur during extinction or maximum population capacity. Though there exists prior steady-state knowledge of the system, the existence of an analytic solution provides more information when specifying the model in the Bayesian framework.

Specifying Equation 1.2.1 as a Bayesian model requires the specification of the likelihood of the experiment and any prior knowledge of the parameters. The logistic equation records the population count of a single species across time. Although the model could be distributed in many ways, the Poisson assumption (Chen, 1987) best represents a sequence of counts

observed over a fixed period of time. This implies a Poisson likelihood for the model

$$N_i|t_i \sim \text{Poisson}(\lambda_i|t_i).$$

$N_i|t_i$ is the population at time $t = i$, and $\lambda_i = \frac{KN_0}{(K-N_0)e^{-rt_i}+N_0}$.

As for prior knowledge, the carrying capacity K represents a maximum population size, which requires a positive parameter value. The growth rate parameter r is assumed to be positive as well to ensure population growth. Due to these constraints, the log-normal distribution can be used, which has been classically studied by Aitchison and Brown (1957). The log-normal distribution is known for being skewed to the right and starting at zero, implying a positive support that increases to the mode and decreases thereafter. Given these unique properties, the prior distributions for the parameters are specified in this research as

$$k \sim \text{Lognormal}(\mu_k, \sigma_k^2)$$

$$K \sim \text{Lognormal}(\mu_K, \sigma_K^2).$$

The mean μ and variance σ^2 for the prior distribution of each parameter can change depending on expert knowledge provided by ecologists. This specified Bayesian model provides a general introduction to logistic growth as it will be employed later in this research.

1.2.2 The Lotka-Volterra Differential Equations

Lotka (1926) and Volterra (1928) developed the Lotka-Volterra differential equations as a model for the interactions between a predator and its prey. This dynamic behavior having such broad applications has been frequently modeled among ecologists. The traditional Lotka-Volterra model captures realistic encounters between one predator and one prey

$$\begin{aligned} \frac{dx}{dt} &= \alpha x - \beta xy \\ \frac{dy}{dt} &= \delta xy - \gamma y \end{aligned} \tag{1.3}$$

where x and y represent the prey and predator populations respectively. The parameter α represents the birth rate of the prey, β is the predation success rate, δ is the efficiency of converting prey into predators and γ represents the mortality rate of the predator. Unlike Equation 1.2.1, the Lotka-Volterra differential equations do not have an analytic solution. However, a stability analysis of the dynamical system (Hale and Koçak, 2012) can ensure specified parameter conditions.

The Lotka-Volterra model is considered stable at certain equilibrium points in the system. Performing a stability analysis finds these stable points mathematically. By setting both the derivatives equal to zero

$$\alpha x - \beta xy = 0$$

$$\delta xy - \gamma y = 0$$

two solutions exist. The first solution to the system is

$$\{x = 0, y = 0\},$$

implying a scenario of extinction. Given extinction of species is a key component in the study of evolution, many ecologists study this steady-state. For instance, Parker and Kamenev (2009) study the stability of the Lotka-Volterra model and how the populations are driven towards extinction. In some cases though, ecologists may consider exponential decay of each species rather than continuing to study this model. The second solution to the system is

$$\left\{x = \frac{\gamma}{\delta}, y = \frac{\alpha}{\beta}\right\},$$

representing a continually oscillating behavior between the two species. This consistent fluctuation between predator and prey is commonly studied as a more realistic solution. This fixed point solution models both populations maintaining a non-zero state indefinitely. The parameter values of α , β , δ and γ can further be investigated to achieve this equilibrium state.

Evaluating the stability of the fixed point $\{x = \frac{\gamma}{\delta}, y = \frac{\alpha}{\beta}\}$ can be done using the Jacobian matrix. The Jacobian of the Lotka-Volterra differential equations is

$$J(x, y) = \begin{bmatrix} \alpha - \beta y & -\beta x \\ \delta y & \delta x - \gamma \end{bmatrix}.$$

Using this matrix, the solutions of the oscillating state are substituted into the Jacobian

$$J\left(\frac{\gamma}{\delta}, \frac{\alpha}{\beta}\right) = \begin{bmatrix} 0 & -\frac{\beta\gamma}{\delta} \\ \frac{\alpha\delta}{\beta} & 0 \end{bmatrix},$$

which can be solved for the eigenvalues

$$\begin{aligned} \lambda_1 &= i\sqrt{\alpha\gamma} \\ \lambda_2 &= -i\sqrt{\alpha\gamma}. \end{aligned}$$

Given purely imaginary conjugate eigenvalues ensures an elliptic or periodic solution for the fixed point (Strogatz, 2018). Based on the knowledge of the system, the Lotka-Volterra differential equations can be specified as a Bayesian model.

Equation 1.2.2 has more complex dynamics that can be specified as a Bayesian model. This model as stated previously is temporal and counts the population size of two species making the likelihood of the predator and prey

$$x_i|t_i \sim \text{Poisson}(\lambda_i|t_i)$$

$$y_i|t_i \sim \text{Poisson}(\lambda_i|t_i).$$

$x_i|t_i$ is the population of the prey at time $t = i$, and $y_i|t_i$ is the population of the predator at time $t = i$. As for the priors, each parameter represents a positive rate of change. Thus, the conjugate prior, Gamma distribution, to the Poisson distribution can be used to define

the priors of the parameters

$$\alpha \sim \Gamma(\alpha_\alpha, \beta_\alpha)$$

$$\beta \sim \Gamma(\alpha_\beta, \beta_\beta)$$

$$\delta \sim \Gamma(\alpha_\delta, \beta_\delta)$$

$$\gamma \sim \Gamma(\alpha_\gamma, \beta_\gamma).$$

The hyper-parameters for the prior distributions will change according to the expert knowledge provided by experts. For now, a general Bayesian model has been defined and can be simulated using Bayesian sampling techniques.

1.3 Simulation Techniques

1.3.1 Metropolis-Hastings Algorithm

Markov Chain Monte Carlo (MCMC) sampling techniques (Gilks et al., 1995) can be used for simulation purposes. Among the MCMC techniques, the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) is broadly used in practice. Martino and Elvira (2014) support the Metropolis-Hastings sampler as the core of MCMC sampling and state that the algorithm consists of three main elements: the candidate density $q(z|x)$, the acceptance probability $\alpha(x, z)$, and the target function $\pi(x)$. Based on the acceptance ratio $\alpha(x, z)$, the algorithm proceeds by randomly attempting to move about the sample space sometimes accepting points and alternatively remaining in place. Ultimately a set of samples $x^{(1)}, \dots, x^{(T)}$ or a subset of them is returned. As t grows ($t \rightarrow \infty$), the density of the current state $x^{(t)}$ converges to the target density $\pi(x)$, implying that large quantities of samples can provide a substantial representation of the model.

Algorithm 1: Metropolis-Hastings

Begin**Initialization:** Choose an initial state $x^{(0)}$ **For** $t = 1, \dots, T$:(a) Draw a sample $z' \sim q(x|x^{(t-1)})$ (b) Accept the new state, $x^{(t)} = z'$, with probability:

$$\alpha(x^{(t-1)}, z') = \min\left[1, \frac{\pi(z')q(x^{(t-1)}|z')}{\pi(x^{(t-1)})q(z'|x^{(t-1)})}\right]$$

Otherwise, set $x^{(t)} = x^{(t-1)}$ **End**

For more on Metropolis-Hastings see Chib and Greenberg (1995); Hitchcock (2003).

1.3.2 Gibbs Sampler

In this research, the *Gibbs-sampler* is used to simulate population dynamics. Wakefield et al. (1994) promote Gibbs sampling as a method that can numerically solve complex linear and non-linear population models. First introduced by Geman and Geman (1984), the Gibbs sampler is a popular choice Markov Chain Monte Carlo algorithm. The algorithm generates posterior samples for each random variable by sampling from a conditional distribution of the remaining variables fixed to their current state. Consider random variables X_1, X_2, \dots, X_n , the algorithm begins by setting initial values of $x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}$ sampled from a prior distribution $q(x)$. At the i^{th} iteration, random variable X_1 is sampled $x_1^{(t)} \sim p(X_1 = x_1 | X_2 = x_2^{(t-1)}, \dots, X_n = x_n^{(t-1)})$, and X_2 is sampled $x_2^{(t)} \sim p(X_2 = x_2 | X_1 = x_1^{(t)}, X_3 = x_3^{(t-1)}, \dots, X_n = x_n^{(t-1)})$ and so on. The process continues until the sample values match the same distribution as the posterior joint distribution. The algorithm is given in detail as follows.

Algorithm 2: Gibbs-Sampler

Begin**Initialization:** Choose an initial state $x^{(0)} \sim q(x)$ **For** $t = 1, \dots, T$:

$$x_1^{(t)} \sim p(X_1 = x_1 | X_2 = x_2^{(t-1)}, X_3 = x_3^{(t-1)}, \dots, X_T = x_T^{(t-1)})$$

$$x_2^{(t)} \sim p(X_2 = x_2 | X_1 = x_1^{(t)}, X_3 = x_3^{(t-1)}, \dots, X_T = x_T^{(t-1)})$$

$$\vdots$$

$$x_T^{(t)} \sim p(X_T = x_T | X_1 = x_1^{(t)}, X_2 = x_2^{(t)}, \dots, X_{T-1} = x_{T-1}^{(t)})$$

End

For more on Gibbs sampling see Casella and George (1992); Smith and Roberts (1993); Gelfand and Smith (1990); Gelfand (2000).

1.3.3 Simulated Annealing

Simulated Annealing applied by Laarhoven and Aarts (1987) is an adaptation of the Metropolis algorithm and is used for a discrete design space to approximate the global optimum of a given function. Given that the Metropolis-Hastings algorithm returns a discrete number of samples, the simulated annealing algorithm by Metropolis et al. (1953) can be implemented to explore the sample space for an optimal design. The algorithm begins by defining an initial state x_0 . Then a temperature T is set, which resembles the boiling point in the metallurgy process of annealing. A neighboring state x_{new} is evaluated using a temperature dependent probability $P(x_0, x_{new}, T)$ which indicates moving to a lower energy state. The algorithm continuously repeats until the discrete budget k has been exhausted.

Algorithm 3: Simulated Annealing

Begin

For a finite set of iterations, choose an initial state x_0

For $k = 1, \dots, k_{max}$:

(a) Let $T = \frac{k}{k_{max}}$

(b) Pick a random neighbor, x_{new}

Jump to the new sample x_{new} with probability $P(x_0, x_{new}, T)$

$$\text{where } P(x_0, x_{new}, T) = \begin{cases} 1 & f(x_{new}) \geq f(x_0) \\ e^{\frac{f(x_{new}) - f(x_0)}{T}} & \text{otherwise} \end{cases}$$

End

Simulated annealing does not guarantee an optimal design but is relatively easy to implement. Typically, specific optimality criteria is incorporated as a condition for the algorithm to converge. For more on simulated annealing see Kirkpatrick et al. (1983); Goffe et al. (1994); Bohachevsky et al. (1986).

1.4 Optimal Experimental Design

Design of experiments was first used in early research by Wald (1943); Hotelling (1944); Elfving et al. (1952). However, Kiefer et al. (1958); Kiefer (1959); Kiefer and Wolfowitz (1959) are known for significant contributions in the area of optimal design. Optimal experimental designs are intended to improve the precision of statistical inferences. Specifically, traditional optimality criteria minimizes the variance of parameter estimates by maximizing the information matrix of the design. Among the various optimality criteria, A , D , and E optimal designs are quite popular. A -optimal designs minimize the trace of the inverse of the information matrix, which results in minimizing the average variance of the estimates of the model parameters. D -optimal designs minimize the covariance matrix, which equivalently maximizes the determinant of the information matrix. Whereas, E -optimal designs maximize the minimum eigenvalue λ of the information matrix. Given a design matrix denoted

by X , the optimality criteria can be written mathematically as

$$A - \text{Optimal: } \min \text{tr} ((X'X)^{-1})$$

$$D - \text{Optimal: } \max |X'X|$$

$$E - \text{Optimal: } \max \min \lambda_i, \text{ where } \lambda_i \text{ is the } i^{\text{th}} \text{ eigenvalue of } X'X.$$

This research focuses on A and D optimal designs since these criteria are easy to implement. Prediction based criteria can also be used for determining optimal experimental designs. I -optimal designs are known to minimize the average prediction variance over the entire design region (Atkinson et al., 2007). Of the possible predictive designs, I -optimality criterion best suits the aims of this research.

Considering the Population Dynamics previously introduced, the design matrices for each experiment may be difficult to calculate. Alternative criteria can be used for the intended purposes of this research. Rather than attempting to calculate the information matrix for each population model, the design criteria can be applied to the parameter covariance matrix. Replacing the information matrix with the covariance matrix updates the criteria denoted by Φ or other notations throughout subsequent chapters. Let $\Phi = \text{cov}(\hat{\beta}) = \text{Var}(\hat{\theta})$, then the criteria can be redefined

$$A - \text{Optimal: } \min \text{tr} (\Phi^{-1})$$

$$D - \text{Optimal: } \max |\Phi|.$$

Now, A -optimal designs minimize the trace of the inverse of covariance matrix, and D -optimal designs maximize the determinant of the covariance matrix. For more on traditional optimal designs see Hardin and Sloane (1993); Goos and Jones (2011); Montgomery (2017).

Bayesian optimal designs are considered solely in Chapter two of this dissertation. Bayesian analysis expresses experimental design as a decision problem that selects optimal designs by maximizing the expected utility Chaloner and Verdinelli (1995). Bayesian optimal designs can be written in the form of a utility function $U(d_{x_f}^*)$, where $d_{x_f}^*$ represents a design chosen from the design region d_{x_f} . The design region is explored using Bayes I, D, and A optimal designs written mathematically as follows.

Bayesian I-Optimal

$$\bar{U}_I(d_{x_f}^*) = \min \int_{\Theta} \min_{d_{x_f}} (Y_2 - E[Y_2|y_1, d_{x_f}])^2 p(Y_2|y_1, d_{x_f}) dY_2$$

Bayesian D-Optimal

$$\bar{U}_D(d_{x_f}^*) = \min \int_{\Theta} \min_{d_{x_f}} |Cov(\theta|Y_2, y_1, d_{x_f})| p(Y_2|y_1, d_{x_f}) dY_2$$

Bayesian A-Optimal

$$\bar{U}_A(d_{x_f}^*) = \min \int_{\Theta} \min_{d_{x_f}} tr(\theta|Y_2, y_1, d_{x_f}) p(Y_2|y_1, d_{x_f}) dY_2$$

Y_2 represents the predicted designs, while y_1 represents the current design. Θ is the parameter space including the parameters θ from the model. Similar to the traditional criteria, the Bayesian optimality criteria are minimizing the posterior predictive distribution across the parameter space Θ and design space d_{x_f} . The Bayesian I-optimal designs minimize the distance between the squared predictions. Bayesian D-optimal designs minimize the determinant of the covariance of the posterior predictive distribution. The Bayesian A-optimal designs minimize the trace of the posterior predictive distribution. For more on Bayesian experimental design see Verdinelli (1992); Clyde (2001).

CHAPTER 2

BAYESIAN EXPERIMENTAL DESIGN WITH THE LOGISTIC EQUATION FOR POPULATION GROWTH APPLICATIONS

2.1 Introduction

Population growth models are commonly used in ecology when studying plants, animals, and organisms as they grow over time and interact with the environment. In earlier research, Hsu et al. (1984) evaluate seedling growth under laboratory and field conditions. According to Huang et al. (2016), early life-cycle events play critical roles in determining the dynamics of plant populations and their interactions with the surrounding environment. Germination patterns modeled by growth help ecologists understand the diversity, variation, and climate change within a system. Gamito (1998) further emphasize the importance of modeling growth and community dynamics using applications to fish populations. Population dynamics of fish are commonly studied to learn about water quality and the environment.

In riverine settings, the health of a waterway can be determined by the abundance of fish present. Kennard et al. (2005) study species that are highly tolerant of human-induced disturbances and consider certain fish as good indicators of river health. Ecologists model the population growth of indicator species to preserve habitats as land is developed along the banks of rivers. Not only can population growth models help preserve the environment, but also ecologists model growth of crops to improve economies. In Vietnam, rice-shrimp farming highly impacts the economic security of the country as these are the two main crops in the region. Leigh et al. (2017) study the environmental conditions of ponds used for

harvesting and address the risks that affect the survival and yields of crops. In order to improve year-round farming, ecologists monitor the growth and abundance of shrimp and rice.

Whether researchers are interested in improving yields for farmers, preserving the health of rivers or studying germination of seedlings, collecting ecological data has always been a time consuming and expensive process. Ecologists fishing in rivers face strenuous terrain, varying water levels, and dangerous wildlife and prefer to limit their time collecting samples. Though farmers raise crops in controlled environments, limiting sampling costs can increase revenue. During germination studies, scientists prefer to optimize their time and resources required to grow samples. In all cases, researchers prefer to reduce costs when collecting data by decreasing their necessary sample size without compromising the ability to accurately track growth.

Rather than traditional sampling methods, this study proposes an approach that designs optimal sampling regimes based on the dynamics associated with a population. Though traditional methods are often convenient, statistically designing an experiment can optimize sampling procedures for ecologists without compromising the information obtained from the data. As seen in Pagendam and Pollett (2009), experimental design is combined with Bayesian techniques to determine optimal designs for population growth models. Similarly, this paper investigates capturing the dynamics of various theoretical ecological growth models by designing optimal sampling regimes. Given the approach is intended to design experiments prior to data collection, simulations are conducted using a well-known ecological model with set parameters that can later be replicated with minimal error.

Using Bayesian inference, the probability of the model parameters are predicted based on simulated data by employing Markov Chain Monte Carlo (MCMC) sampling techniques demonstrated by Gilks et al. (1995). The intent of this study is to learn about dynamical systems in a sequential manner by using various criteria to optimize the system. The Bayesian model combined with design of experiment techniques can be used to determine the optimal

sampling frequency that minimizes costs associated with data collection while obtaining the maximum amount of information from the data. The statistical methods in this paper are applied to various theoretical models and compared across optimality criteria to evaluate the performance. The goal of this research is to develop a new methodology that provides ecologists with optimal times to collect data that accurately captures the population growth dynamics of a system.

2.2 Logistic Growth Model

Logistic curves are used to model growth in various contexts. Originally developed by Verhulst (1838), logistic curves were used to model population growth as an adaptation of exponential growth. Verhulst developed the logistic growth model by adding a multiplicative factor to exponential growth, which stabilizes the model at a particular population size due to competition for resources. The self-limiting growth of the population is referred to as the carrying capacity. The carrying capacity is the maximum population that terminates the growth rate, at which point the system enters a steady-state.

The logistic growth model has been implemented in numerous fields. Logistic growth has been used to estimate parameters and inform intervention strategies during a pandemic involving influenza as demonstrated by Hedge et al. (2013). Furthermore, coal production peaks and trends in China have been modeled with logistic growth by Lin and Liu (2010). Similarly, the logistic curve has taken the form of Hubbert's model (Hubbert et al., 1956) to forecast the production of petroleum. In this paper, the proposed model represents ecological modeling of population growth.

The rediscovery of the logistic equation by Reed and Pearl (1927) is often used in ecology to represent the population dynamic where the rate of reproduction is proportional to both the existing population and the amount of available resources. This model is popular in ecology due to the realistic dynamic of the carrying capacity and existence of the analytic

solution. The equation is written in the mathematical form

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) \quad (2.1)$$

with analytic solution

$$N(t) = \frac{KN_0}{(K - N_0)e^{-rt} + N_0}.$$

The variables N and t represent population size and time respectively. N_0 represents the population at time $t = 0$ referred to as the initial population. Parameter r is the population growth rate, and parameter K is the carrying capacity.

The parameters of logistic growth can be difficult to estimate since the model does not always provide a good fit to true data. Thus, parameter estimation techniques are commonly studied in the literature. Oliver (1964) demonstrates methods for estimating model parameters across a variety of logistic growth models. In more recent times, Bayesian inference has become more accessible to ecologists and used for parameter estimation. Heydari et al. (2014) use Bayesian inference to determine fast estimation methods for logistic growth. In general, ecologists are shifting towards Bayesian modeling of population dynamics based on the increasing resources for ecological Bayesian models (Hobbs and Hooten, 2015). In this paper, logistic growth is specified probabilistically in the Bayesian framework in order to perform a new method for parameter estimation.

2.3 Methods

2.3.1 Statistical Model

Given that ecologists have expertise in modeling logistic growth, the Bayesian approach can be used to incorporate prior knowledge of the system. Thus, the model in equation 2.1 can be specified probabilistically in the Bayesian framework. The Bayesian approach is implemented using Bayes' Theorem (Bayes, 1763). The theorem combines prior beliefs with the likelihood of an experiment to determine a posterior belief. The prior belief of a

parameter θ is noted by $\pi(\theta)$. The likelihood of the experiment is a conditional probability of the data x given the parameter θ noted by $L(x|\theta)$. The posterior belief is given by the continuous case of Bayes' Theorem, also known as the posterior distribution. The posterior distribution defines the conditional probability of a parameter θ given the observance of data x , which is written mathematically as

$$P(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta}.$$

Specifying equation 2.1 as a Bayesian model requires the specification of the likelihood of the experiment and the prior distributions for each parameter. The nature of the logistic equation tracks the population of a species N across a fixed time t where the growth rate parameter r and carrying capacity K are independent fixed values. The population at a specified time $N_i|t_i$ is assumed to have a Poisson likelihood given that the observations N_i are independent counts occurring at known times t_i . The analytic solution to the logistic equation provides the expected value of the likelihood set as $\lambda_i = N(t_i)$. This provides the conditional likelihood of the experiment

$$N_i|t_i \sim \text{Poisson}(\lambda_i|t_i),$$

where λ_i depends on the parameters r and K . Thus, the prior distributions must be specified for the growth rate and carrying capacity parameters.

When modeling logistic growth, ecologists are aware that a population can only reach a maximum capacity by increasing at a positive rate. This implies that the growth rate r and carrying capacity K must be positive values. Though many distributions have positive supports, the log-normal distribution easily specifies numerical information by using the mean and variance. Less informative priors such as the gamma distribution can cause model instability, which defeats the purpose of incorporating prior beliefs. Thus, informative priors

are specified as

$$r \sim \text{Lognormal}(1, 10)$$

$$K \sim \text{Lognormal}(2000, 0.10),$$

where the expected growth rate is centered about one and the carrying capacity is assumed to reach a large quantity. During implementation various priors were explored to test the sensitivity of these set values. As the variance σ^2 of both parameters r and K increased the priors became more informative. Though expert knowledge of the system needs to be incorporated, providing too much information would limit the ability to test the robustness of the proposed methods in this paper. For example when selecting priors $r \sim \text{Lognormal}(0.5, 20)$ and $K \sim \text{Lognormal}(2000, 100)$, the predicted model more closely fit the true model during implementation and did not fully allow the likelihood of the data to inform the decision-making process. On the other hand, decreasing the variance created uninformative priors even with a mean close to the true value. Rather than providing less information, it would be more beneficial in this case to change the distribution and use an uninformative prior such as the uniform distribution. Thus, the above explicit priors were implemented to represent some form of prior knowledge of the system and are used for all simulations in this paper. The Bayesian approach emphasizes combining informative prior beliefs with experimental knowledge to determine a posterior belief. Thus, the specified model will be used to predict the parameters of various theoretical population growth models that can be optimized by assessing various criteria.

2.3.2 Optimal Designs

Optimal designs are used to estimate statistical models while reducing experimental costs. The objective of optimal design is to eliminate uncertainty by minimizing the variability of the parameter estimates. Traditionally, optimal designs (Kiefer, 1959) minimize the variance of the parameter estimates while maximizing the information matrix. Specifically, the Fisher

information matrix is defined as the negative expectation of the second derivative of the log-likelihood function with respect to the parameter θ

$$F(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log L(x|\theta)\right],$$

where the expectation is taken over the sample space of the observations x and parameter space of θ . Optimality criteria are applied to the Fisher information matrix $F(\theta)$ and provide measures of fit to assist with model selection. The popular A and D optimal designs are commonly used in experimentation due to their ability to limit computational expenses. A optimality criterion minimizes the trace of the inverse of the Fisher information matrix, which equivalently minimizes the average variance of the parameter estimates of a model. Whereas, D optimal designs consist of maximizing the determinant of the Fisher information matrix, again minimizing the parameter estimates of the model.

$$A - \text{Optimal: } \min \text{tr}(F^{-1}(\theta))$$

$$D - \text{Optimal: } \max |F(\theta)|$$

Though these designs are commonly used in practice, $F(\theta)$ can be difficult to calculate when the model parameters are unknown. Instead of estimating the Fisher information matrix using parameter estimates noted by $\hat{\theta}$, the variance of the parameter estimates are considered. The inverse of the estimated Fisher information matrix $F(\hat{\theta})$ is an estimator of the asymptotic covariance matrix (Abt and Welch, 1998).

$$\Phi = \text{Var}(\hat{\theta}) = [F(\hat{\theta})]^{-1}$$

The covariance matrix is much easier to calculate than the Fisher information matrix for certain models. Thus, defining $F(\hat{\theta})$ in terms of Φ redefines the optimality criteria in terms of the estimated covariance matrix. I-optimal designs (Atkinson et al., 2007) are known for minimizing the average prediction variance over the entire design region and are also

considered.

$$\begin{aligned}
 A_\Phi - \text{Optimal: } & \min \text{tr}(\Phi) \\
 D_\Phi - \text{Optimal: } & \min |\Phi| \\
 \text{I - Optimal : } & \min \bar{\Phi}_{pred}
 \end{aligned}
 \tag{2.2}$$

$\bar{\Phi}_{pred}$ notes the average prediction variance of the model. Rather than minimizing the trace of the inverse of the Fisher information matrix, A_Φ -optimal designs minimize the trace of the covariance matrix. D_Φ -optimal designs minimize the determinant of the covariance matrix. While, I-optimal designs minimize the average prediction variance over the design space. Again, optimality criteria are measures of fit used to guide optimization processes. A_Φ , D_Φ and I optimality criteria are used in the proposed sequential optimality algorithm presented in the next section.

2.3.3 Sequential Optimality

In practice, design techniques are used prior to collecting data in order to optimize the amount of information obtained during experimentation. For instance, simulated annealing applied by Laarhoven and Aarts (1987) is a commonly used probabilistic technique that searches the design region for an approximation of the global optimum of a given function. Though easy to implement, simulated annealing can be computationally expensive and does not learn about a system in a sequential manner. We compare simulated annealing to our proposed adaptation of the method that predicts the optimal future design point based on the current information of the system. We prefer this Bayesian approach to incorporate prior knowledge into the decision-making process. The proposed algorithm sequentially searches subsets of the design space to determine the best future time for ecologists to collect data. Sequential optimality is written below as Algorithm 4.

Algorithm 4: Sequential Optimality

Begin

Choose an initial design t_1, \dots, t_n

Set a design budget, b

Set a design window, w

Set criteria, C :

(i.) $A_\Phi = \min \text{tr}(\Phi)$

(ii.) $D_\Phi = \min |\Phi|$

(iii.) $I = \min \bar{\Phi}_{pred}$

For $D = t_1, \dots, t_n$:

(a) Draw a sample $t^* = \{t_{n+1}, \dots, t_{n+w}\}$

(b) Accept the new state $t_{new} = \arg C(t^*)$

Repeat until $D = t_1, \dots, t_b$

End

The algorithm begins by setting an initial design of size n typically set as the number of parameters in the model plus one. Then, a design point budget b is chosen according to the number of runs the experimenter can afford. Once the initial data is collected and a design budget is set, the design window w can be set. The design window can be established by dividing the planned time interval by the design point budget. The window of points following the current design D are explored as candidate samples. Each candidate is evaluated by running the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953) guided by specified optimality criterion, C . The optimal point within the window t_{new} is added to the design, and the process repeats until the design point budget is exhausted. These steps are demonstrated and explained further through a worked example.

Worked Example

This worked example uses Algorithm 4 to find an optimal design for a normal 10% logistic growth model using A-optimality criterion. Each step is provided in detail as follows.

Step 1: Choose an initial design t_1, \dots, t_n

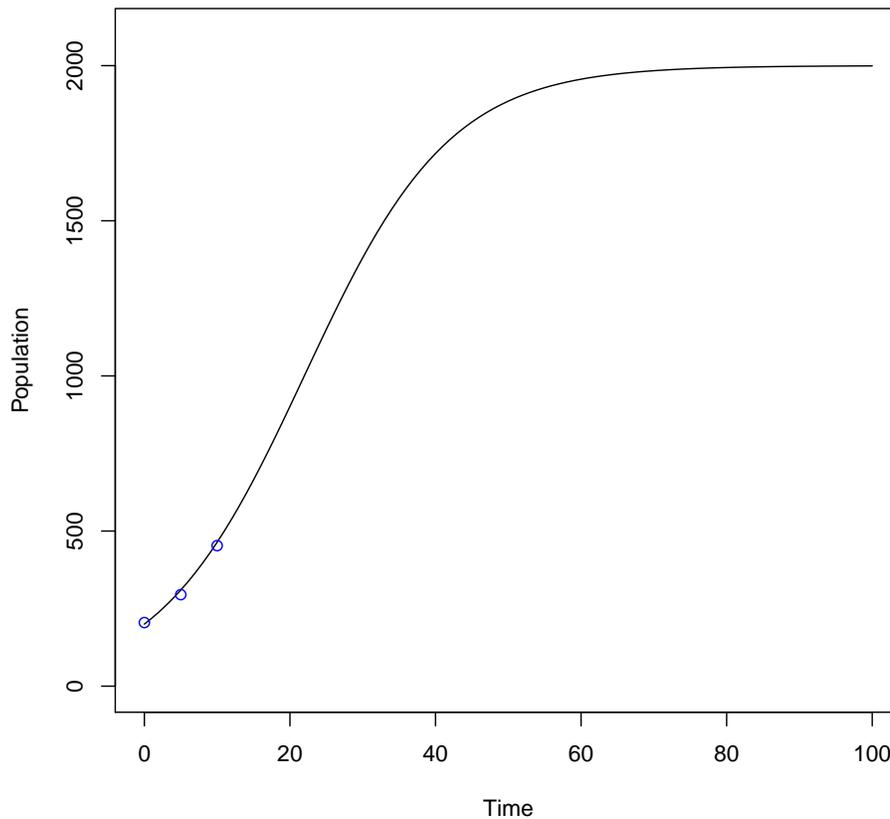


Fig. 1.: An initial design of three points is plotted in blue. The ground truth model of normal 10% growth is plotted as a black curve across one hundred time points with a carrying capacity of two thousand and initial population of two hundred.

Logistic growth has two parameters, growth rate and carrying capacity. Therefore, three initial design points, t_1, t_2 and t_3 , are selected as shown in Figure 1 representing samples collected on days 0, 5 and 10 of the season. The population size is calculated at these time points using the analytic solution with random Poisson noise. In this example the coordinates are $(0, 205)$, $(5, 295)$ and $(10, 453)$.

Step 2: Set a design budget, b

The design budget is chosen at the discretion of the expert based on available resources. In this example, we select a budget of ten, $b = 10$.

Step 3: Set a design window, w

This study focuses on sampling across a one hundred day season with a design budget of size ten. A practical design window would be to search ten days into the future. Thus, we set $w = 10$.

Step 4: Set criteria, C

For demonstration purposes, A-optimal criterion is selected as $C = A_{\Phi} = \min tr(\Phi)$.

Step 5: Draw a sample $t^* = \{t_{n+1}, \dots, t_{n+w}\}$

The initial design consists of three points. Therefore, t^* represents a window of the ten consecutive points following the last point sampled from $t_4, \dots, t_{13} = \text{Day 11}, \dots, \text{Day 20}$.

Step 6: Accept the new state $t_{new} = \arg C(t^*)$

Here, we calculate the trace of the covariance matrix for each point in $t^* = t_4, \dots, t_{13}$, which comes from the MH algorithm producing estimates for the carrying capacity and growth rate.

$$tr(\Phi(t^*)) = (1.09 \times 10^{14}, 1.16 \times 10^{14}, 1.94 \times 10^{14}, 8.75 \times 10^{13}, 1.01 \times 10^{14}, \\ 7.45 \times 10^{13}, 6.32 \times 10^{14}, 2.37 \times 10^{14}, 1.40 \times 10^{14}, 1.39 \times 10^{14})$$

$$\min tr(\Phi(t^*)) = (7.45 \times 10^{13})$$

$$\arg \min tr(\Phi(t^*)) = 16$$

Based on the calculations, the optimal point selected in the design window is Day 16, which leads to the selected design point (16, 661) calculated by the analytic solution with random

Poisson noise. Figure 2 illustrates the candidate sample region in red and the selected point among the candidates is shown in Figure 3.

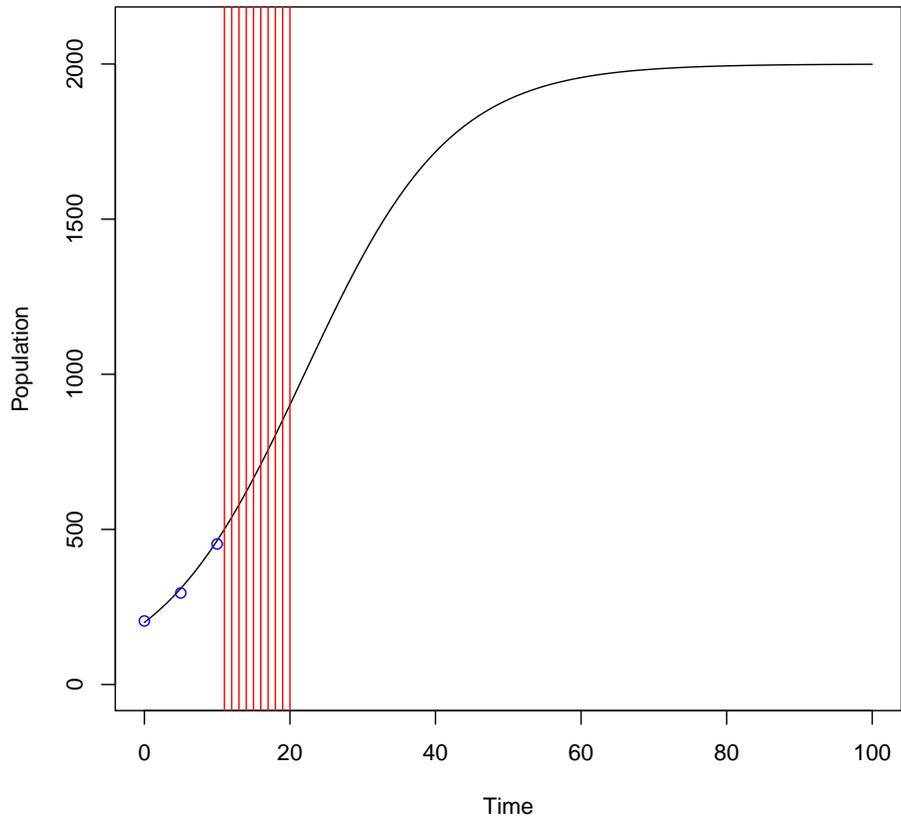


Fig. 2.: A window of ten candidate points are selected from the last point sampled illustrated by the red region.

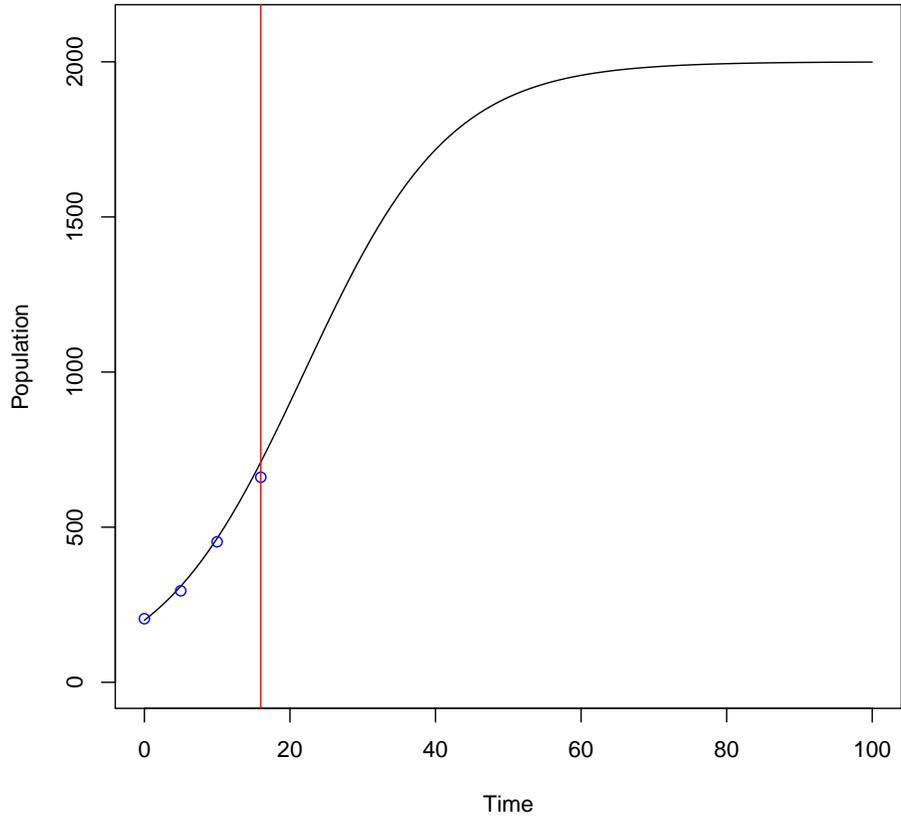
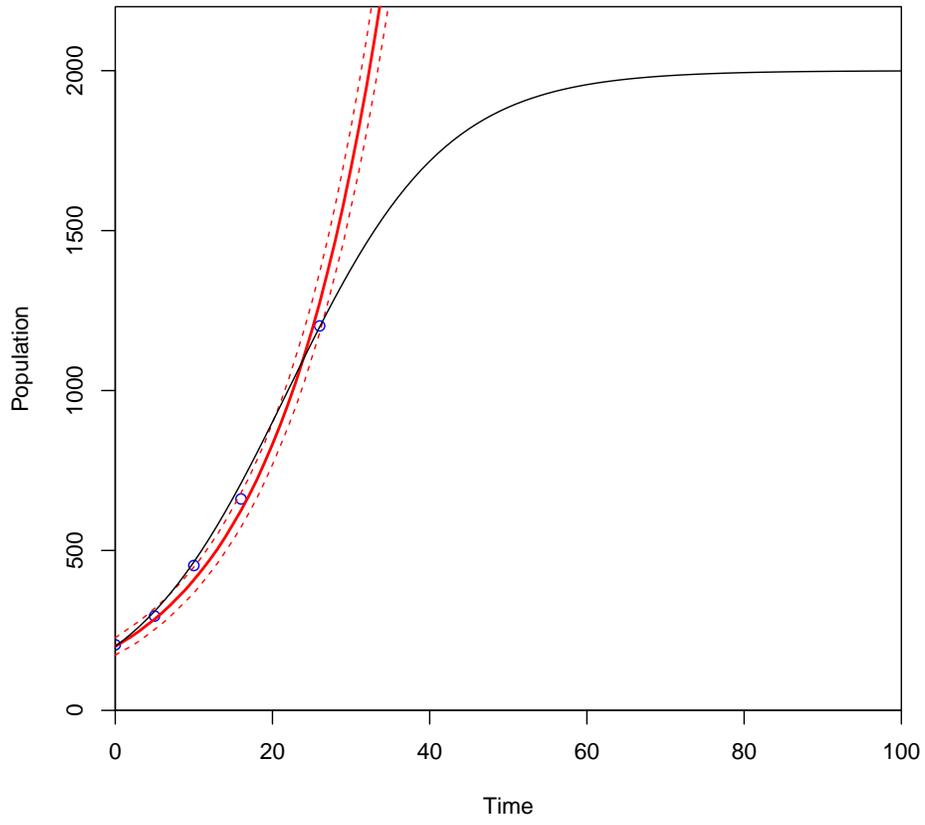


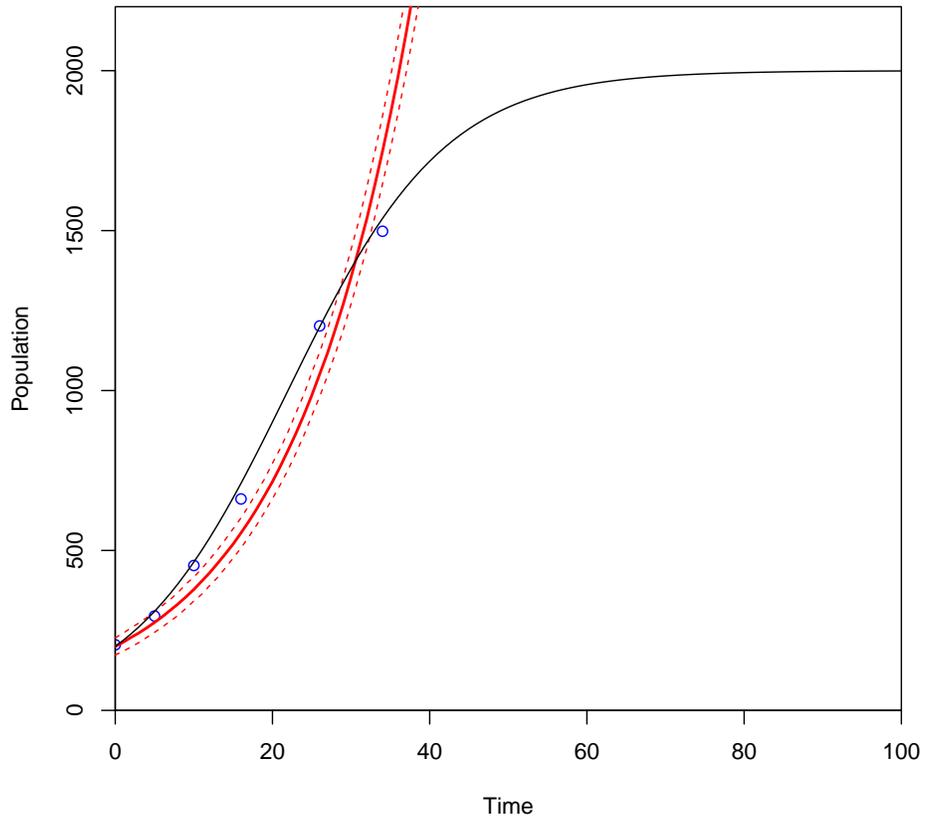
Fig. 3.: Each point is evaluated using A optimality criterion in the window from Figure 2 and the argument with the minimum trace in the window is selected.

Step 7: Repeat until $D = t_1, \dots, t_b$

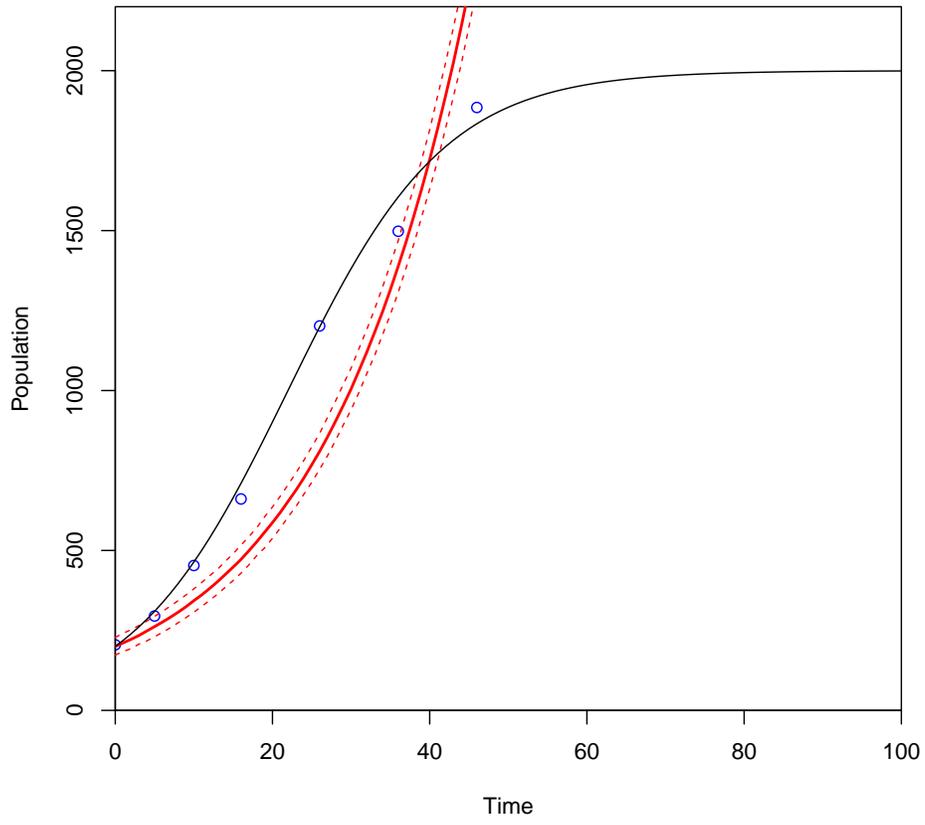
Steps 5 and 6 are repeated until all ten points in the budget are exhausted. The parameter estimates are updated as each new design point is added as shown in Figure 4. The final design consists of ten points calculated by the analytic solution as follows: (0, 205), (5, 295), (10, 453), (16, 661), (26, 1202), (36, 1498), (46, 1885), (56, 1927), (64, 2019) and (68, 1961).



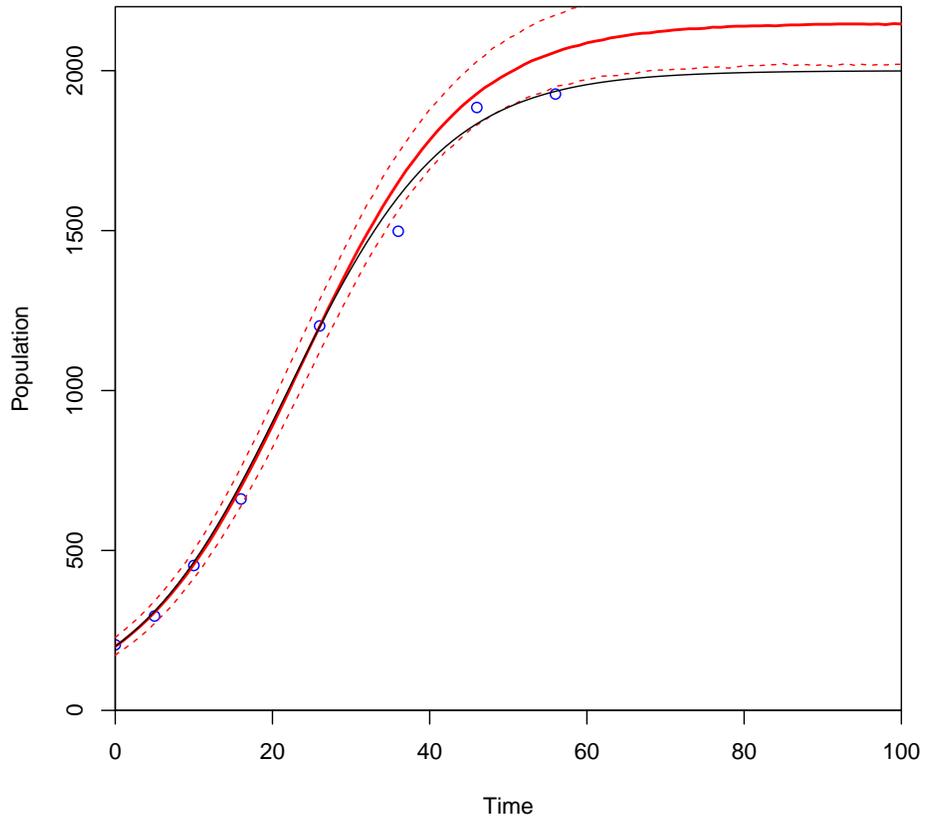
(a)



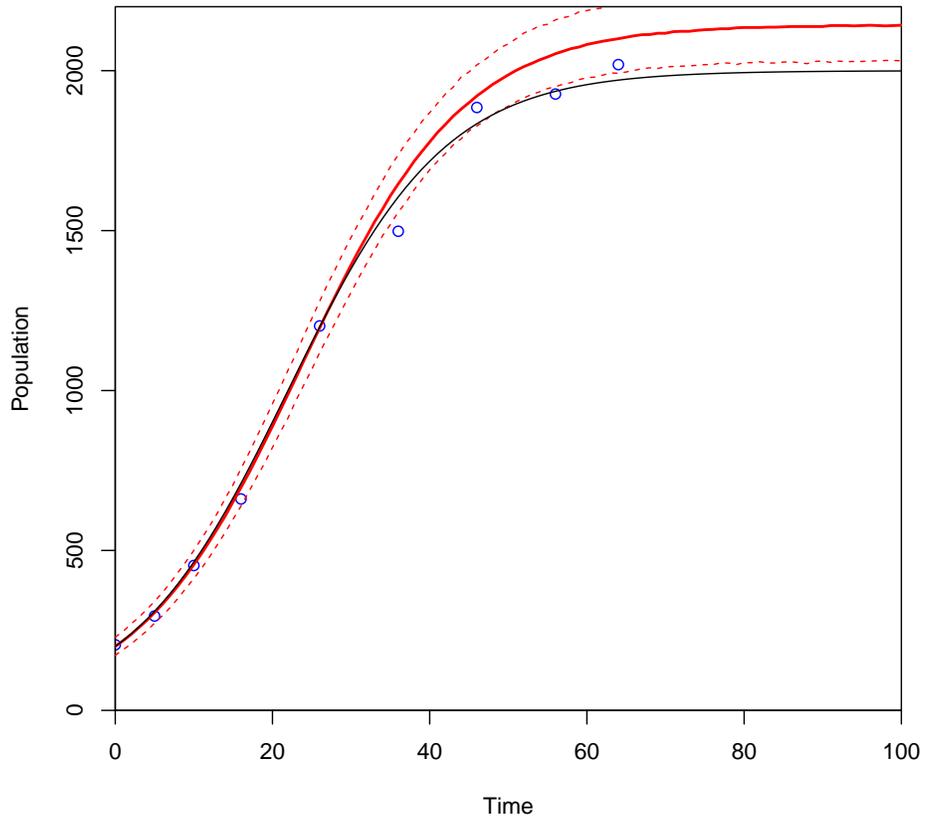
(b)



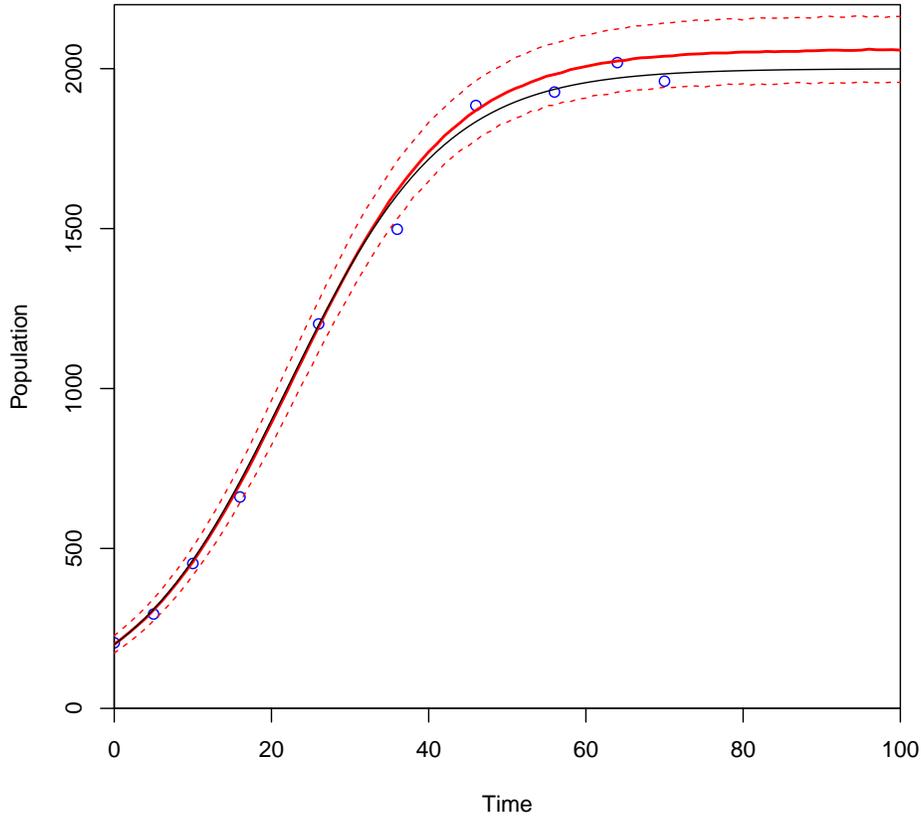
(c)



(d)



(e)



(f)

Fig. 4.: The fifth design point is selected in panel (a) found in a window following the design from Figure 2: panel (b). Panels (b)-(f) represent each additional design point added until the budget of ten points is exhausted. Each panel plots the ground truth model in black, the design points in blue and fits the design with a red curve based on the parameter estimates. The red dotted lines represent the 2.5% and 97.5% prediction quantiles.

The sequential optimality algorithm is intended to search the design space of temporal models. The above example demonstrates the procedure learning about a temporal system in a sequential manner. The proposed method will be implemented later across various logistic growth models using all design criteria to compare the technique in reference to simulated an-

nealing. The purpose of this design study is to develop and demonstrate statistical methods that can optimize sampling procedures for ecologists.

2.4 Simulation Study

This simulation study demonstrates methods that can be used to design sampling schedules for ecologists applied to three theoretical scenarios of logistic growth provided by Table 1. All simulations are programmed in R (R Core Team, 2013). Data is simulated for normal, fast and slow logistic growth models across one hundred time points. The normal growth rate is set to 10%, the fast rate is set to 100%, and the slow rate is set to 5% growth. All models have a maximum carrying capacity of two thousand, and ten thousand MCMC samples are used to predict the probability of the model parameters. Based on the planned time interval of one hundred days, all optimal designs will consist of ten points that search windows of size ten. This allows the entire budget to be used should the farthest point in each window be selected. Simulated annealing is implemented first to demonstrate the global optimum for each scenario as a reference when comparing sequential optimality. Therefore, ten design points are also selected during simulated annealing for comparison purposes. Then, sequential optimality is demonstrated as a novel approach to designing optimal sampling regimes.

Label	Growth Rate	Carrying Capacity	Initial Population
Normal	0.10	2000	200
Fast	1.00	2000	200
Slow	0.05	2000	200

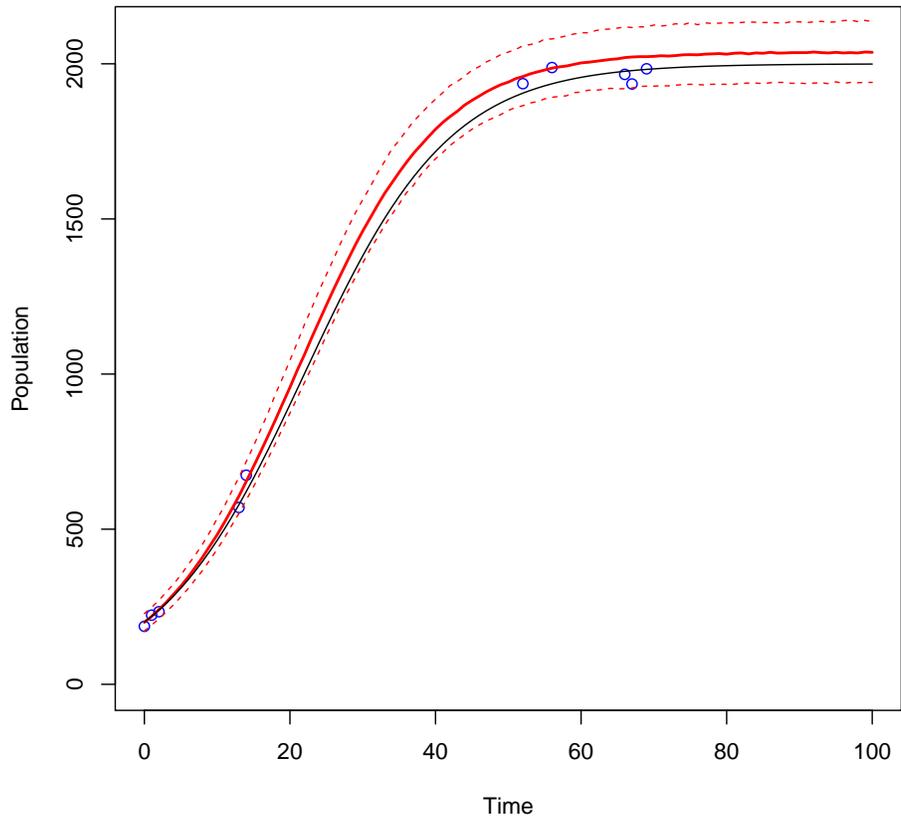
Table 1.: Logistic Growth Scenarios

All ground truth models are generated using the logistic equation and analytic solution provided by Equation 2.1. Each simulation plots the ground truth model as a black curve

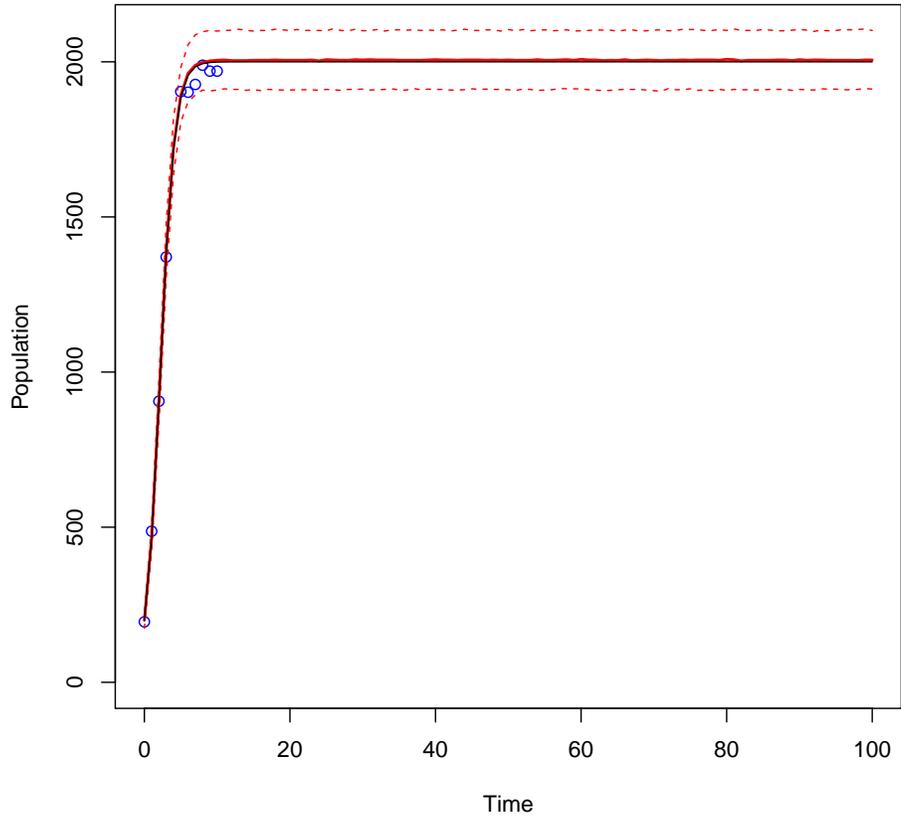
tracking the population size across time. 10,000 MCMC samples are used for each simulation, and the optimal design points selected are plotted as blue open circles. The designs are fit by a red curve with the prediction interval plotted as red dotted lines calculated by the 2.5% and 97.5% quantiles of the predicted values. It will become clear by the results that there are many ways to successfully design experiments for logistic growth models.

2.4.1 Simulated Annealing

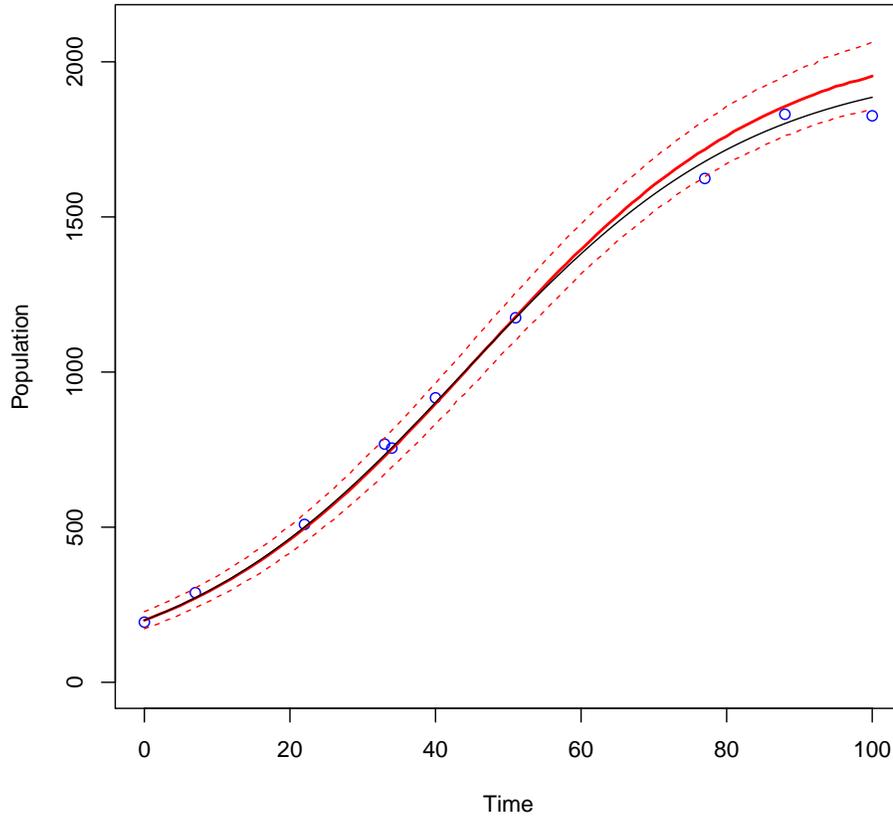
Simulated annealing is implemented as an adaptation of the MH algorithm to explore the design space for an optimal solution. Considering normal, fast and slow growth rates, the specified optimality criteria are used to guide the algorithm and find global optimal solutions. As stated previously, the models simulate growth at 10%, 100%, and 5% rates. Each criterion guides the simulated annealing process to produce optimal designs for each model.



(a)

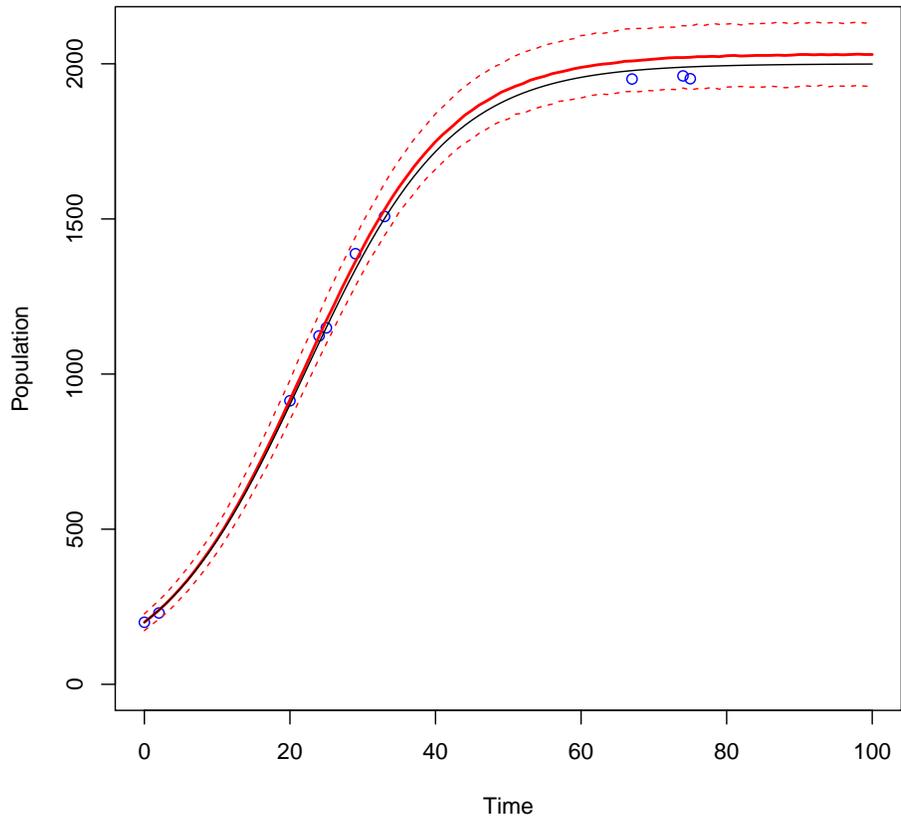


(b)

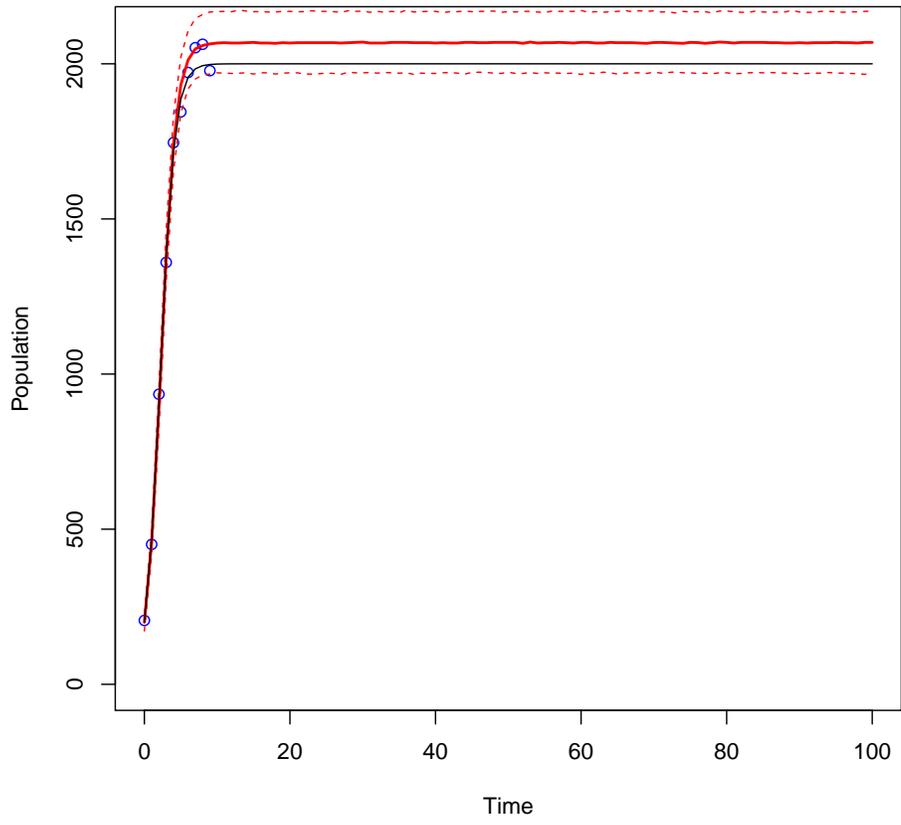


(c)

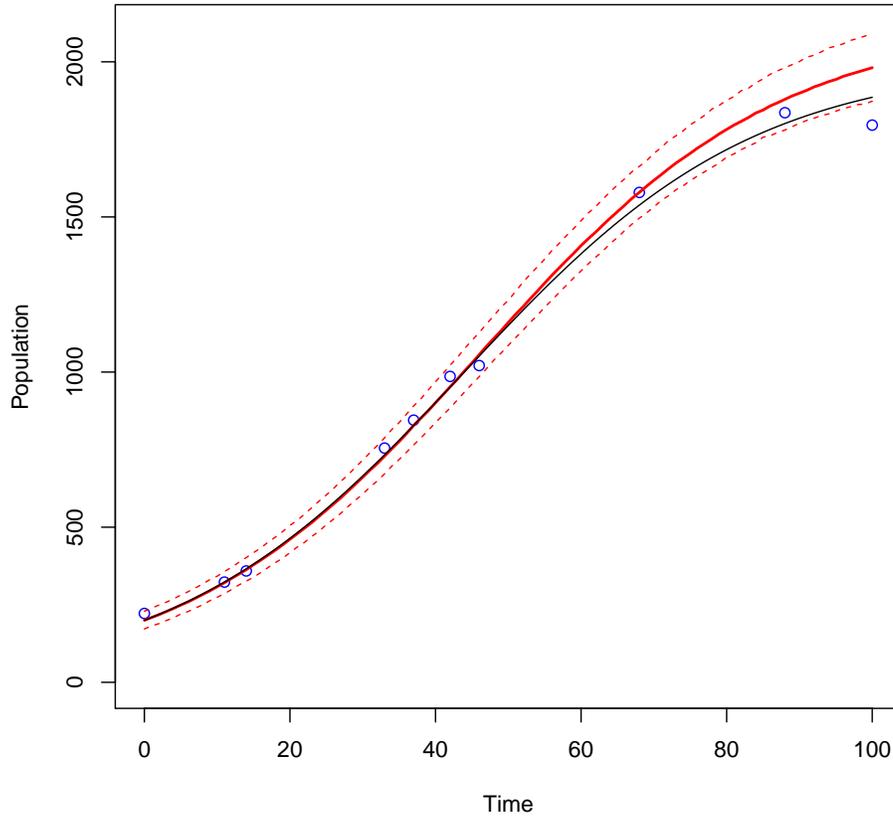
Fig. 5.: 10,000 MCMC samples are used to simulate A_Φ - optimal designs for (a) normal, (b) fast and (c) slow growth models. All models are simulated across 100 time points with a carrying capacity of 2000. The ground truth models are plotted in black. The optimal design points are plotted in blue and fit by a solid red curve, where the 2.5% and 97.5% quantiles of the predicted values are plotted as red dotted lines.



(a)

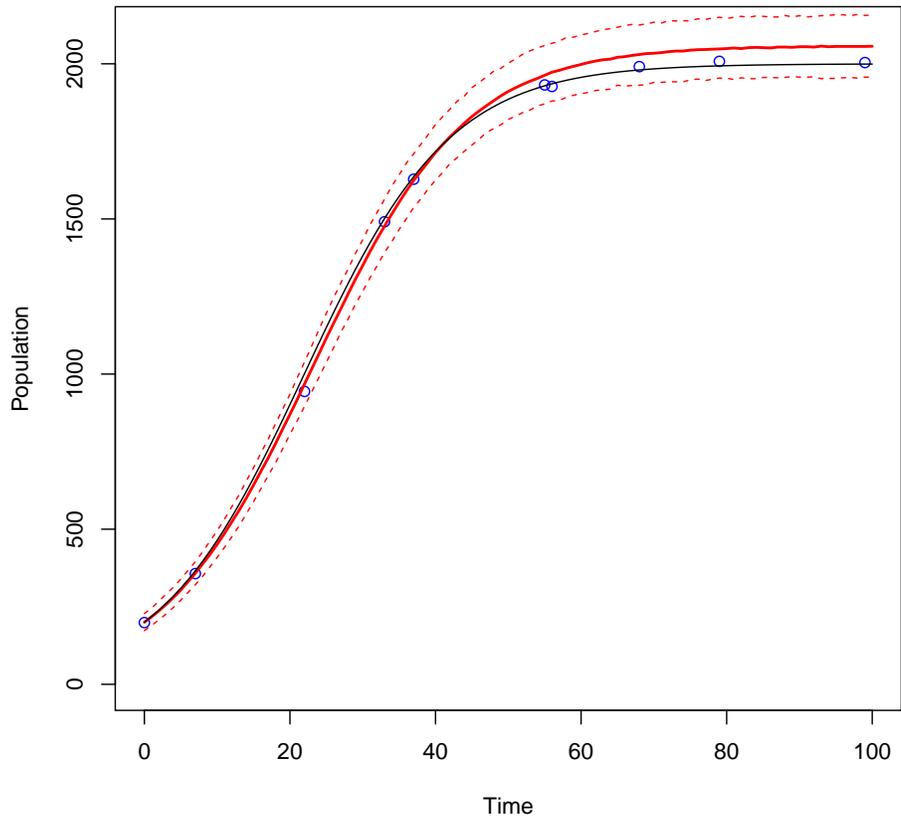


(b)

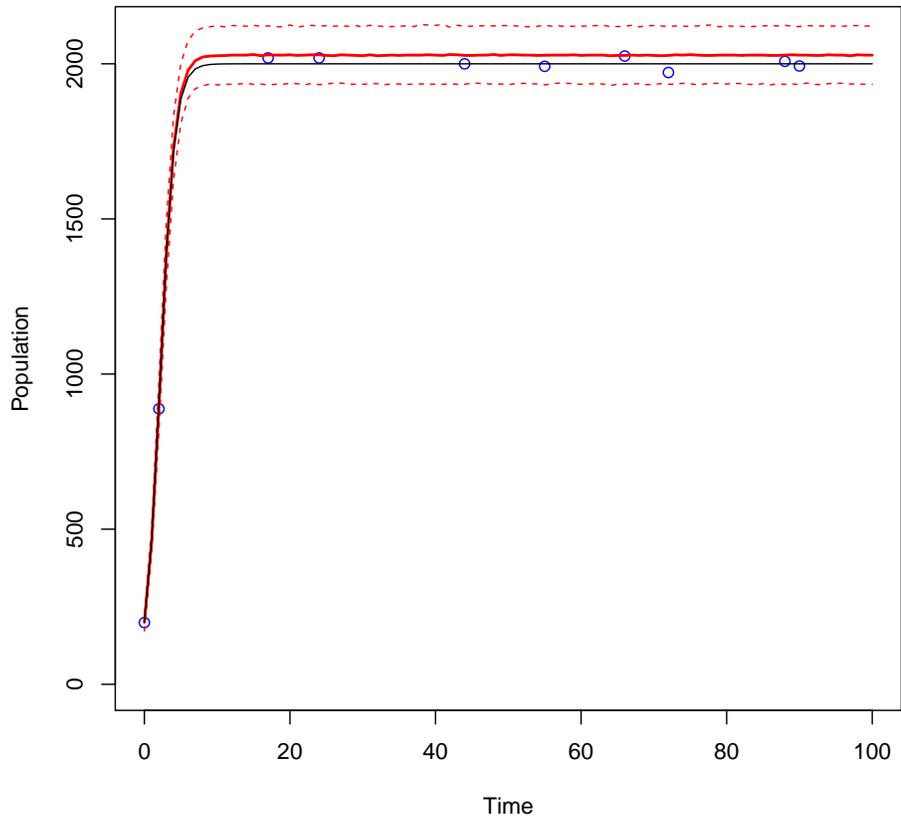


(c)

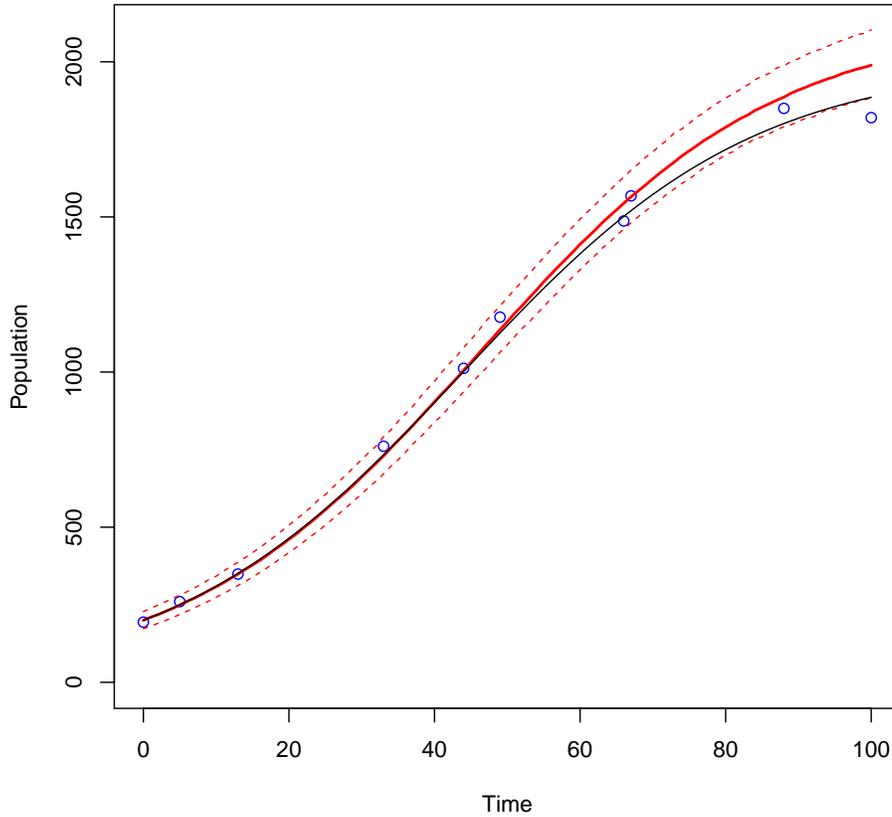
Fig. 6.: 10,000 MCMC samples are used to simulate D_Φ - optimal designs for (a) normal, (b) fast and (c) slow growth models. All models are simulated across 100 time points with a carrying capacity of 2000. The ground truth models are plotted in black. The optimal design points are plotted in blue and fit by a solid red curve, where the 2.5% and 97.5% quantiles of the predicted values are plotted as red dotted lines.



(a)



(b)



(c)

Fig. 7.: 10,000 MCMC samples are used to simulate I-optimal designs for (a) normal, (b) fast and (c) slow growth models. All models are simulated across 100 time points with a carrying capacity of 2000. The ground truth models are plotted in black. The optimal design points are plotted in blue and fit by a solid red curve, where the 2.5% and 97.5% quantiles of the predicted values are plotted as red dotted lines.

Simulated annealing successfully captures the dynamics of all growth models using each criterion. Figure 5 demonstrates the designs generated using A_Φ optimality criterion. Figure 6 fits the D_Φ -optimal designs. Figure 7 illustrates the fit of I-optimal designs. In the next section, simulation studies are conducted using sequential optimality to demonstrate

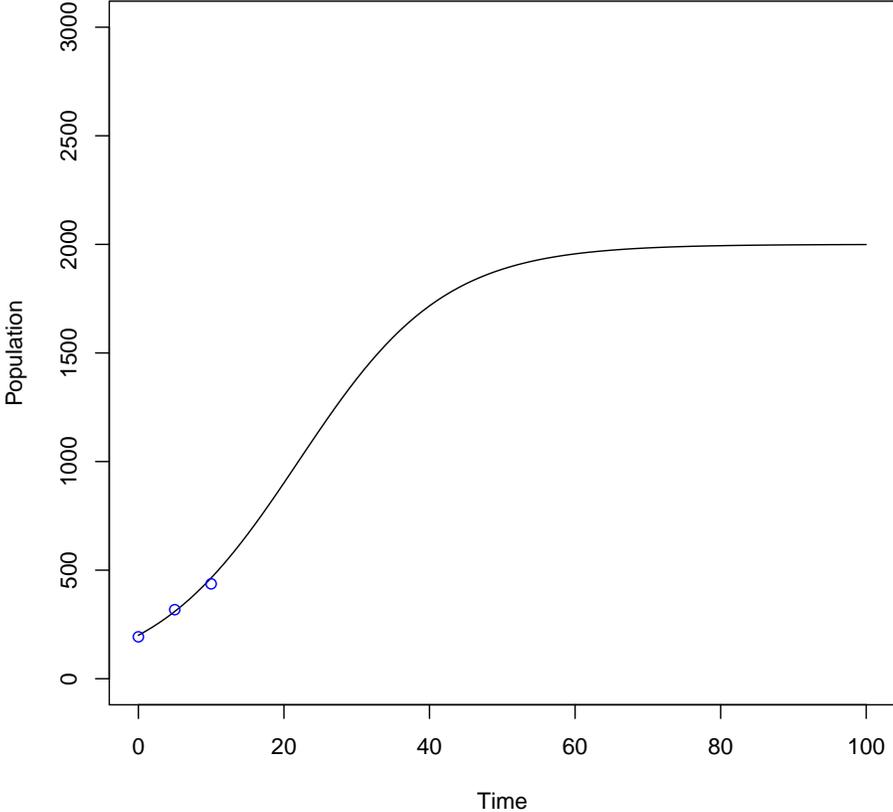
a new method for prediction based optimization. Sequential optimality is an adaptation of simulated annealing that predicts optimal design points.

2.4.2 Sequential Design

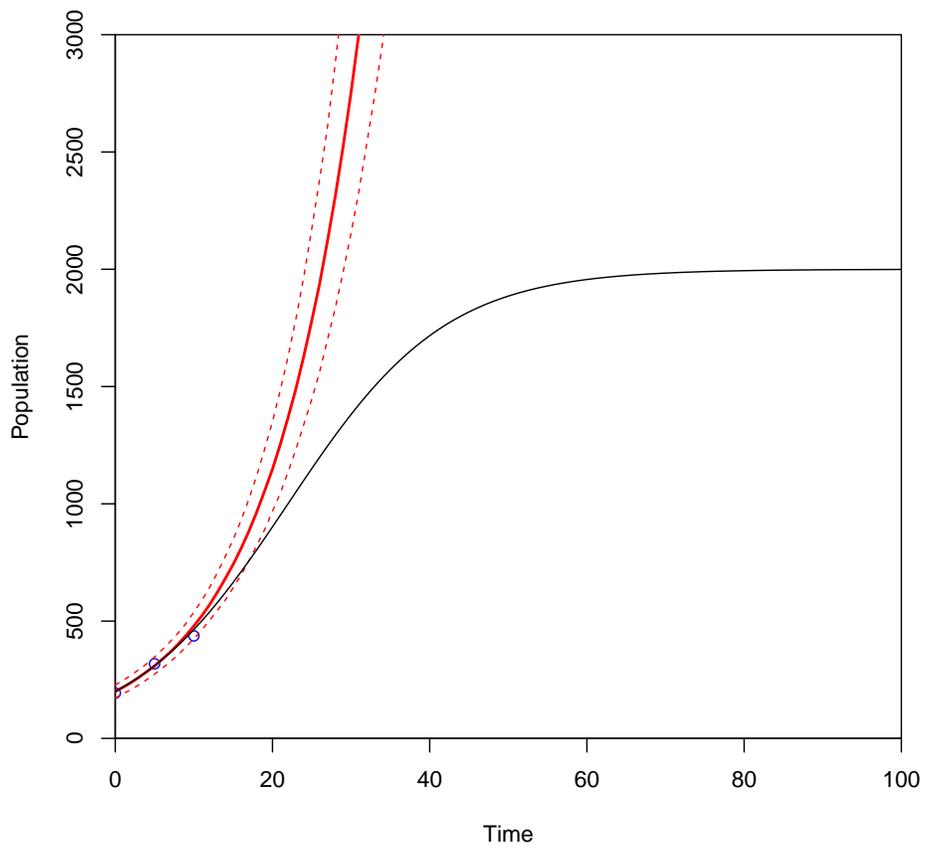
Sequential optimality searches subsets of the design space to find the next best point. Normal, fast and slow growth at 10%, 100% and 5% rates respectively are simulated. All simulations use 10,000 MCMC samples plotted across one hundred time points. The carrying capacity is two thousand. The design point budget consists of ten points, and the algorithm uses the optimality criteria. As each criterion guides the sequential optimality process, the final optimal designs are compared for each model. Again, the ground truth model is plotted as a black curve. The optimal design points are plotted as blue points fit by a solid red curve with the 2.5% and 97.5% quantiles of the predicted values plotted as red dotted lines. As the system updates, our goal is to fit the ground truth model as accurately as possible with the final optimal design. I, A_{Φ} , and D_{Φ} optimal designs are compared across normal, fast, and slow growth rates.

The simulations begin with the I-optimality criterion to demonstrate prediction based designs. Figure 8 demonstrates sequential optimality using the prediction based I-optimality criterion to search a normal growth model. Given the model specification and priors, there is expert knowledge informing the growth rate parameter but limited knowledge regarding the carrying capacity. This simulation specifically demonstrates the ability of the algorithm to explore the uncertainty of the carrying capacity and capture the dynamic. An initial design is set of three points given the logistic equation has two parameters, carrying capacity and growth rate. One hundred day seasons are modeled with a design point budget of ten. The ten candidate points following the base design are evaluated using the I-optimality criterion to select the point with the minimum prediction variance. Once the optimal point is added to the design, the process repeats until all ten design points are sampled. Notice that the change from step 6 in panel (e) to step 7 in panel (f) of Figure 8 demonstrates the method

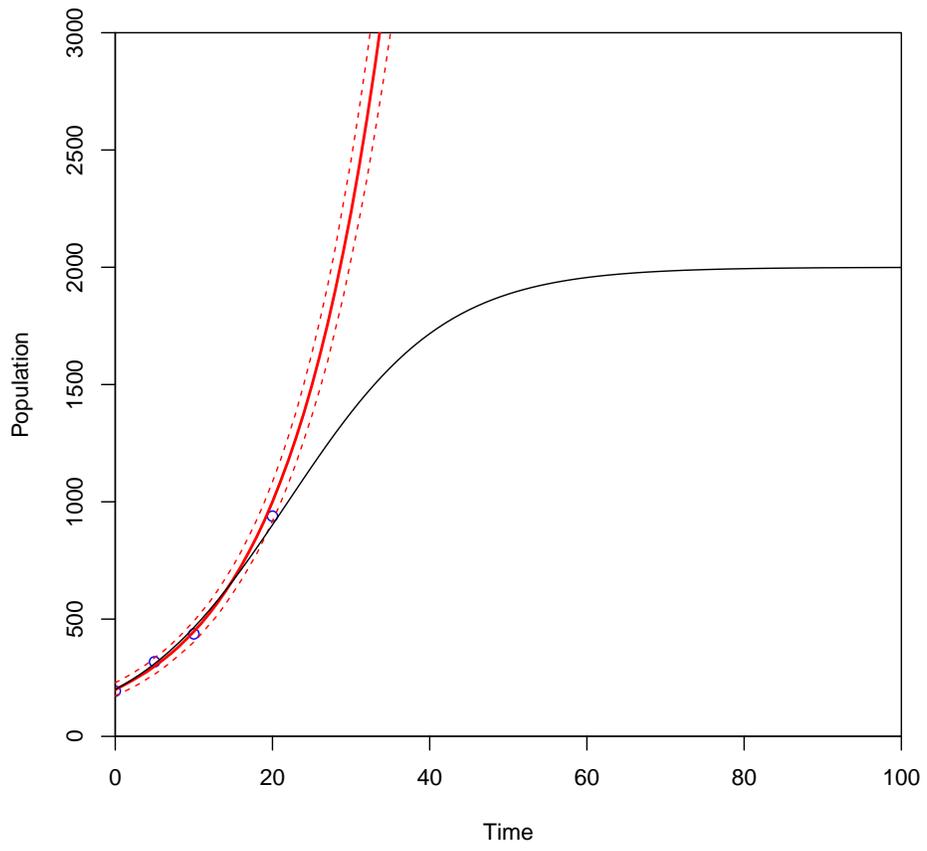
searching uncertainty of the plateau level of the curve. As the process updates, the width of the decision boundary decreases showing more certainty around the steady state of the system. Based on the simulation, the prediction based I optimality criterion is successful in capturing the dynamics of a normal growth model. However, different combinations of criteria and growth models are simulated to test the robustness of the technique.



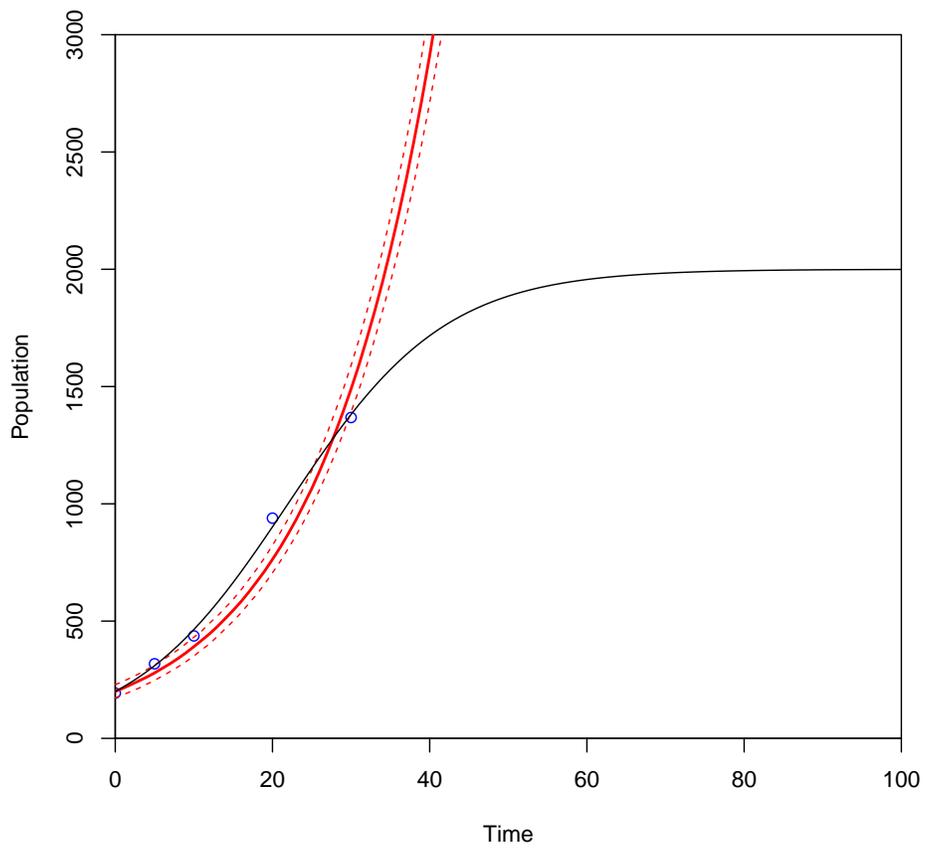
(a)



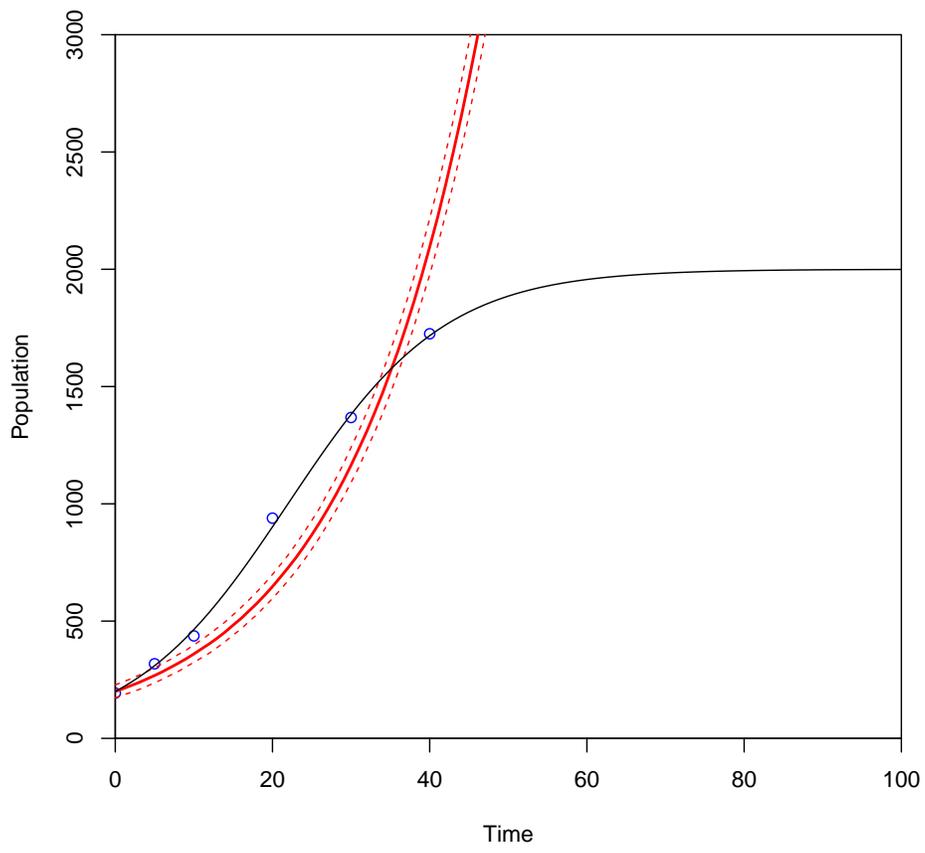
(b)



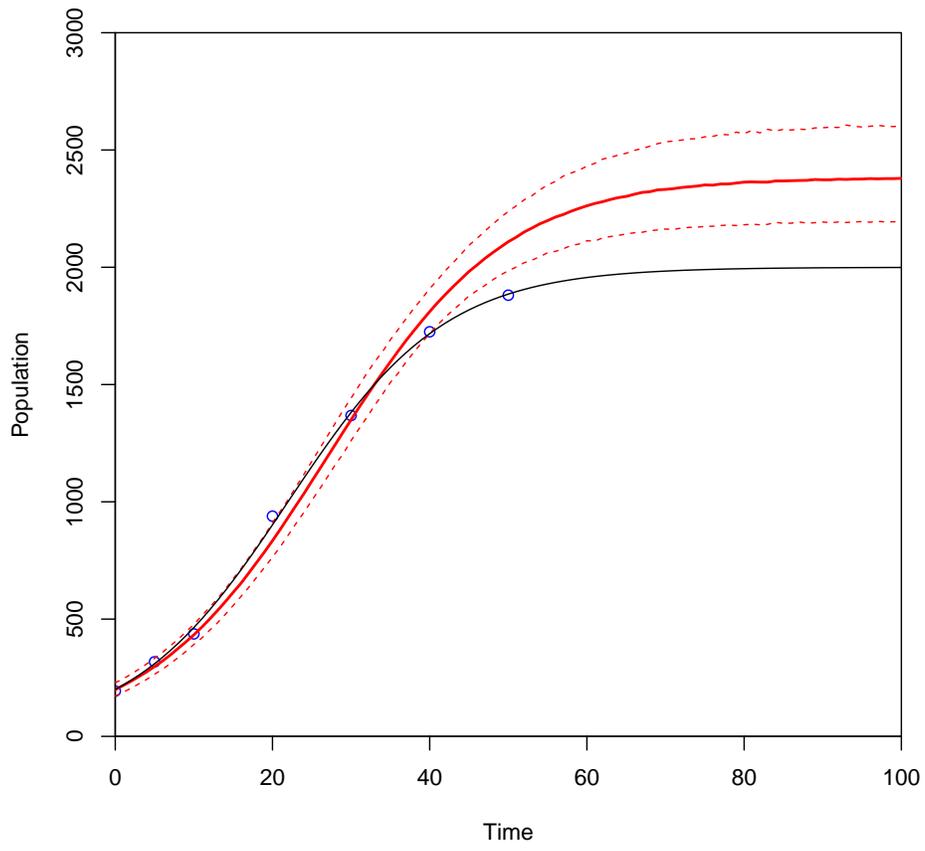
(c)



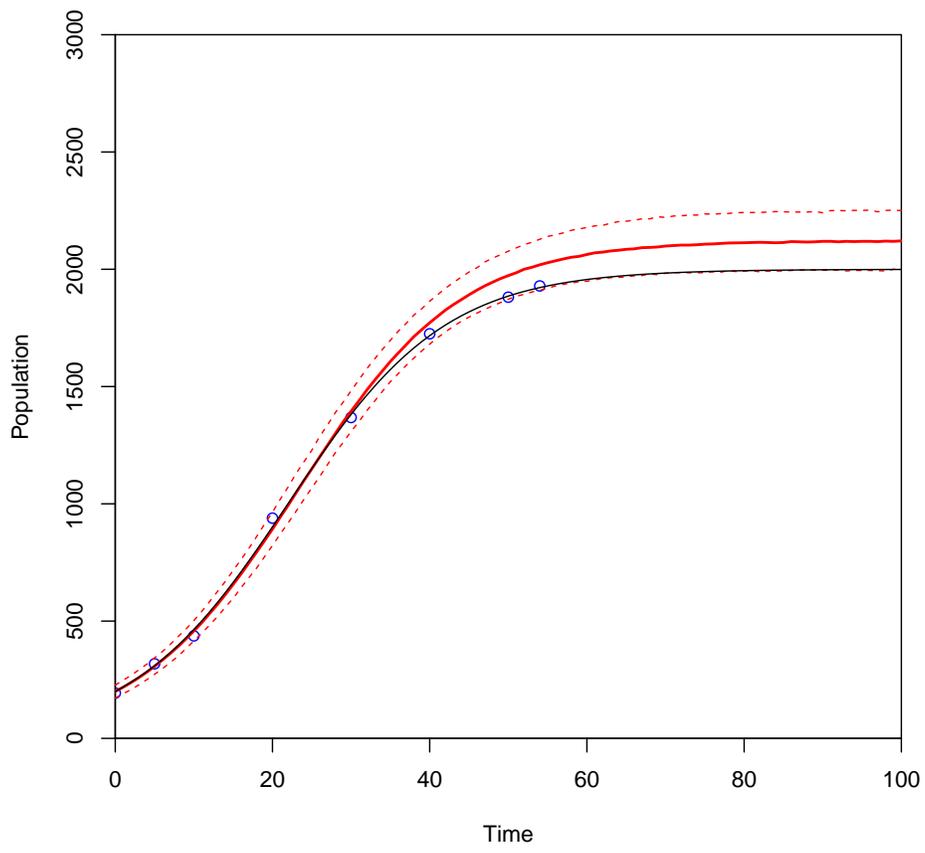
(d)



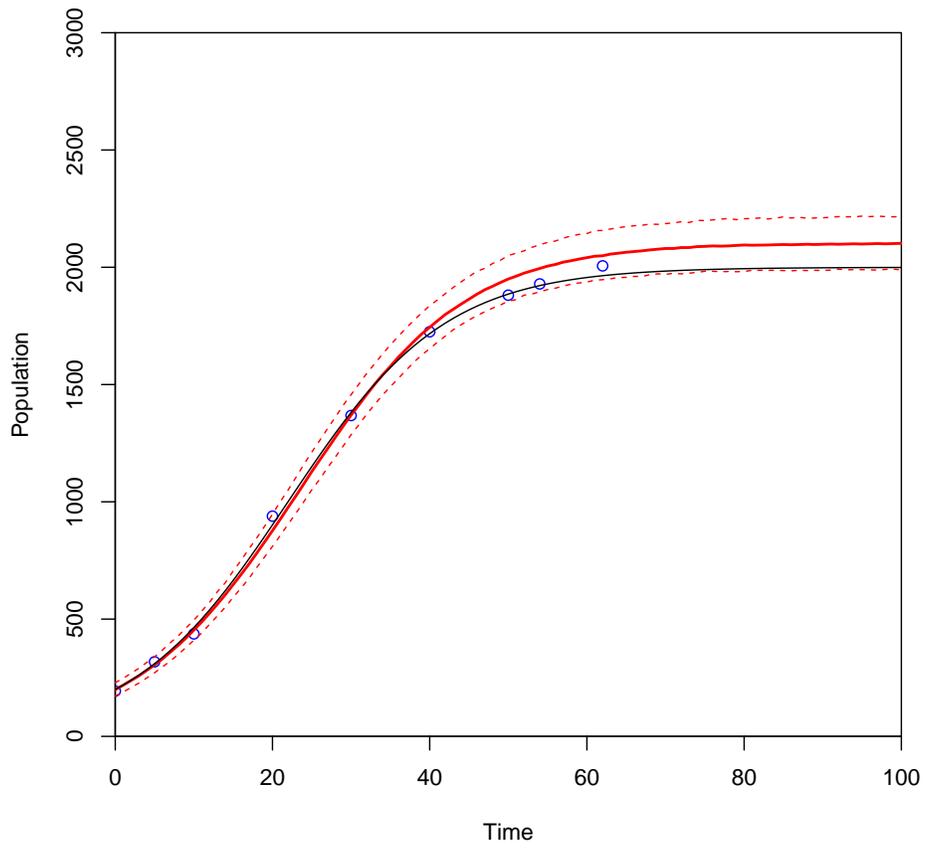
(e)



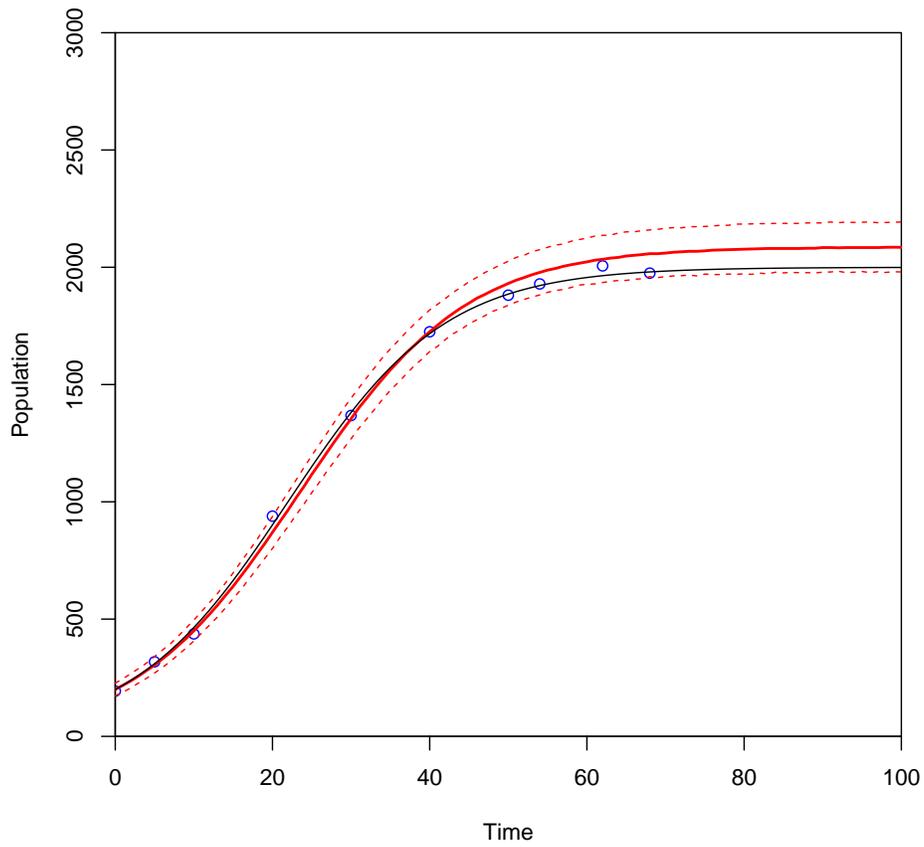
(f)



(g)

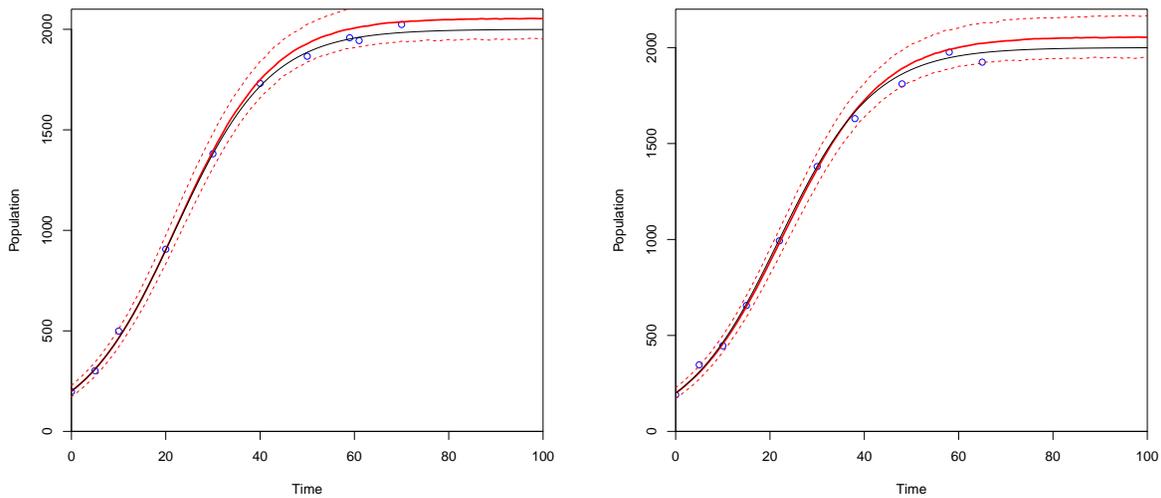


(h)



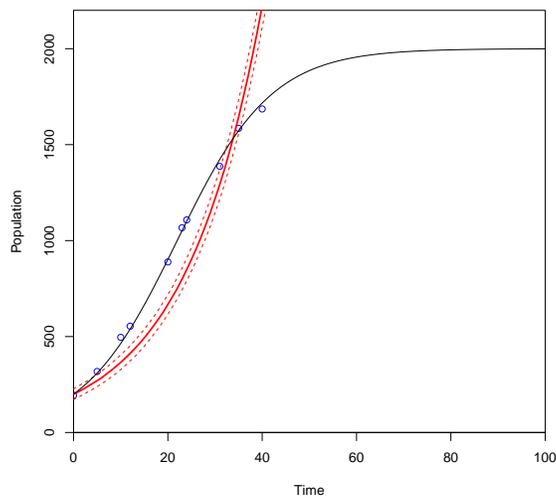
(i)

Fig. 8.: The ground truth normal growth model is plotted as a black curve. I-optimality criterion guides the sequential optimality process. The design points are plotted in blue and are fit by a solid red curve. The red dotted lines represent the 2.5% and 97.5% predicted quantiles. Panel (a) plots the base design, panel (b) fits the design, and panels (c)-(i) plot the optimal points as they are fit.



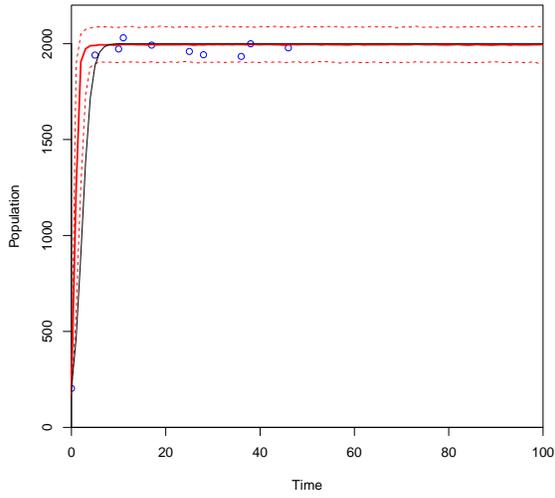
(a)

(b)

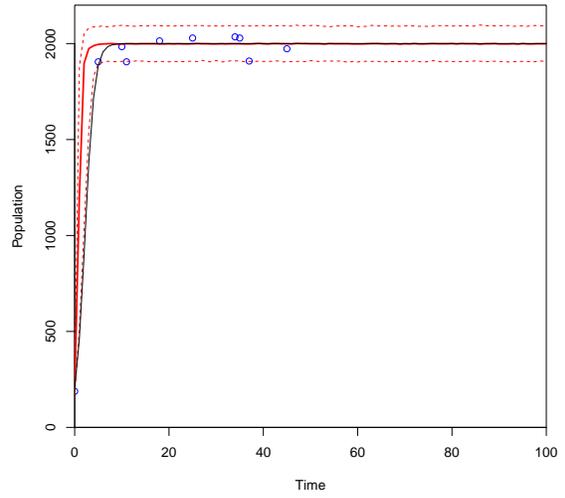


(c)

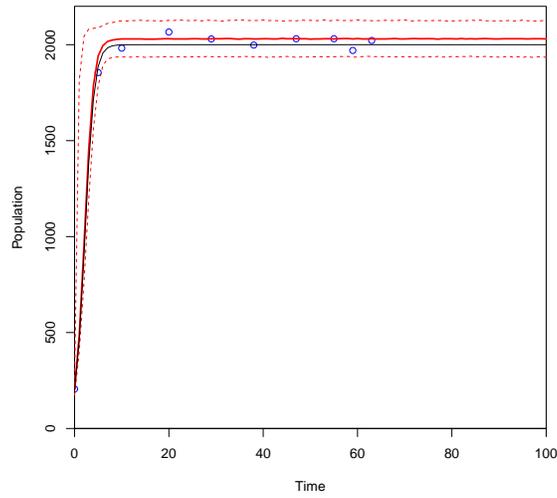
Fig. 9: 10,000 MCMC samples are used in this simulation of a normal growth curve increasing at a 10% rate across 100 time points with a carrying capacity of 2000. Sequential optimality is used to find the ten best points for (a) I , (b) A_Φ and (c) D_Φ optimal designs. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.



(a)

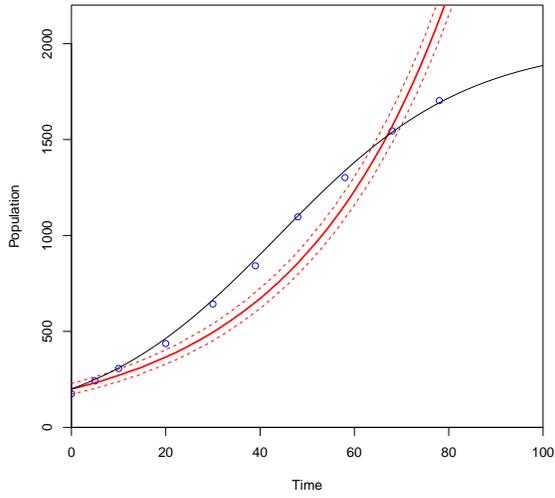


(b)

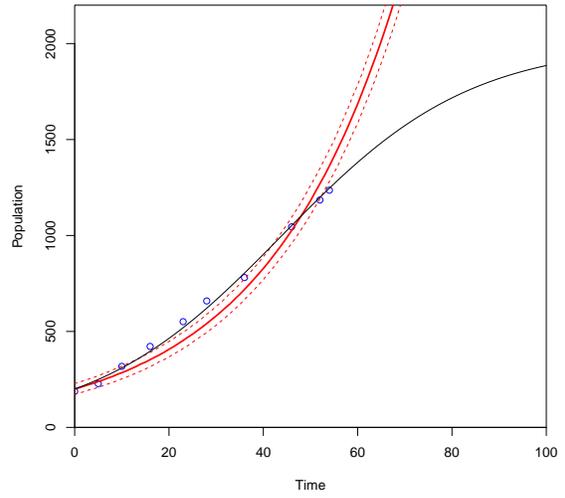


(c)

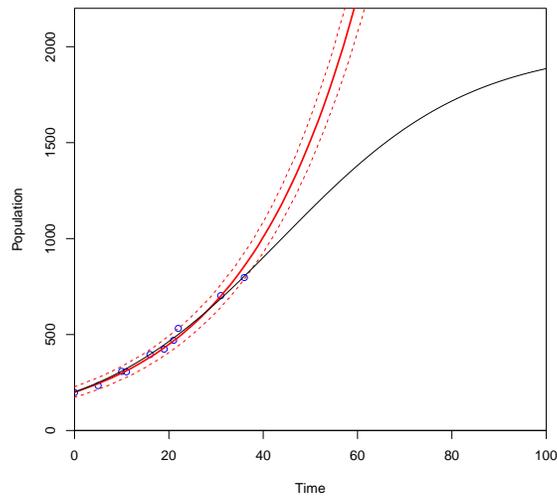
Fig. 10.: 10,000 MCMC samples are used in this simulation of a rapid growth curve increasing at a 100% rate across 100 time points with a carrying capacity of 2000. Sequential optimality is used to find the ten best points for (a) I , (b) A_Φ and (c) D_Φ optimal designs. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.



(a)



(b)



(c)

Fig. 11.: 10,000 MCMC samples are used in this simulation of a slow growth curve increasing at a 5% rate across 100 time points with a carrying capacity of 2000. Sequential optimality is used to find the ten best points for (a) I , (b) A_{Φ} and (c) D_{Φ} optimal designs. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.

The simulations of normal growth show varying results in Figure 9 across the optimality criteria. The I and A_Φ optimal designs successfully capture the dynamics of the population with ten points. Whereas, the D_Φ optimal design did not capture the dynamic of the carrying capacity. This indicates the need for a larger design point budget or smaller design window when using D_Φ optimality criterion with a normal growth curve. On the other hand, all three optimality criteria capture the dynamics of the fast growth rate in Figure 10, which is expected given that the model plateaus rapidly. This implies that the design point budget could be decreased in this case. As for the slow growth rate model in Figure 11, no criteria could capture the population dynamics. This could indicate a need to increase the design point budget or decrease the design window for slow growth models. The size of the design point budget was set to ten points with a window of ten points to simply demonstrate this novel approach of sequential optimality.

Bayesian optimal designs (Chaloner and Verdinelli, 1995) are also examined for comparison purposes. Bayesian optimal designs can be written in the form of a utility function $U(d_{x_f}^*)$, where $d_{x_f}^*$ represents a design chosen from the design region d_{x_f} . The design region is explored using Bayes I, D, and A optimal designs written mathematically as follows.

Bayesian I-Optimal

$$\bar{U}_I(d_{x_f}^*) = \min \int_{\Theta} \min_{d_{x_f}} (Y_2 - E[Y_2|y_1, d_{x_f}])^2 p(Y_2|y_1, d_{x_f}) dY_2$$

Bayesian D-Optimal

$$\bar{U}_D(d_{x_f}^*) = \min \int_{\Theta} \min_{d_{x_f}} |Cov(k, K|Y_2, y_1, d_{x_f})| p(Y_2|y_1, d_{x_f}) dY_2$$

Bayesian A-Optimal

$$\bar{U}_A(d_{x_f}^*) = \min \int_{\Theta} \min_{d_{x_f}} tr(k, K|Y_2, y_1, d_{x_f}) p(Y_2|y_1, d_{x_f}) dY_2$$

Y_2 represents the predicted design points, while y_1 represents the current design. Θ is the parameter space including parameters k and K from the model in Equation 2.1. Similar to

the criteria from Equation 2.2, the Bayesian optimality criteria are minimizing the posterior predictive distribution across the parameter space Θ and design space d_{x_f} . The Bayesian I-optimal designs minimize the distance between the squared predictions. Bayesian D-optimal designs minimize the determinant of the covariance of the posterior predictive distribution. The Bayesian A-optimal designs minimize the trace of the posterior predictive distribution. Implementing the Bayesian criteria into our process of sequential optimality gives various results. The frequency tables illustrate the precision of each criteria based on the probabilities assigned to each candidate point.

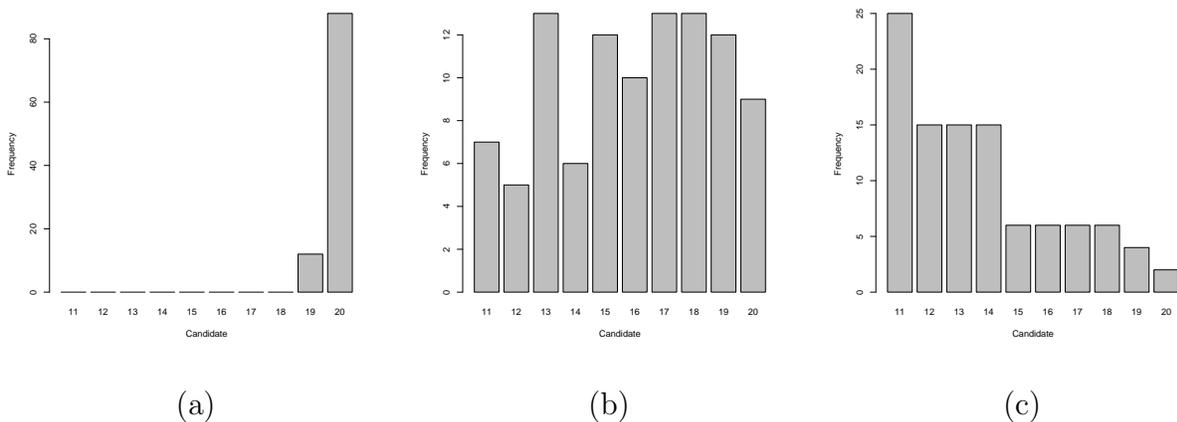
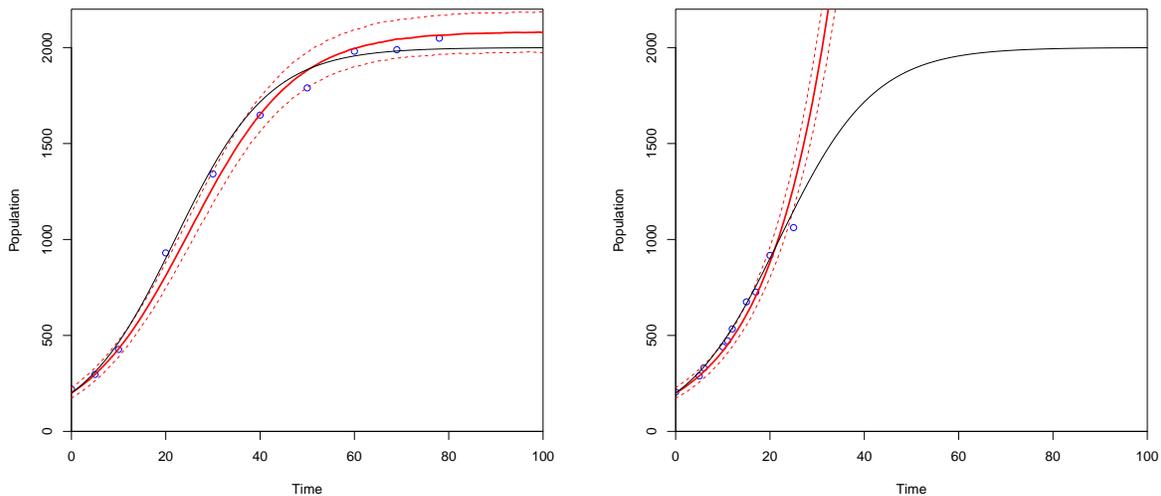


Fig. 12.: Bayesian optimality criteria are used to guide the sequential optimality process. The frequency tables plot the probabilities assigned to the first ten candidate points following the initial base design for a normal growth model. Panel (a) represents candidate probabilities under the \bar{U}_I criterion. Panel (b) plots probabilities of the candidates for \bar{U}_A criterion. Panel (c) provides the frequencies associated with \bar{U}_D criterion.

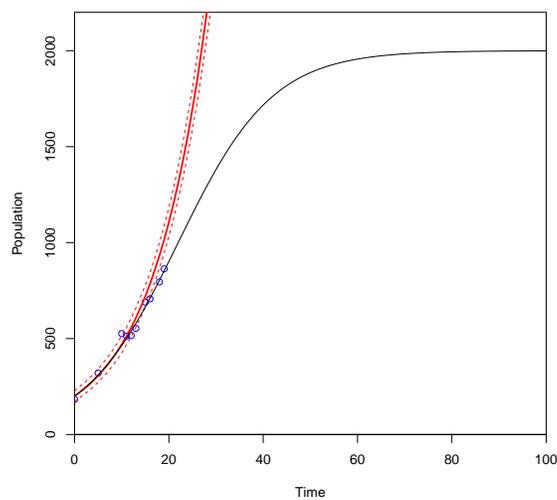
The frequency plots in Figure 12 show the probability densities associated with the first set of candidate points evaluated in the sequential optimality algorithm. The Bayesian I-optimal design gives weight to specific candidate points, whereas the Bayesian A and D optimality criteria provide a wide range of optimal candidates. Based on the frequency charts, it is clear that the Bayesian I-optimal design can provide a precise optimal design

point. Whereas, the Bayesian A and D optimal designs appear to lack precision. These results are further visualized by plotting the Bayesian sequential designs in Figure 13 for the normal growth model.



(a)

(b)



(c)

Fig. 13.: 10,000 MCMC samples simulate a normal growth curve increasing at a 10% rate across 100 time points with a carrying capacity of 2000. Sequential optimality is used with the Bayesian design criteria to find the ten best points for (a) \bar{U}_I , (b) \bar{U}_A and (c) \bar{U}_D optimal designs. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.

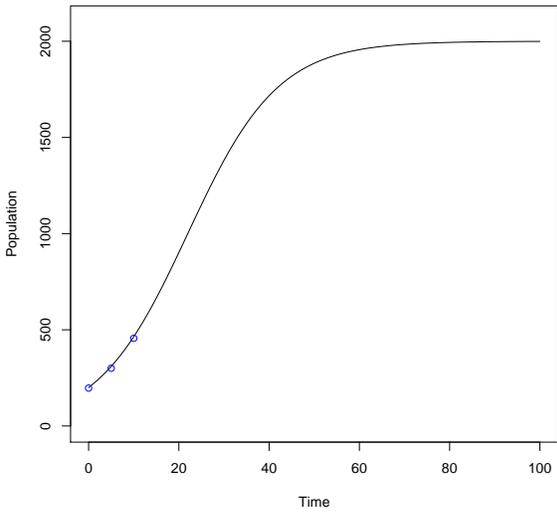
Given all of the criteria and varying growth rates, the simulations using sequential optimality resulted in different designs. Some designs are able to capture the dynamics of the system while others require a larger design point budget. Ultimately, this indicates that there are several ways to design experiments for dynamic models. Using the simple example of population growth modeled by the logistic equation, the sequential optimality simulations demonstrate an adaptation of the commonly performed simulated annealing algorithm. Thus, sequential optimality can be recommended when limitations arise for sampling.

2.4.3 Potential Difficulties

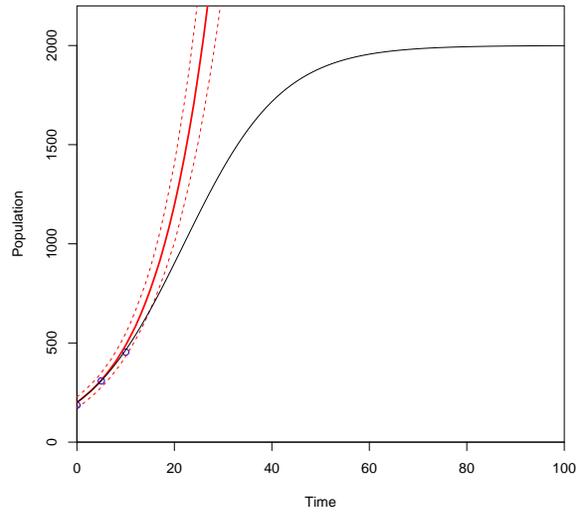
As demonstrated previously, sequential optimality predicts optimal future design points in a sequential manner by learning about the dynamics of a system. Thus far, the simulations use the analytic solution to the logistic equation with Poisson noise to produce simulated data that closely resembles a specified model. This raises concern when evaluating the proposed method given that scenarios exist where logistic growth may not be appropriate. The following simulation details potential difficulties with the sequential optimality algorithm in a scenario that tests if the method can detect a model lack of fit. The simulated data is generated using a piecewise function where the population count suddenly decreases rather than reaching the carrying capacity.

This simulation is performed under the assumption that the logistic growth model is appropriate to fit the simulated data. Figure 14 illustrates sequential optimality being informed by data that does not follow the logistic growth trend. The prediction based I-optimality criterion is used for demonstration purposes. First, an initial design is set of three points given the logistic equation has two parameters, carrying capacity and growth rate. Then, the ten candidate samples following the initial design are evaluated using the I-optimality criterion to select the point with the minimum prediction variance. Once the optimal design point is added, the process repeats and updates three times. However on the fourth run, the

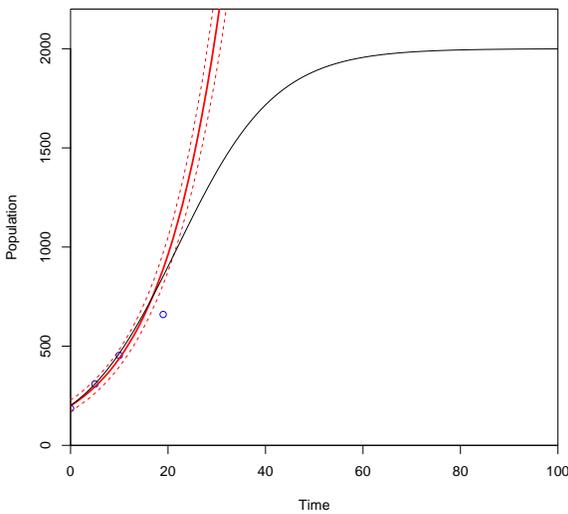
logistic growth curve is poorly fit, and we see a decrease in the population size. By the fifth run and addition of the seventh design point, the I-optimality criterion creates a range that becomes extremely large, which signals an incorrect model. Though the budget is set for ten design points, the process fails to continue once the population reaches zero. This implies that the process can detect a lack of fit and abnormality in the system.



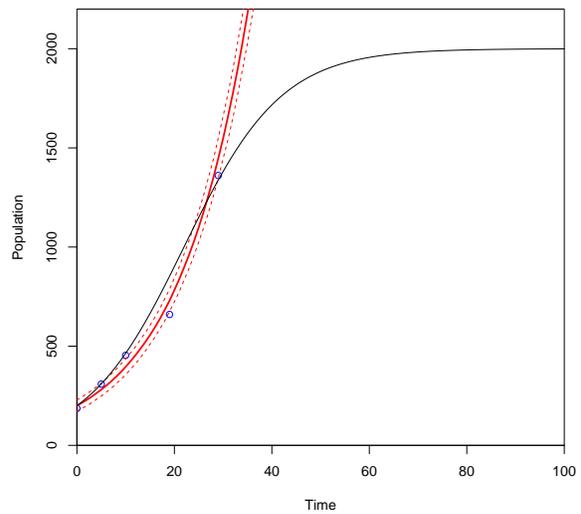
(a)



(b)



(c)



(d)

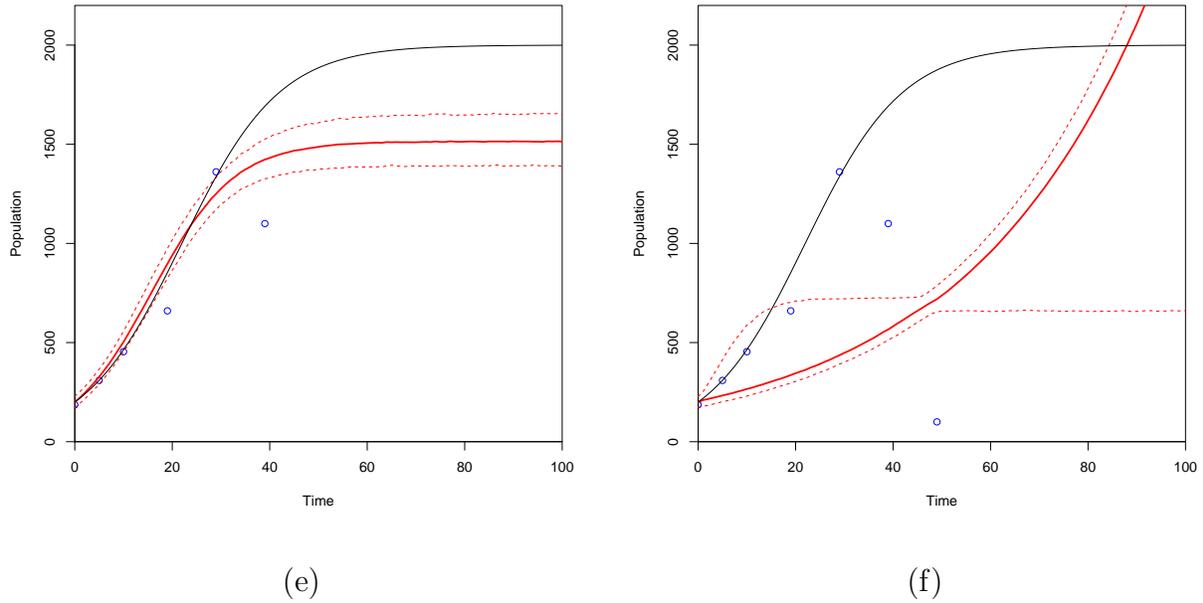


Fig. 14.: 10,000 MCMC samples are used to simulate normal 10% logistic growth across a 100 day period with a carrying capacity of 2000. The ground truth model is plotted as a black curve. I-optimality criterion guides the sequential optimality process. The optimal design points are plotted in blue and are fit by a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values. Panel (a) plots the base design of three points, panel (b) plots the fit of the base design, and panels (c)-(f) plot the fourth through seventh optimal points in the design as they are fit. This simulation demonstrates a lack of fit scenario.

2.5 Discussion

In this study, the dynamics of various logistic growth models were investigated in order to design optimal sampling regimes for ecologists. Efforts were focused on a simple model with limited external factors to demonstrate the approach of sequential optimality. For the purpose of this study, the logistic equation provided a straightforward model used to compare sampling techniques. Simulated annealing was implemented to explore the design

space using various optimality criteria. Despite the growth rate and criterion used, simulated annealing was able to capture the dynamics of the models by exploring the entire design region. However, sequential optimality found sampling regimes that capture the dynamics of a system in a sequential manner.

The proposed method of sequential optimality incorporates prior information into the design process. Rather than exploring the entire design space at once, the sampling method evaluates subsets of the region using parameter based criteria. Sequential optimality captured the dynamics of a normal growth rate using the I optimality criterion. However, implementing this method across criteria and growth rates led to different results. The normal growth rate model was captured by I and A_Φ optimal designs but required more design points for the D_Φ optimal design. Models with fast growth rates were always captured leading to the belief that a smaller design point budget could be explored. On the other hand, slow growth rates required a larger design point budget no matter the criteria used. The algorithm had a more difficult time distinguishing the dynamics of the slower growth rate.

Bayesian optimal designs were also considered as a method for comparison. The \bar{U}_I -optimality criterion had more precise candidate points than the \bar{U}_A or \bar{U}_D optimality criteria. This led to the results that only the \bar{U}_I Bayesian optimal design was able to capture the dynamics of the curve with a set of ten design points. The other two designs required more design points to capture the dynamics of the model. Comparing Bayesian sequential designs to the traditional criteria provided yet another technique available for determining the optimal design points in the region. Comparing these processes across criteria and models suggests that there are several ways to design experiments. Numerous approaches exist for Bayesian experimental design including studies by Müller et al. (2007), Williamson and Goldstein (2012) and Jones et al. (2018). Furthermore, research by Ford et al. (1989) demonstrates advancements in experimental design for nonlinear systems. Though alternative methods are available to help ecologists allocate resources, our approach of sequential optimality is particularly beneficial when predicting the next step, or optimal time, to collect

data.

2.6 Conclusions

In this paper, the ecological model known as the logistic equation was used based on the various applications and popularity of the model. The proposed design space tracked the change in population dynamics across time and was explored using the novel approach of sequential optimality. When using prediction based criteria for a normal growth model, the method captured the dynamics of the system and found an optimal design that translates to an optimal sampling regime for ecologists. Rather than sampling out of convenience or across an equal interval, the simulations demonstrate that the method can provide the next optimal time to sample given the current state of the environment. Real data cannot be incorporated at this time given that this paper introduces a new method for designing sampling regimes. However, the developed method does learn about theoretical processes in a sequential manner and has the potential to assist ecologists when planning sampling schedules.

Though the analysis is performed on a straightforward univariate model, we acknowledge that more complex dynamics exist and can represent more realistic environmental encounters. Incorporating external factors into the model could lead to substantial improvements to the method. Furthermore, there are many design problems that can be investigated by a Bayesian approach. Specifically related to sequential optimality, this study focuses on designs of a set size that explore a set window of time into the future. The method could be expanded by studying varying design sizes and design windows. Also, sampling techniques used by ecologists could be invasive and change the process. These changes could be taken into consideration as they may affect the model. In this paper, we focus on providing optimal designs based on temporal models. However, in the future spatial models could be explored along with a hierarchical network of temporal models. Given these additional conditions, there are many levels of uncertainty that could be incorporated to reflect the reality of an

ecological system.

CHAPTER 3

BAYESIAN EXPERIMENTAL DESIGN WITH THE LOTKA-VOLTERRA DIFFERENTIAL EQUATIONS FOR PREDATOR-PREY DYNAMICS

3.1 Introduction

Community ecologists are interested in studying the temporal and spatial changes in species resembling community structures. The most fundamental relationship within a community is the temporal predator-prey dynamic. Predator-prey interactions highly influence population and community dynamics (Lima, 1998). The removal of prey or flooding of predators can drastically change an entire ecosystem. Hence, ecologists prefer to study predation effects on populations, communities and ecosystems (Crawley, 2009). The most popular representation of predator-prey dynamics can be modeled by the Lotka-Volterra differential equations. Lotka (1926) and Volterra (1928) developed the Lotka-Volterra differential equations as a model for the interactions between a predator and its prey.

The Lotka-Volterra equations have broad applications and are frequently modeled among ecologists. Extensions of the deterministic Lotka-Volterra system have been used by Gard and Hallam (1979) to study food webs as well as Hening and Nguyen (2018b) to study intraspecific competition. Hening and Nguyen (2018a) further study persistence and extinction of species based on the food-chain version of the Lotka-Volterra model. The possibility of two species competing for the same food supply is modeled as a special case by Lotka (1978). However, the traditional model most simply captures realistic encounters between two species in a predator-prey relationship. In this study, the traditional Lotka-Volterra

differential equations are employed to model predator-prey population dynamics.

Many efforts have gone towards estimating the parameters of differential equations. In the Bayesian framework, estimation can be performed using various Markov Chain Monte Carlo (MCMC) sampling techniques (Girolami, 2008). Though ample sampling produces a more accurate representation of dynamical systems, in practice ecologists do not have the resources to excessively sample. Thus, Sequential Optimality (Atanga et al., 2020) is performed to learn about the system in a sequential manner and design optimal sampling regimes according to the population dynamics. Modeling in the Bayesian framework will inform the process and maximize the information obtained from the data. The goal of this simulation study is to demonstrate how sequential optimality can inform ecologists of when to collect data based on the population dynamics associated with the Lotka-Volterra differential equations.

3.2 Predator-Prey Dynamics

The Lotka-Volterra differential equations model the interactions between predators and prey (Lotka, 1926; Volterra, 1928). Though applied frequently in ecology, the model is quite versatile. Variations of the Lotka-Volterra system have been implemented across disciplines. Competitive behavior in the marketplace has been modeled using autonomous Lotka-Volterra systems by Marasco et al. (2016). Gatabazi et al. (2019) use the grey Lotka-Volterra model of higher dimensions to study the interactions between cryptocurrencies. Furthermore, the interaction terms in the Lotka-Volterra equations have represented the effect of technology interaction as demonstrated by Zhang et al. (2018a). Though the model can adapt to specific applications, in this study the traditional model is used to represent an ecological relationship between predator and prey.

The traditional mathematical model for predator-prey interactions are given by the Lotka-Volterra equations (Murray, 2002). The simple first order nonlinear pair of differential

equations are written mathematically as

$$\begin{aligned}\frac{dx}{dt} &= \alpha x - \beta xy \\ \frac{dy}{dt} &= \delta xy - \gamma y\end{aligned}\tag{3.1}$$

where x and y represent the prey and predator respectively. The coefficient α denotes the birth rate of the prey, β is the predation success rate, δ is the efficiency of converting prey into predators and γ represents the mortality rate of the predator. Given the complex features of the model, the oscillating steady-state better represents a realistic relationship between two species. Though ecologists study extinction of species as it affects evolution, the oscillating behavior of the system can be simulated using numerical methods.

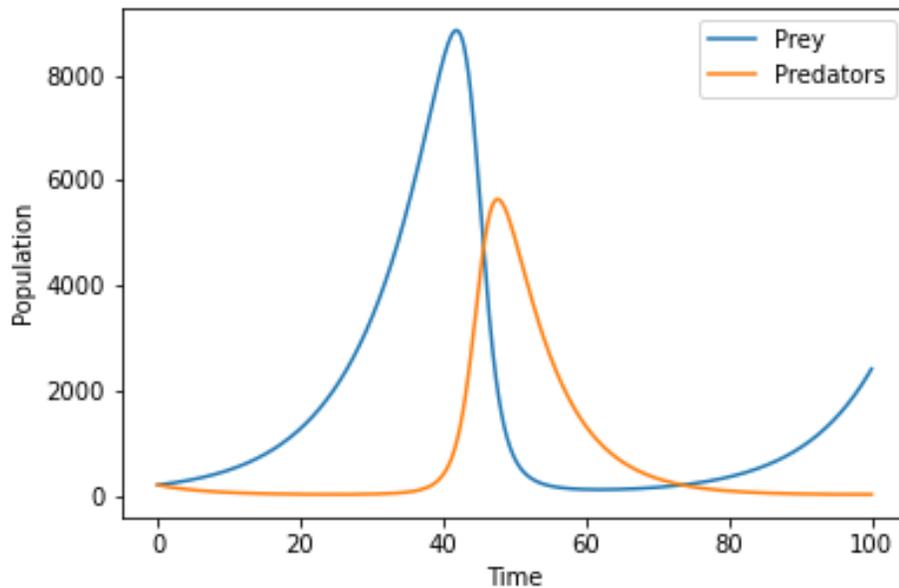


Fig. 15.: The Lotka-Volterra differential equations are simulated using Runge-Kutta methods with parameter values $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The numerical solutions of the prey and predator populations are plotted as blue and orange curves respectively ranging from zero to ten thousand across one hundred time points.

The model is simulated in Figure 15 using the Runge-Kutta method famously developed by Runge (1895) and Kutta (1901). A sequence of approximate solutions is obtained from the differential equations to simulate data that continually oscillates with parameter values $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The defined parameters serve as the true values that need to be estimated in this study using Bayesian inference and design of experiment techniques. In the next section, the Lotka-Volterra model is specified in the Bayesian framework.

3.3 Statistical Model

Ecologists are shifting towards Bayesian modeling of natural systems based on the increasing resources for ecological Bayesian models (Hooten and Hobbs, 2015). Bayesian analysis combines prior beliefs with sample information to make inferences about the *posterior* belief of the population parameter. The posterior is a conditional probability of parameter θ given observed data x defined by Bayes' Theorem (Bayes, 1763)

$$\pi(\theta|x) = \frac{h(x, \theta)}{m(x)},$$

where $h(x, \theta)$ is the joint density of the prior $\pi(\theta)$ and likelihood $f(x|\theta)$

$$h(x, \theta) = \pi(\theta)f(x|\theta),$$

and $m(x)$ is the marginal density of the data

$$m(x) = \int_{\Theta} \pi(\theta)f(x|\theta).$$

Equation 3.1 is a temporal model that tracks the populations of two species across time implying a Poisson likelihood for the data. Thus, the likelihood can be written for both predator and prey as

$$x_i|t_i \sim \text{Poisson}(\lambda_{1i}|t_i)$$

$$y_i|t_i \sim \text{Poisson}(\lambda_{2i}|t_i)$$

where $x_i|t_i$ is the population of the prey at time $t = i$, $y_i|t_i$ is the population of the predator at time $t = i$, λ_{1i} is the numerical solution of the prey population at time t_i and λ_{2i} is the numerical solution of the predator population at time t_i . The numerical solutions are both dependent on parameters α , β , δ , and γ , where each parameter represents a positive rate of change. The gamma distribution has a positive support and is conjugate to the Poisson likelihood. Therefore, informative priors are specified using the gamma distribution as

$$\alpha \sim \Gamma(1.0, 1)$$

$$\beta \sim \Gamma(0.1, 1)$$

$$\delta \sim \Gamma(1.5, 1)$$

$$\gamma \sim \Gamma(0.75, 1).$$

The true parameters for the simulated data are $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. In reality, ecologists at this stage will not know the true parameter values of the model. Though maximum likelihood estimation methods are available for the gamma distribution as demonstrated by Harter and Moore (1965) and Johnson (1970), the specified priors are chosen to represent limited expert knowledge of the system. When incorporating too much information, the algorithm relies heavily on the priors rather than the information provided by the likelihood of the data. During implementation the following priors were tested, $\alpha \sim \Gamma(1, 1)$, $\beta \sim \Gamma(0.01, 0.1)$, $\delta \sim \Gamma(2.25, 1.5)$ and $\gamma \sim \Gamma(0.56, 0.75)$. The hyper-parameters were calculated using method of moments to test more informative priors. However, the simulations performed with great precision when limited data was available implying that experts should know the behavior of the system prior to sampling. Given the purpose of this study is to quantify the uncertainty of a dynamical system, less informative priors were tested. Alternatively, uninformative priors were tested where all priors for α , β , δ and γ were set to $\Gamma(0.1, 0.1)$. This created too much uncertainty in the simulation and caused the algorithm to perform poorly since all four priors provided no information. Thus, the Bayesian model is specified so that MCMC sampling can be employed combined with design

of experiment techniques to estimate the true model parameters.

3.4 Experimental Design

Design of experiments is intended to reduce costs by establishing designs prior to experimentation. Specifically, optimal designs estimate statistical models and improve the precision of statistical inference. Traditional optimal designs applied to predator-prey models by Zhang et al. (2018b) typically optimize the design region by minimizing the variance of the parameter estimates. In this study, traditional optimality criteria is applied to the estimated covariance matrix. The covariance matrix is much easier to estimate in the Bayesian framework using MCMC sampling for complex models compared to calculating the Fisher information matrix. The estimated covariance matrix is written mathematically as

$$Var(\hat{\theta}) = Cov(\hat{\theta}, \hat{\theta}) = E((\hat{\theta} - E(\hat{\theta}))(\hat{\theta} - E(\hat{\theta}))^T),$$

where $\hat{\theta}$ represents a vector of the parameter estimates of the model. The diagonal of the estimated covariance matrix contains the variances of the estimated parameters, and the off-diagonals give the covariances between each of the parameter estimates.

Traditional optimality criteria (Goos and Jones, 2011) are applied to the estimated covariance matrix. Specifically, *A*, *D* and *I* optimal designs are employed in this study. *A* optimality criterion minimizes the trace of the estimated covariance matrix. *D*-optimal designs minimize the determinant of the estimated covariance matrix. *I*-optimal designs are prediction based and minimize the average prediction variance over the entire design region.

$$\begin{aligned} A - \text{Optimal: } & \min tr(Var(\hat{\theta})) \\ D - \text{Optimal: } & \min |Var(\hat{\theta})| \\ I - \text{Optimal : } & \min \bar{Var}(\hat{\theta})_{pred} \end{aligned} \tag{3.2}$$

$\bar{Var}(\hat{\theta})_{pred}$ denotes the average prediction variance of the model parameter estimates. Optimality criteria are measures of fit used to guide optimization processes. Thus, all three of

these criteria are implemented in this simulation study to guide the process of sequential optimality.

3.5 Sequential Optimality

Sequential optimality proposed by Atanga et al. (2020) begins by setting an initial design n , a design point budget b , and window size w . The initial design size is calculated as the number of parameters in the model plus one. The design point budget is up to the discretion of the expert based on available resources, and the window size is determined by dividing the timeline of the experiment by the design point budget. The sequential optimality algorithm is then guided by the set criteria C provided by Equation 3.2. As each design point is accepted, the process updates and continues until a finalized design D is produced.

Algorithm 5: Sequential Optimality

Begin

Choose an initial design t_1, \dots, t_n

Set a design budget, b

Set a design window, w

Set criteria, C :

(i.) $A = \min \text{tr}(\text{Var}(\hat{\theta}))$

(ii.) $D = \min |\text{Var}(\hat{\theta})|$

(iii.) $I = \min \bar{\text{Var}}(\hat{\theta})_{\text{pred}}$

For $D = t_1, \dots, t_n$:

(a) Draw a sample $t^* = \{t_{n+1}, \dots, t_{n+w}\}$

(b) Accept the new state $t_{\text{new}} = \arg C(t^*)$

Repeat until $D = t_1, \dots, t_b$

End

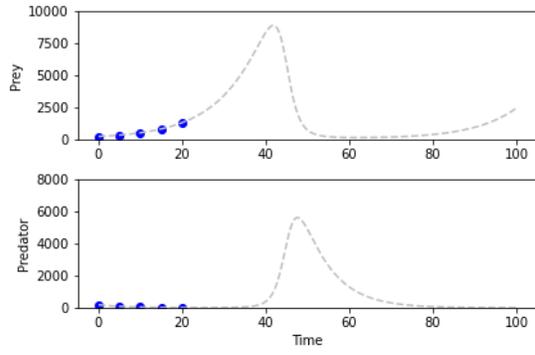
The Bayesian model tracks the population of predators and prey across one hundred time points with ground truth parameters of $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The

simulations use the numerical solution with MCMC sampling to predict the probability of the model parameters. Each design criteria is implemented to find designs with varying budgets and window sizes for comparison purposes. The goal of this design study is to optimize sampling procedures for ecologists by sequentially learning about the predator-prey dynamics of the Lotka-Volterra differential equations.

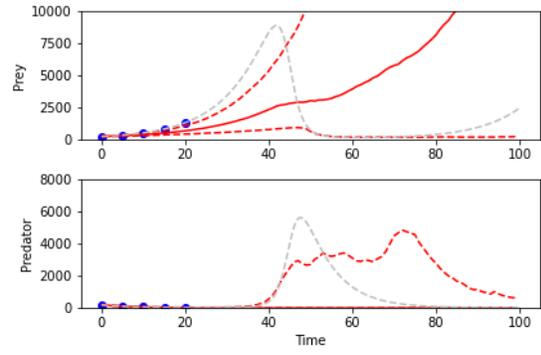
3.6 Simulation Study

The Lotka-Volterra differential equations model the fluctuation of predator and prey populations as they pass through time. All simulations are programmed in Python (Van Rossum and Drake Jr, 1995). The dynamics are simulated where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The simulated data is illustrated in Figure 15 where the blue curve represents the prey and the orange curve represents the predators interacting across a one hundred day period. However in this section, the simulation results are shown on two separate plots to provide clear graphics for each population. The ground truth curves for both predator and prey are plotted as grey dashed lines. The optimal design points are plotted in blue and are fit by a solid red curve calculated by the 50% quantiles of the predicted values of the model. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values and serve as a decision boundary.

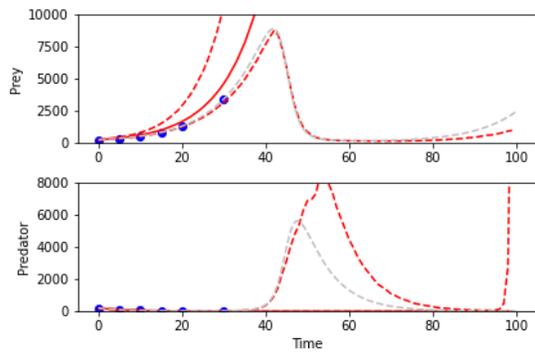
It is of interest to select an optimal design using each criteria across varying budgets and window sizes. The simulation results in Figure 16 demonstrate sequential optimality selecting an I-optimal design of size fifteen. First, the initial design of five points is plotted. Then, the design is fit using the predicted parameter values. The sequential optimality algorithm then searches a design window of size ten to select the next optimal design points and updates the estimates. The process continues until the budget of fifteen points is exhausted.



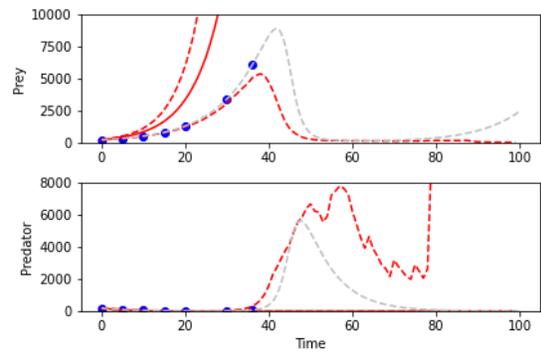
(a)



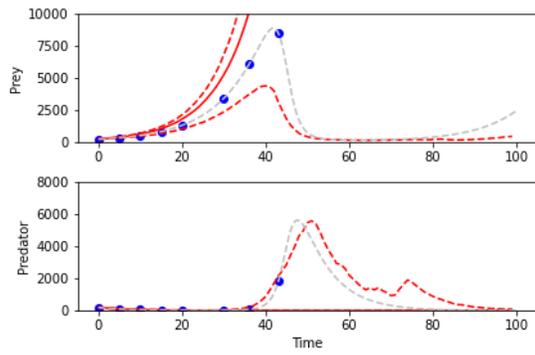
(b)



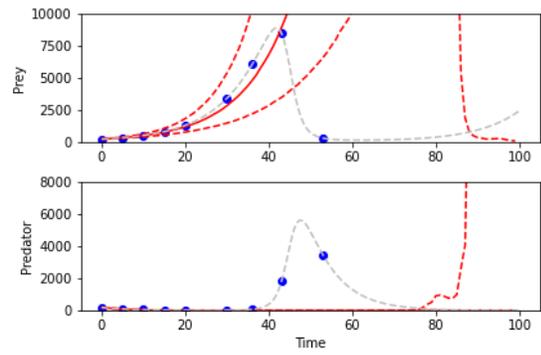
(c)



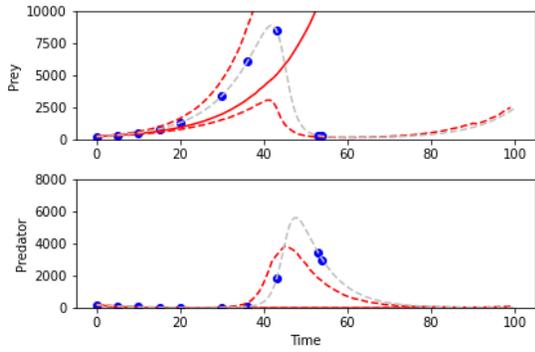
(d)



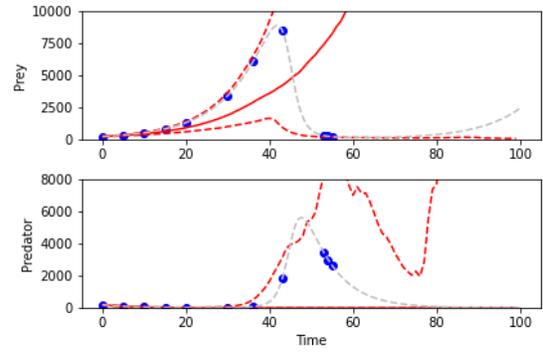
(e)



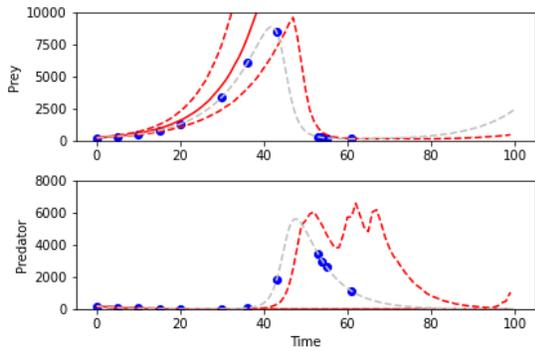
(f)



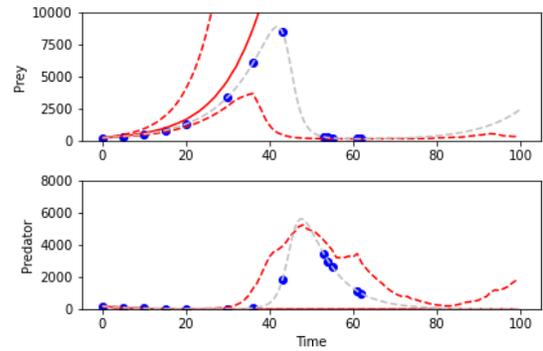
(g)



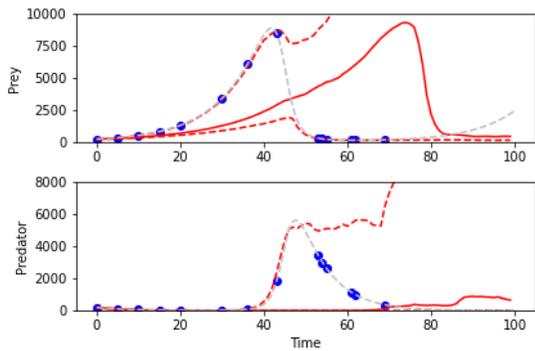
(h)



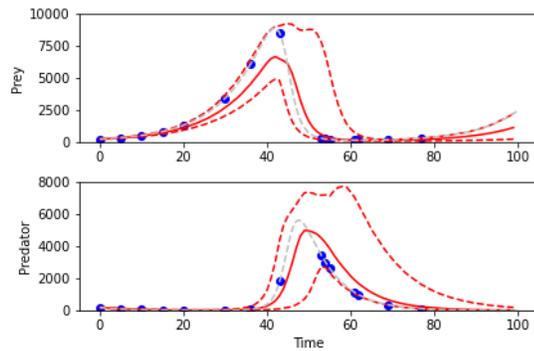
(i)



(j)



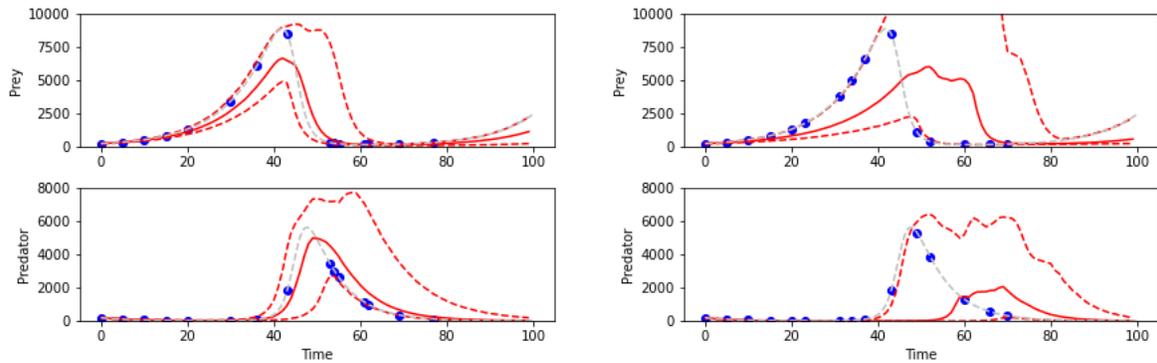
(k)



(l)

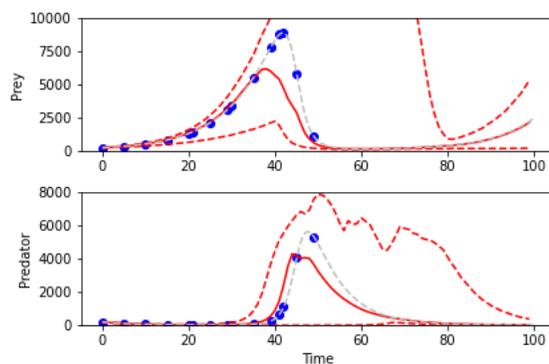
Fig. 16.: I-optimality criterion is used to select the blue design points. The design is fit by a solid red curve and the true values are plotted as dashed grey curves. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Panel (a) plots the initial design, panel (b) fits the design, and panels (c)-(l) represent each optimal point added to the design selected from a window of size ten. The final design consists of fifteen points.

Atanga et al. (2020) demonstrate sequential optimality using various logistic growth models and compare criteria across designs of size ten. However in this study, the model parameters remain the constant and the criteria are compared across varying design budgets and window sizes. The simulation results in Figure 17 compare the optimality criteria across designs with budgets of fifteen points and windows of size ten.



(a)

(b)

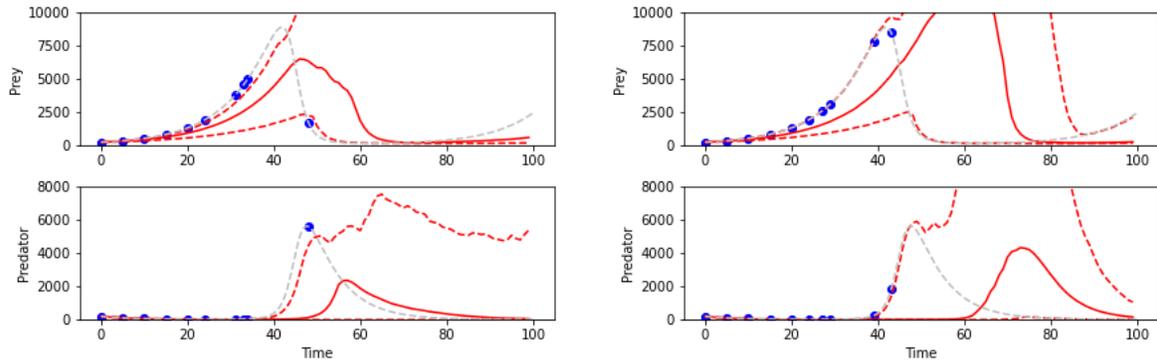


(c)

Fig. 17.: The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size ten to select a final design of size fifteen. The panels plot the final designs using (a) I-optimality, (b) A-optimality, and (c) D-optimality criteria.

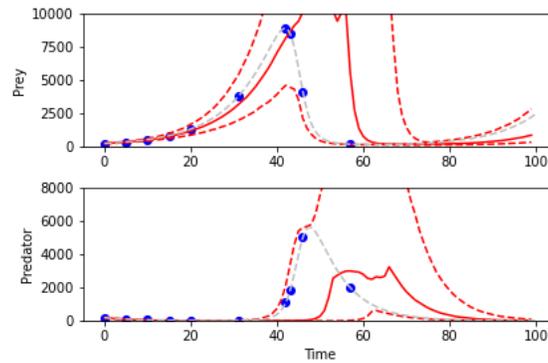
The next set of results simulate optimal designs of size ten as shown in Figure 18. The decrease in the design budget represents limited resources available to ecologists. Given a smaller design point budget, the window size is increased to evaluate fifteen candidate samples. The larger window size allows for optimal placement of design points as well as

another scenario to compare each criterion. Notice in these results the loss in dynamics.



(a)

(b)



(c)

Fig. 18.: The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size fifteen to select a final design of size ten. The panels plot the final designs using (a) I-optimality, (b) A-optimality, and (c) D-optimality criteria.

Ideally, the design point budget should increase so more knowledge can inform the model. In another theoretical scenario, there is potential to design sampling regimes with twenty design points. As the design budget increases, ecologists can explore smaller windows of time

in the future to collect data. In the simulation results provided by Figure 19, the sequential optimality algorithm evaluates windows of size five to select optimal designs consisting of twenty points.

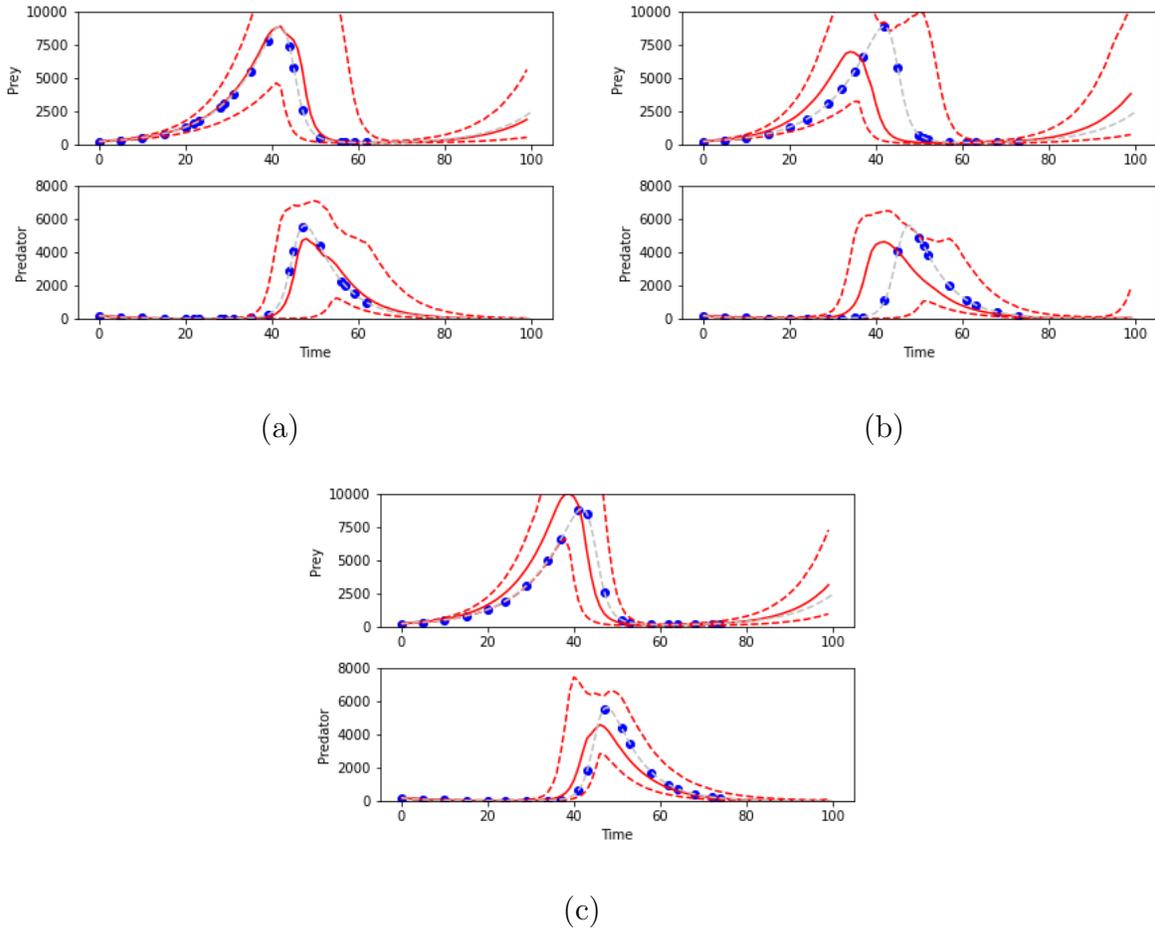


Fig. 19.: The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size five to select a final design of size twenty. The panels plot the final designs using (a) I-optimality, (b) A-optimality, and (c) D-optimality criteria.

The simulations appear to provide a wide range of results. In the first scenario shown by

Figure 17, the I-optimal design captures the dynamics more closely than the A or D-optimal designs. However, decreasing the design budget as shown in Figure 18 causes a loss in dynamics overall. All three criteria provide low budget designs that poorly capture the true parameter values of the specified Lotka-Volterra model. On the other hand, increasing the design budget and evaluating smaller window sizes as shown in Figure 19 creates favorable circumstances. In this case, the larger I, A and D-optimal designs outperform the results from the other simulation scenarios.

This study theoretically designs sampling schedules for ecologists based on simulated predator-prey dynamics. Real data cannot be used given the method of sequential optimality has not yet been implemented in practice. However, each result in this study further validates the process by demonstrating the robustness of the technique. It is clear that there are several ways to design experiments for dynamical systems. Sequential optimality serves as a new and novel approach to sampling that can predict optimal future points when designing experiments.

3.7 Conclusions

The intention of this study is to expand the applications of the sequential optimality algorithm. Implementing the Lotka-Volterra model refines the method by considering more realistic dynamics within an ecosystem. Designs were selected using I, A and D optimality criteria. The designs were compared by altering the budgets and windows evaluated in the process. Based on the results, smaller designs poorly captured the dynamics of the system while larger designs produced more accurate parameter estimates. The main purpose of these simulations was to demonstrate that the method can determine the next optimal time to collect data rather than sampling according to convenience over an extended period of time.

Sequential optimality has thus far focused on temporal models. Specifically in this design study, predator-prey dynamics are explored using the algorithm. However, ecologists

are interested in studying communities involving temporal and spatial changes in ecosystems. Thus, there is potential to appeal to a larger audience by studying spatiotemporal models in the future. The purpose of this design study was to demonstrate a new way to improve sampling methods for ecologists. Therefore as the technique expands and becomes more versatile, it is believed that researchers can implement this method in practice.

CHAPTER 4

RESOURCE-BASED SEQUENTIAL BAYESIAN EXPERIMENTAL DESIGN FOR DYNAMIC MODELING

4.1 Introduction

Ecological population and community studies rely heavily on sampling procedures. Sampling is the first stage of ecological research and directly impacts the entire study. Therefore, it is imperative to develop sampling techniques that ecologists can use when making sampling decisions. General data collection methods are recommended in the ecological literature such as those reviewed by Henderson (2001). Though many sampling procedures exist and vary depending on species and habitat, all methods have underlying expenses that limit the process. Thus, research involving cost-effectiveness such as studies by Caughlan and Oakley (2001) and Mode et al. (1999) are of the utmost importance. The increasing number of sampling design methods are expanding practices for ecologists and evolving ecological sampling techniques.

Bourdeau (1953) studies random and systematic ecological sampling methods to improve sampling efficiency. Similarly, Dennis et al. (2010) focuses on increasing sampling efficiency through replication when monitoring biological populations. Furthermore Schweiger et al. (2016) optimize sampling approaches using reproducible, statistical simulations to combat erroneous conclusions made by studies with flawed sampling design. Given the many criticisms against the reliability of ecological sampling designs, the purpose of this simulation study is to further develop design algorithms that can determine the optimal budgets needed

when allocating resources and monitoring species composition and abundance.

Ecologists mathematically model ecological systems and are interested in estimating various population dynamics. As seen in McCallum (2008), there are various methods for population parameter estimation. Though the frequentist approach is favorable, the Bayesian framework provides Markov Chain Monte Carlo sampling techniques (Brooks et al., 2011) that can be used to estimate dynamical systems. Given raised concerns with cost-effectiveness, design of experiment techniques are incorporated with Bayesian inference to further extend Sequential Optimality as proposed by Atanga et al. (2020). This design study focuses on developing a stopping criterion for the algorithm that determines the optimal budget size necessary to capture the dynamics of ecological systems. The logistic equation and Lotka-Volterra differential equations are modeled to demonstrate the technique across simple and complex ecological systems. The intention of this simulation study is to broaden statistical methods available to ecologists for designing sampling regimes.

4.2 Statistical Models

Ecologists have expertise in modeling ecological systems. Thus, the Bayesian approach can be used to combine prior knowledge with the likelihood of an experiment to probabilistically specify various dynamic models. The Bayesian approach comes from Bayes' Theorem (Bayes, 1763). The theorem requires the specification of prior beliefs and merges this knowledge with the likelihood of an experiment to form a posterior belief. Prior knowledge is always assumed to exist about a parameter θ and can be written as $\pi(\theta)$. The conditional probability of the data x given the parameter θ is known as the likelihood of the experiment noted as $L(x|\theta)$. The posterior distribution is then defined as the conditional probability of parameter θ given the observance of data x , which can be written mathematically as

$$P(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta}.$$

In the next two sections, the ecological models used for simulation purposes are intro-

duced and specified in this Bayesian framework.

4.2.1 The Logistic Equation

The logistic equation was proposed by Verhulst (1838) as a solution to the dilemma of exponential growth. The rediscovery of the equation by Reed and Pearl (1927) was introduced as a self-regulating population growth model. The logistic equation has since inspired much ecological research and has grown in popularity as seen in works by Feller (1940) and Hutchinson (1978). As a simple and straightforward ecological model, the logistic equation is written mathematically as

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) \quad (4.1)$$

with solution

$$N(t) = \frac{KN_0}{(K - N_0)e^{-rt} + N_0}.$$

The equation tracks the size of a population N across time t . Parameter r is the population growth rate, parameter K is the carrying capacity, and the initial population N_0 represents the population at time $t = 0$. The logistic equation models species abundance across time. Hence, the model can be specified in the Bayesian framework with a Poisson likelihood

$$N_i|t_i \sim \text{Poisson}(\lambda_i|t_i),$$

where $\lambda_i = N(t)$ which depends on the parameters r and K . The prior distributions for the growth rate and carrying capacity parameters are specified as

$$r \sim \text{Lognormal}(1, 10)$$

$$K \sim \text{Lognormal}(2000, 0.10),$$

where these specified priors represent expert prior knowledge of the system. This Bayesian model is used for all simulations involving logistic growth. Given the logistic equation is commonly used to model the growth of a single species system, in the next section a more

complex model is specified to represent species interactions.

4.2.2 The Lotka-Volterra Differential Equations

The Lotka-Volterra differential equations are implemented as a more realistic model demonstrating the relationship between multiple species. Lotka (1926) and Volterra (1928) developed these equations to represent predator-prey interactions. The simple first order nonlinear pair of differential equations are written mathematically as

$$\begin{aligned}\frac{dx}{dt} &= \alpha x - \beta xy \\ \frac{dy}{dt} &= \delta xy - \gamma y\end{aligned}\tag{4.2}$$

where x represents the prey population, and y represent the predator population. The parameters α , β , δ and γ represent the birth rate of the prey, predation success rate, efficiency of converting prey into predators and mortality rate of the predator respectively. The Lotka-Volterra model can be specified in the Bayesian framework. The abundance of each species is tracked across time implying a Poisson likelihood

$$\begin{aligned}x_i|t_i &\sim \text{Poisson}(\lambda_{1i}|t_i) \\ y_i|t_i &\sim \text{Poisson}(\lambda_{2i}|t_i)\end{aligned}$$

where $x_i|t_i$ is the population of the prey at time t_i and $y_i|t_i$ is the population of the predator at time t_i . λ_{1i} and λ_{2i} represent the numerical solutions of the prey and predator populations at time t_i respectively. The numerical solutions to the differential equations depend on α , β , δ , and γ . The gamma distribution has a positive support and is conjugate to the Poisson likelihood, which can serve as a prior that ensure positive rates of change for each parameter.

Informative priors are specified as

$$\alpha \sim \Gamma(1.0, 1)$$

$$\beta \sim \Gamma(0.1, 1)$$

$$\delta \sim \Gamma(1.5, 1)$$

$$\gamma \sim \Gamma(0.75, 1).$$

The real parameters for the simulations are $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. Given ecologists at this stage will not know the true parameter values of the model, the above priors represent limited expert knowledge of the system.

4.3 Sequential Optimality

Sequential optimality (Atanga et al., 2020) sets an initial design n , a design point budget b , window size w and predicts the optimal design points in a sequential manner guided by optimality criteria. The set of criteria C consists of choosing either I, A, and D optimality criterion defined traditionally by Montgomery (2017). I-optimal designs minimize the average prediction variance of the of the parameter estimates. A-optimal designs minimize the trace of the estimated covariance matrix. D-optimal designs minimize the determinant of the estimated covariance matrix. Using this same criteria and algorithm, this study updates the method by creating a stopping criterion that eliminates the need to set a budget b .

Instead, an initial design is chosen of size n along with a window size w . Rather than proceeding through the algorithm until the budget is exhausted, the process halts once the decision boundary or predictive region stabilizes. Though a similar method is performed by Pagendam and Pollett (2009), the standard errors cannot be used in practice to stop the process since ecologists will not know the true model and parameter values. Thus, the average distance between the 2.5% and 97.5% prediction quantiles (B_{mean}) is calculated for each time step, and the criterion is set to stop the process when the difference in the mean boundaries stabilizes or minimizes between time steps, $S = |(B_{mean}(t_i) - B_{mean}(t_{i-1}))| < V$.

The stopping criterion must fall below a specified variability V that may change depending on the model. Indicating that the variation V is small implies that the process has stabilized. Once the process stabilizes, the final design point is added and a final design D is produced. Thus, this proposed algorithm provides an optimal budget that captures dynamic models.

Algorithm 6: Sequential Optimality

Begin

Choose an initial design t_1, \dots, t_n

Set a design window, w

Set optimality criteria, $C = I, A,$ or D

For $D = t_1, \dots, t_n$:

(a) Draw a sample $t^* = \{t_{n+1}, \dots, t_{n+w}\}$

(b) Accept the new state $t_{new} = \arg C(t^*)$

Repeat until $S = |(B_{mean}(t_{new}) - B_{mean}(t_{new-1}))| < V$

End

Sequential optimality is implemented on the statistical models specified earlier in the Bayesian framework. In the next section, the ground truth models are discussed and optimal designs are chosen for various scenarios.

4.4 Simulation Study

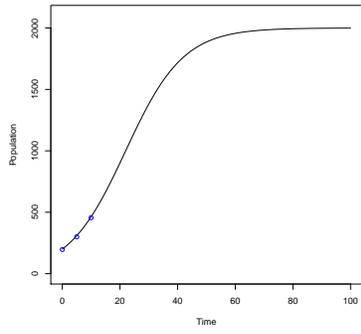
4.4.1 Logistic Growth

In this section, the sequential optimality algorithm is implemented across various logistic growth models. Normal, Fast and Slow growth are simulated with rates of 10%, 5%, and 100% respectively. All models have a carrying capacity of two thousand. Each simulation plots the ground truth dynamics as black curves across a one hundred day period. For comparison purposes, the window size w in the algorithm is set as ten and the selected designs are chosen using I, A and D optimality criteria. The stopping criterion for these models is set to be lower than a variability V of ten. This level of variability is chosen based

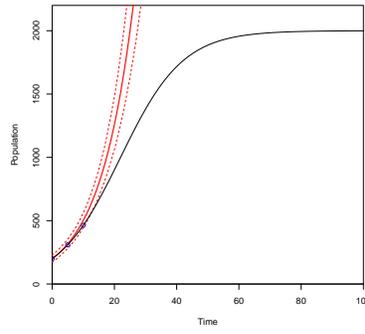
on the simplicity in the model dynamics. The design points are plotted in blue and all base designs consist of three points. The designs are fit with a solid red curve and a decision boundary plotted as red dashed lines calculated by the 2.5% and 97.5% prediction quantiles. The first simulation in Figure 20 demonstrates the process of sequential optimality guided by I optimality criterion on a normal growth model. Table 2 provides the corresponding design step, design size, parameter estimates and stopping criterion for each panel represented in Figure 20.

Design Step	Design Size	Parameter Estimates		$S < V = 10$
Base Design	3	$r = 0.09199$	$K = 10384280.00$	
Step 1	4	$r = 0.07806$	$K = 27105724.00$	$S = 227027.46$
Step 2	5	$r = 0.06768$	$K = 19204298.00$	$S = 26523.50$
Step 3	6	$r = 0.05798$	$K = 18348205.00$	$S = 4799.73$
Step 4	7	$r = 0.07686$	$K = 17342.41$	$S = 2468.94$
Step 5	8	$r = 0.09448$	$K = 2137.59$	$S = 4118.63$
Step 6	9	$r = 0.09904$	$K = 2072.92$	$S = 17.45$
Step 7	10	$r = 0.09743$	$K = 2080.52$	$S = 5.65$

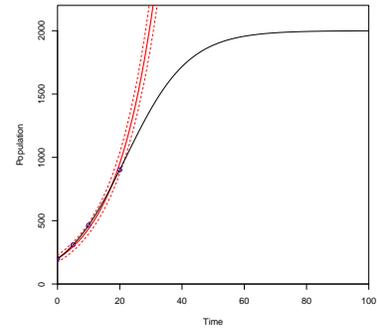
Table 2.: I-Optimal Design Steps for Normal Growth



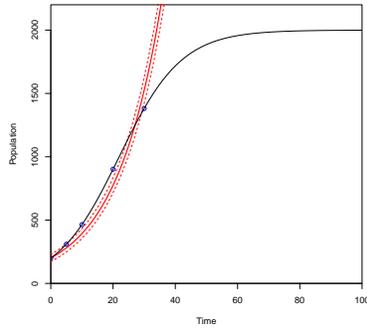
(a)



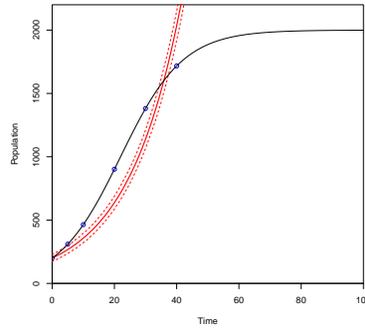
(b)



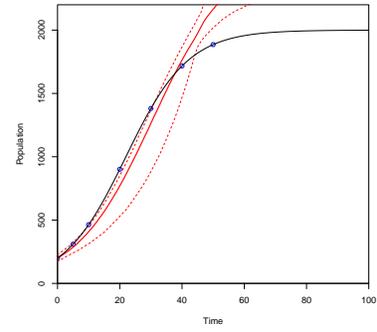
(c)



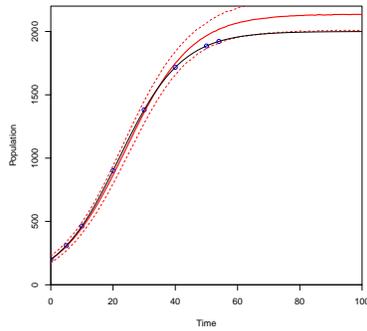
(d)



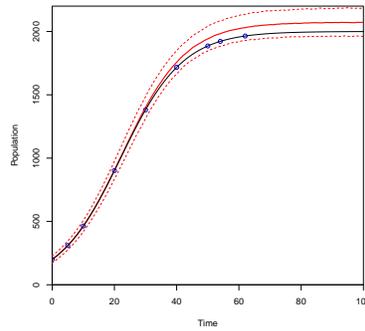
(e)



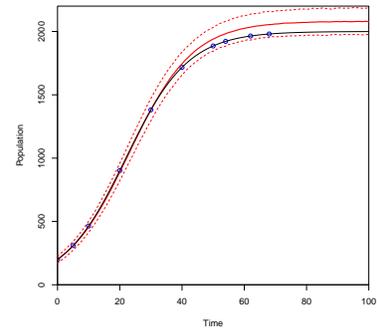
(f)



(g)



(h)



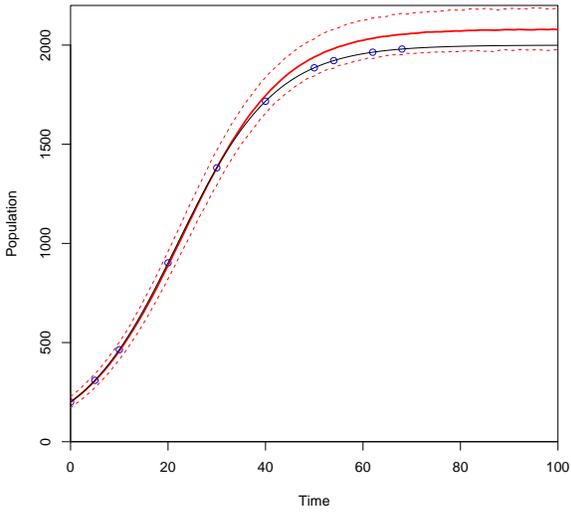
(i)

Fig. 20.: Sequential optimality plots the base I-optimal design (a), the fit of the base design (b), and the fourth to tenth points as they are added to the design (c)-(i) for a normal growth model. The ground truth model is plotted as a black curve. The design points are plotted in blue and are fit by a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values.

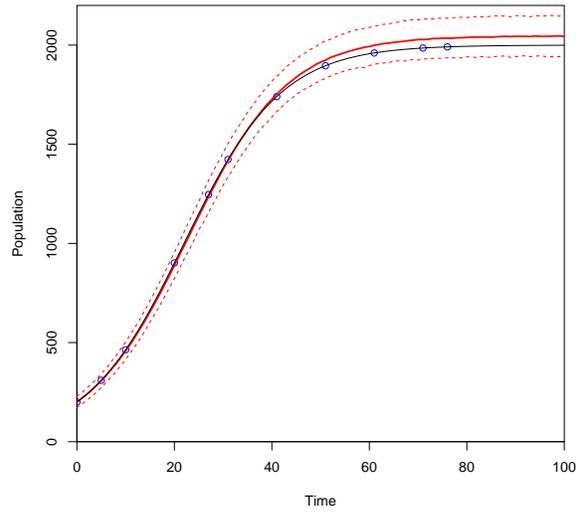
The next set of simulations in Figure 21 compare the final sequential designs of a normal growth model for I, A and D optimality criteria. Table 3 compares the designs simulated in Figure 21 by recording the budget sizes, parameter estimates and corresponding stopping criterion. Notice that the I-optimal design captures the dynamics with a budget of size ten, whereas A and D optimal designs require larger budgets of sizes eleven and fifteen to stop the process. All of the parameter estimates are relatively close to the true values and stabilize similarly by the final step.

Design Criterion	Budget Size	Parameter Estimates		$S < V = 10$
I	10	$r = 0.09743$	$K = 2080.52$	$S = 5.65$
A	11	$r = 0.09791$	$K = 2046.45$	$S = 4.37$
D	15	$r = 0.09591$	$K = 2099.31$	$S = 4.91$

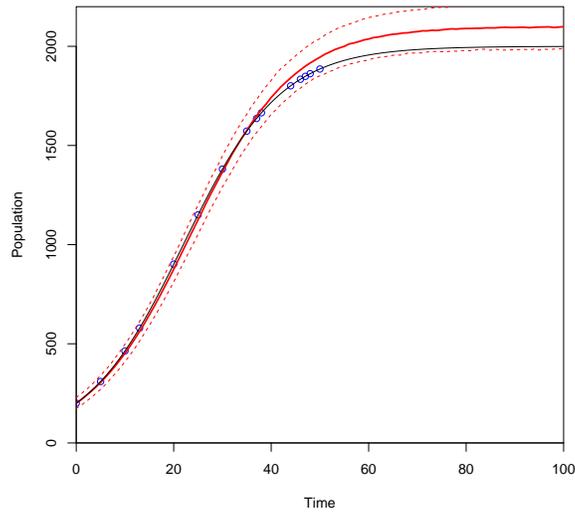
Table 3.: Normal Logistic Growth Designs



(a)



(b)



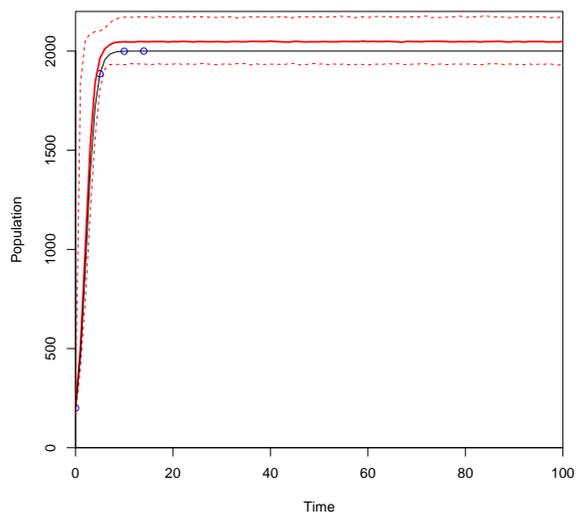
(c)

Fig. 21.: Sequential optimality is used to find (a) I, (b) A and (c) D optimal designs across normal logistic growth models. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.

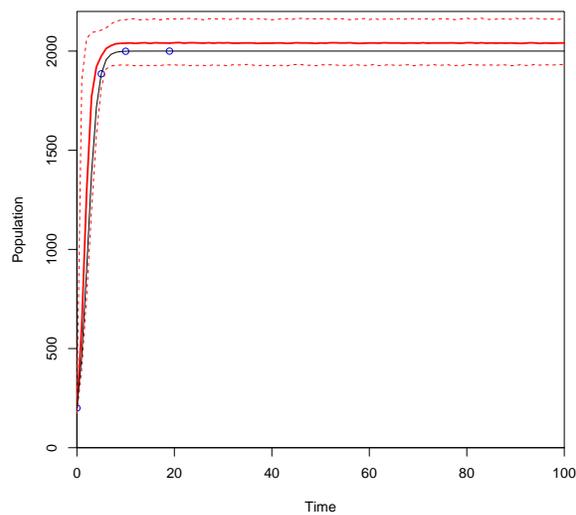
The simulations illustrated in Figure 22 provide the optimal designs chosen for fast growth models across all optimality criteria. Table 4 compares the designs in Figure 22 with the recorded budget sizes, parameter estimates and stopping criterion. Notice that fast growth models require less design points to capture the population dynamics. Though the I and A-optimal designs only require four design points, the D-optimal design needed eight points to capture the dynamics.

Design Criterion	Budget Size	Parameter Estimates		$S < V = 10$
I	4	$r = 1.75723$	$K = 2049.17$	$S = 2.93$
A	4	$r = 1.91222$	$K = 2041.99$	$S = 5.50$
D	8	$r = 2.75906$	$K = 1997.40$	$S = 4.24$

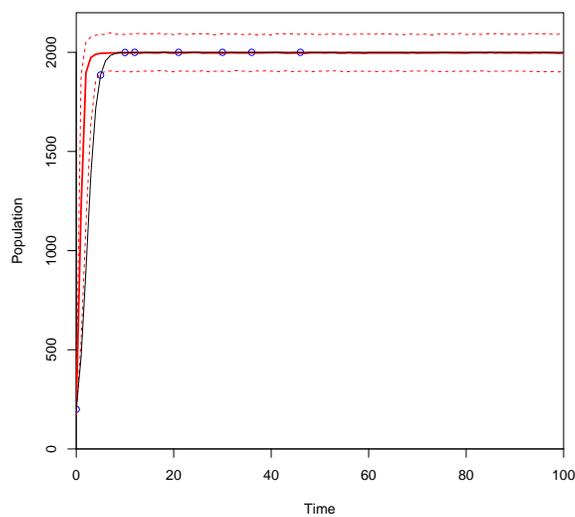
Table 4.: Fast Logistic Growth Designs



(a)



(b)



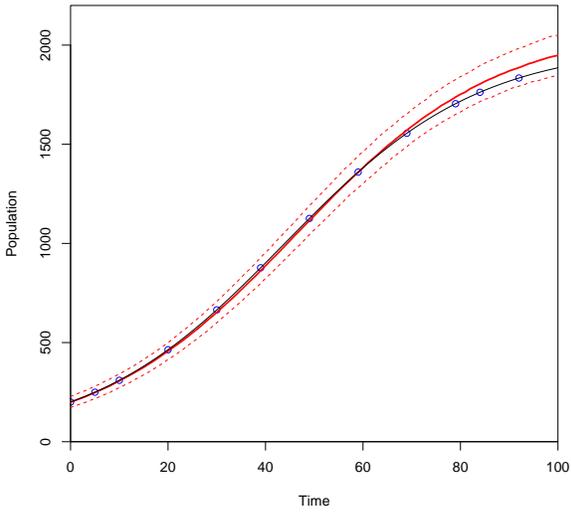
(c)

Fig. 22.: Sequential optimality is used to find (a) I, (b) A and (c) D optimal designs across fast logistic growth models. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.

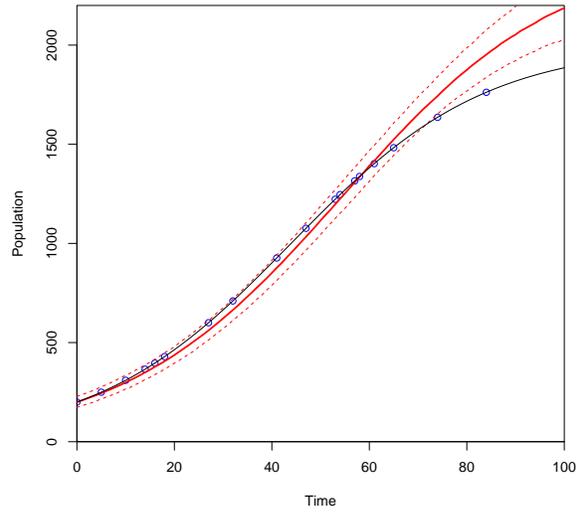
Figure 23 illustrates I, A and D optimal designs selected for slow growth models. Table 5 compares the designs in Figure 23 reports the information about each design with the budget size, parameter estimates and stopping criterion. It is clear that the slow growth models have subtle dynamics that require larger design point budgets. Again, the I-optimal design requires the smallest budget of thirteen compared to the A and D-optimal designs which require budgets of size eighteen and nineteen. This trend across all scenarios and designs demonstrates that the I optimality criterion most efficiently captures the model dynamics.

Design Criterion	Budget Size	Parameter Estimates		$S < V = 10$
I	13	$r = 0.04855$	$K = 2094.80$	$S = 6.07$
A	18	$r = 0.04468$	$K = 2482.20$	$S = 2.63$
D	19	$r = 0.04783$	$K = 2174.10$	$S = 2.19$

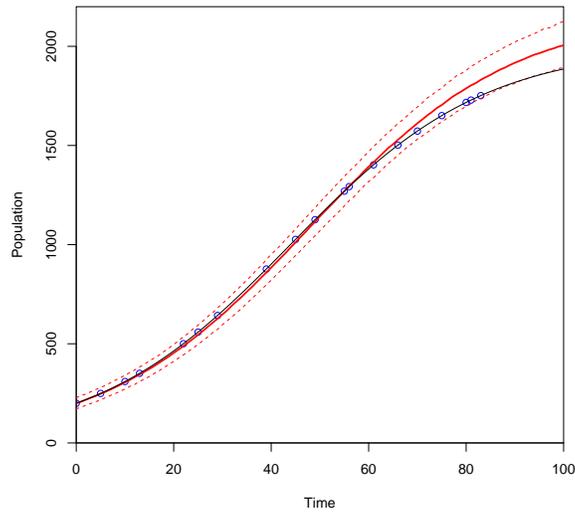
Table 5.: Slow Logistic Growth Designs



(a)



(b)



(c)

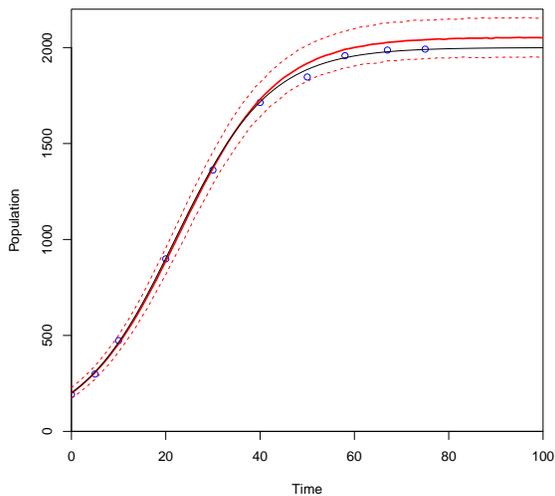
Fig. 23.: Sequential optimality is used to find (a) I, (b) A and (c) D optimal designs across fast logistic growth models. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model.

Additional Comparison of Sampling Regimes

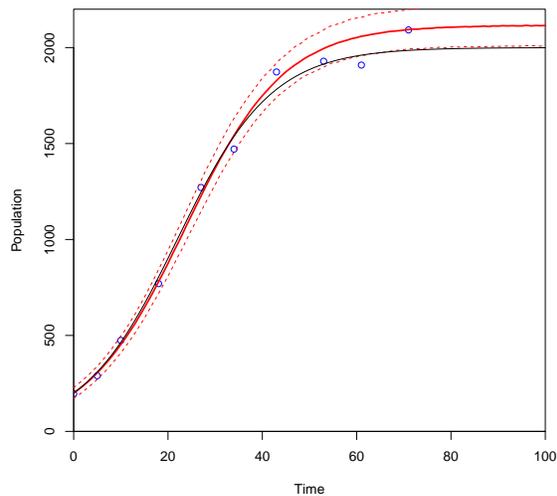
The simulation results thus far compare sampling regimes produced by Algorithm 6 based on the allocated budgets. The intended goal of this study is to convince ecologists to use the proposed statistical method over current practices. Therefore, the next set of simulations compare the designs produced by sequential optimality with a random design produced by randomly sampling within windows of time. The random design represents convenience sampling as a current method that ecologists use when collecting data. Since this is a simulation study and the behavior of the true model is known, the coefficient of determination, also known as R-squared, can be calculated for each predicted model to provide a more robust and straightforward comparison for ecologists. Rather than using the algorithm with a stopping criteria to compare sampling budgets, each predicted model will be tested by fit. For comparison purposes, the results in this section produce designs of size ten for a normal logistic growth model.

Design	Parameter Estimates		R^2
I	$r = 0.0981$	$K = 2040.35$	0.9967
A	$r = 0.0896$	$K = 2232.20$	0.9855
D	$r = 0.0889$	$K = 3754.77$	0.8881
Random	$r = 0.0578$	$K = 24887376$	0.0261

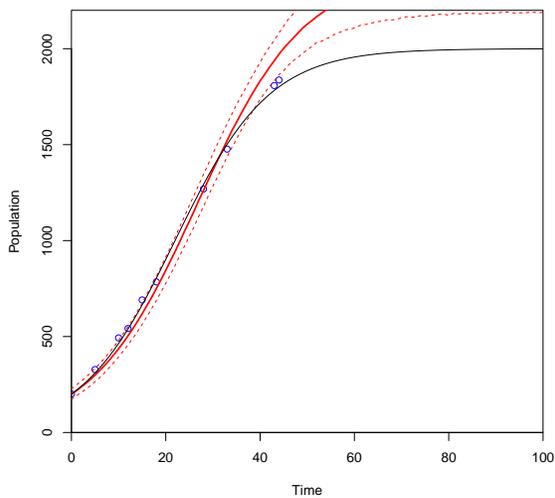
Table 6.: Comparing Fit of Designs



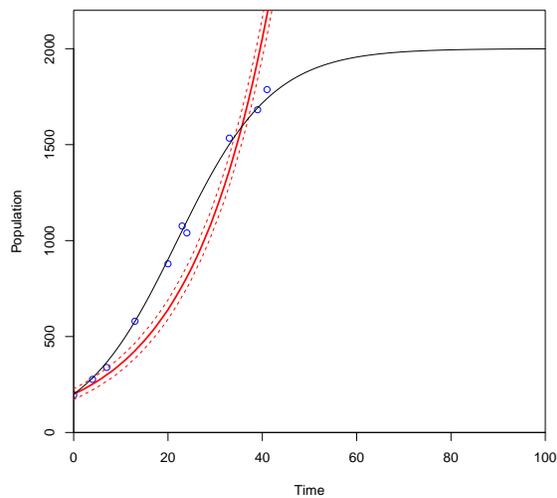
(a)



(b)



(c)



(d)

Fig. 24.: Sequential optimality is used to find (a) I, (b) A and (c) D optimal designs for a normal logistic growth model. Panel (d) compares convenience sampling by plotting a randomly generated design. The optimal points are plotted in blue and fit with a solid red curve. The red dotted lines represent the 2.5% and 97.5% quantiles of the predicted values of the model. The black curve represents the ground truth model. The designs are compared in Table 6.

Previous simulations indicated that the I-optimal designs outperformed the other criteria in terms of cost efficiency. Though final budget size indicates a better performance, the designs can also be compared by the model goodness-of-fit. Table 6 compares the designs by their R-squared values and provides the corresponding parameter estimates. For each design of size ten illustrated in Figure 24, the R-squared values reflect the fit of each model.

Notice that the I-optimal and A-optimal designs have good fits with 99.67% and 98.55% of the variability from the true model accounted for by the predicted models. Whereas, the D-optimal design has less of a fit to the true model with an R-squared of 88.81%. This is not surprising given all previous results indicated a need for a larger budget when using D optimality criterion. However, the randomly collected data in comparison to the others also performed poorly with an extremely low R-squared of 2.61%. If ecologist were to randomly collect data, it is clear by these results that convenience sampling would require a larger sampling budget. Though the designs are similar, these results show that using a specified criterion to select the optimal points is preferred to randomly selecting days in a season to sample. All of the results in this paper thus far demonstrate the benefits of the sequential optimality algorithm. In the next section, more complex dynamics will be incorporated into the process to provide more realistic simulation results.

4.4.2 Predator-Prey Dynamics

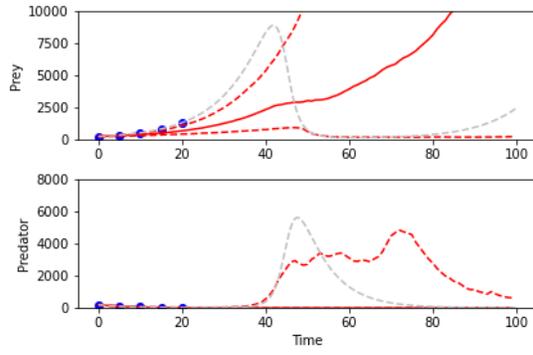
Predator-Prey dynamics are explored in this section using the Lotka-Volterra differential equations with true parameter values of $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The simulations plot the ground truth model as grey dashed lines where the top plot represents the prey and the lower plot illustrates the predators. The selected design points are blue and fit with a solid red curve. Again, the 2.5% and 97.5% prediction quantiles are plotted as red dashed lines. Rather than creating scenarios with varying parameter values, these simulations are compared by setting different windows w in the algorithm. Given the complex dynamics, there is more variance in the decision boundary. Thus, the stopping criterion in

this section is set to fall below one hundred.

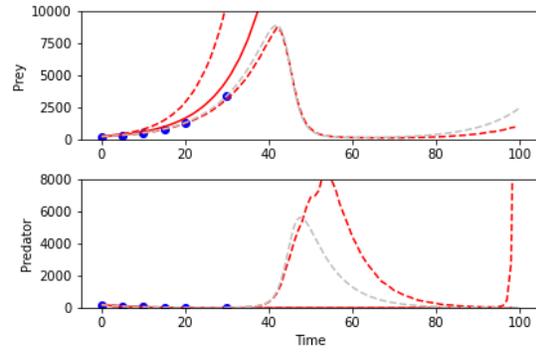
The first simulation in this section demonstrates sequential optimality selecting an I-optimal design for the provided model searching windows of size ten. Figure 25 illustrates each step in the process while Table 7 provides the details of each step including the Design Step, budget size b , parameter estimates and stopping criterion. Sequential optimality captures the dynamics of the Lotka-Volterra with an I-optimal design of size fifteen. Unlike the logistic growth results, the parameter estimates begin relatively close to the true values. However, there is still a decrease in the size of the decision boundary, which leads to the developed stopping criterion.

Design Step	b	Parameter Estimates				$S < V = 100$
Base Design	5	$\alpha = 0.6504$	$\beta = 0.000034$	$\delta = 1.2896$	$\gamma = 0.8585$	
Step 1	6	$\alpha = 1.0444$	$\beta = 0.000028$	$\delta = 1.2789$	$\gamma = 0.6685$	$S = 1547.59$
Step 2	7	$\alpha = 1.4026$	$\beta = 0.000002$	$\delta = 1.2831$	$\gamma = 0.7925$	$S = 15846376$
Step 3	8	$\alpha = 1.0873$	$\beta = 0.000002$	$\delta = 1.2476$	$\gamma = 0.8317$	$S = 15847395$
Step 4	9	$\alpha = 0.8810$	$\beta = 0.000003$	$\delta = 1.4689$	$\gamma = 0.7702$	$S = 13307.16$
Step 5	10	$\alpha = 0.7593$	$\beta = 0.000044$	$\delta = 1.7771$	$\gamma = 0.6778$	$S = 13806.63$
Step 6	11	$\alpha = 0.7451$	$\beta = 0.000048$	$\delta = 1.2507$	$\gamma = 1.0021$	$S = 18533.48$
Step 7	12	$\alpha = 1.0212$	$\beta = 0.000014$	$\delta = 1.2125$	$\gamma = 0.8595$	$S = 17586.48$
Step 8	13	$\alpha = 1.1189$	$\beta = 0.000005$	$\delta = 1.4858$	$\gamma = 0.4799$	$S = 148.43$
Step 9	14	$\alpha = 0.6359$	$\beta = 0.055385$	$\delta = 1.4942$	$\gamma = 0.8496$	$S = 4055.88$
Step 10	15	$\alpha = 0.9171$	$\beta = 0.085404$	$\delta = 1.3222$	$\gamma = 0.9341$	$S = 3616.31$
Step 11	15	$\alpha = 0.9718$	$\beta = 0.102767$	$\delta = 1.5043$	$\gamma = 0.8988$	$S = 18.77$

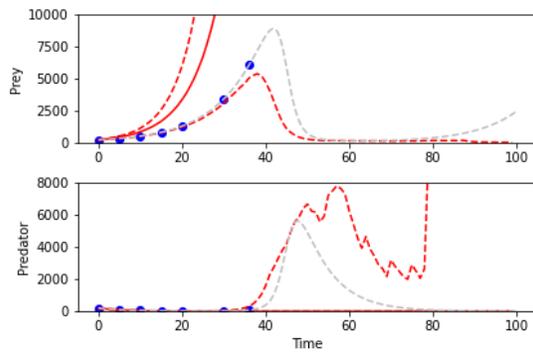
Table 7.: I-Optimal Design Steps Exploring Window of Size 10



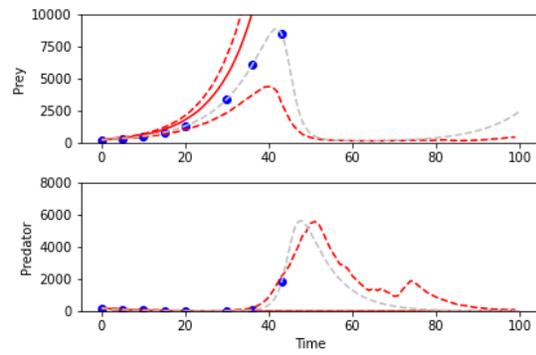
(a)



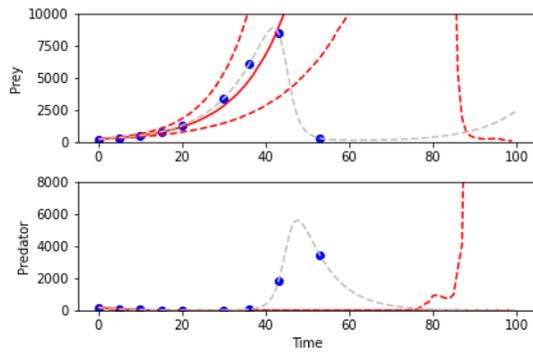
(b)



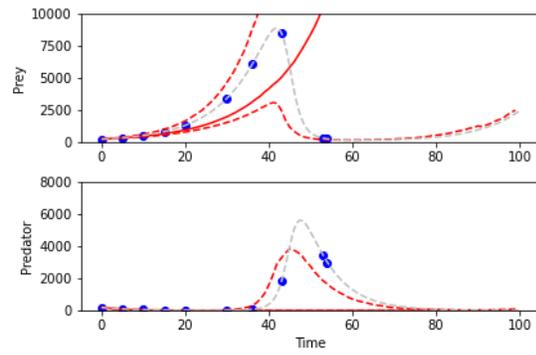
(c)



(d)



(e)

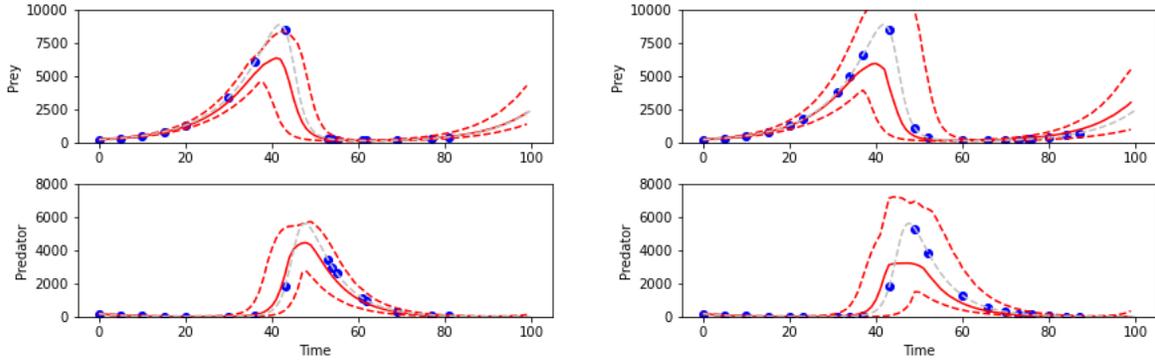


(f)

The next set of simulations illustrate sequential optimality selecting designs searching windows of size ten. Figure 26 illustrates each of the designs per panel, and Table 8 records the corresponding budget size b , parameter estimates and stopping criterion. Notice when searching windows of size ten, the designs require large budgets. The I-optimal design captures the dynamics with sixteen points, the A-optimal design captures the dynamics with twenty-one points, and the D-optimal design captures the dynamics with twenty-three design points.

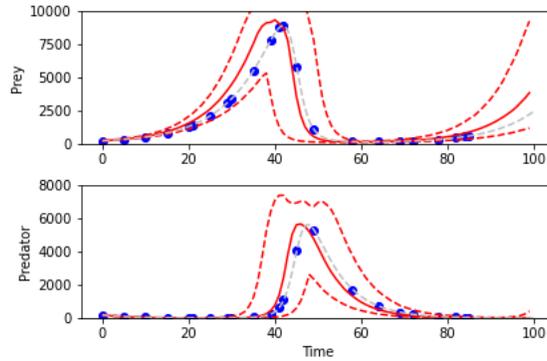
Design Criteria	b	Parameter Estimates				$S < V = 100$
I	16	$\alpha = 0.9718$	$\beta = 0.102767$	$\delta = 1.5043$	$\gamma = 0.8988$	$S = 18.77$
A	21	$\alpha = 1.001$	$\beta = 0.110406$	$\delta = 1.5225$	$\gamma = 0.6685$	$S = 12.95$
D	23	$\alpha = 1.4026$	$\beta = 0.000002$	$\delta = 1.2831$	$\gamma = 0.8157$	$S = 67.6$

Table 8.: Lotka-Volterra Designs Exploring Windows of Size 10



(a)

(b)



(c)

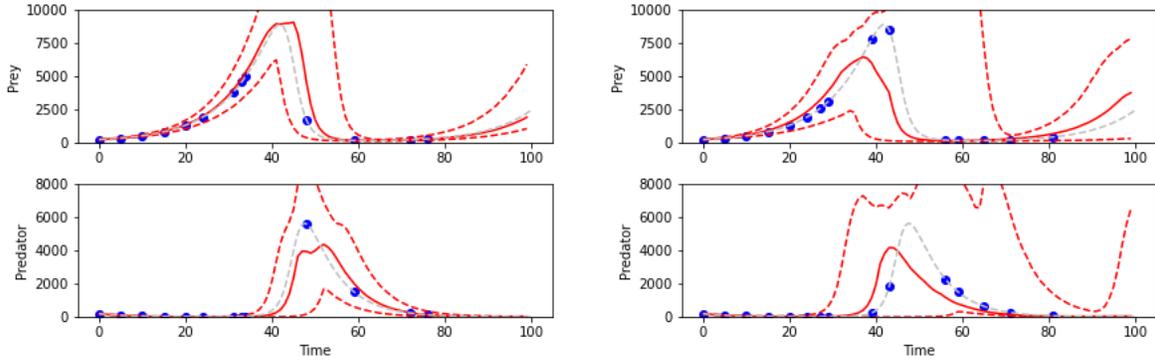
Fig. 26.: The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size ten. The panels plot the final designs using (a) I, (b) A, and (c) D optimality criteria.

The simulations in Figure 27 search windows of size fifteen to select the optimal design points. Table 9 compares the designs with the corresponding budget size, parameter estimates and stopping criterion. In these results, less design points were necessary to capture the dynamics due to the spacing of the points. However, it is clear that there is more variance in the decision boundaries of the final designs. Again, the I-optimal design requires the

least number of points with a final design of size thirteen. The A and D-optimal designs stop with designs of size fifteen and sixteen respectively.

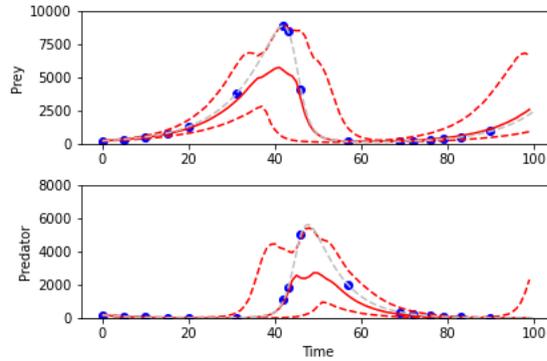
Design Criteria	b	Parameter Estimates				$S < V = 100$
I	13	$\alpha = 1.0172$	$\beta = 0.1108$	$\delta = 1.610545$	$\gamma = 0.6262$	$S = 50.83$
A	15	$\alpha = 1.0787$	$\beta = 0.092871$	$\delta = 1.5108$	$\gamma = 0.8741$	$S = 78.88$
D	16	$\alpha = 0.9697$	$\beta = 0.130643$	$\delta = 1.5508$	$\gamma = 0.7256$	$S = 24.31$

Table 9.: Lotka-Volterra Designs Exploring Windows of Size 15



(a)

(b)



(c)

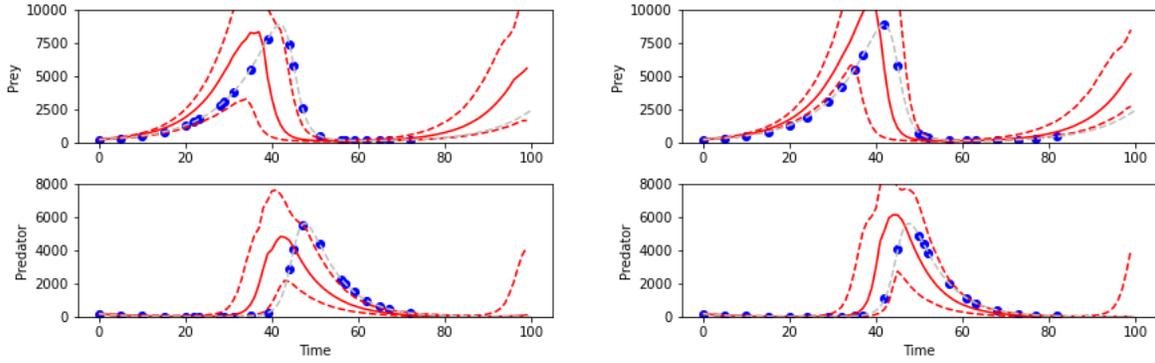
Fig. 27.: The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size fifteen. The panels plot the final designs using (a) I, (b) A, and (c) D optimality criteria.

The larger design windows represent limited resources for ecologists. However, the next set of simulations represent ample resources and exploring smaller windows of size five. Figure 28 illustrates the optimal designs for windows of size five, and Table 10 compares these designs with the corresponding budget size, parameter estimates and stopping criterion. As expected, smaller windows and larger budgets create less variability and closer parameter

estimates. Notice the I-optimal and A-optimal designs require twenty-three and twenty-two points. Whereas the D-optimal design requires nineteen to capture the dynamics.

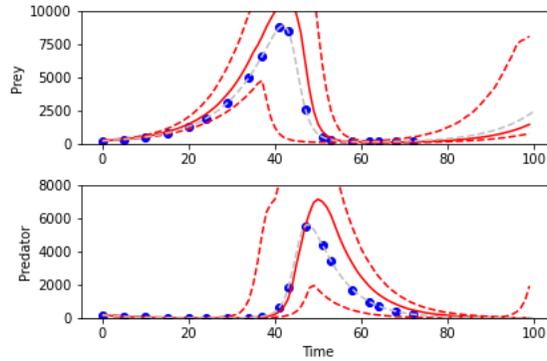
Design Criteria	b	Parameter Estimates				$S < V = 100$
I	23	$\alpha = 1.2028$	$\beta = 0.108392$	$\delta = 1.3875$	$\gamma = 0.6925$	$S = 22.42$
A	22	$\alpha = 1.1408$	$\beta = 0.09559$	$\delta = 1.5800$	$\gamma = 0.7285$	$S = 16.81$
D	19	$\alpha = 1.0492$	$\beta = 0.075292$	$\delta = 1.4101$	$\gamma = 0.7592$	$S = 17.66$

Table 10.: Lotka-Volterra Designs Exploring Windows of Size 5



(a)

(b)



(c)

Fig. 28.: The simulated predator and prey populations are plotted using grey dashed lines where $\alpha = 1.0$, $\beta = 0.1$, $\delta = 1.5$ and $\gamma = 0.75$. The design points are plotted in blue and fit by the 50% quantile of the predicted parameters represented by a solid red line. The 2.5% and 97.5% quantiles of the predicted values are plotted as red dashed lines. Each criteria evaluated candidate points in windows of size five. The panels plot the final designs using (a) I, (b) A, and (c) D optimality criteria.

4.5 Discussion

The purpose of this paper is to adapt the sequential optimality algorithm by creating a stopping criterion that can inform ecologists of the optimal budget sizes required to capture various population dynamics. The simulation results for the Logistic Equation and the Lotka-

Volterra differential equations represent various scenarios for ecologists to consider. First, the simple logistic growth model demonstrated a need to increase the budget size when dealing with slower growth rates while minimizing the budget when modeling faster growth rates. This information is beneficial for ecologists when assessing resource availability. The same can be said when examining more complex dynamics as the simulations of the predator-prey dynamics reveal that larger windows decrease the necessary samples required to capture the dynamics. Whereas, smaller windows create larger budgets with more accurate parameter estimates. In all cases, the developed algorithm is able to determine the optimal budget size, which ultimately helps with resource allocation and cost-efficiency for ecologists. As for comparing the criteria, the simulations show that the I optimality criterion consistently outperforms the other criteria in terms of efficiency. This is slightly expected given that it determines the optimal points based on the average prediction variance. However, again this is useful for ecologists when considering implementing this algorithm in practice to develop sampling designs.

CHAPTER 5

CLOSING REMARKS AND FUTURE WORK

Sequential optimality is developed in this dissertation as an algorithm intended to improve sampling design for ecologists. Temporal models were studied to represent ecological relationships and behaviors present in ecosystems. The logistic equation was first introduced as a simple model that can demonstrate the technique in a straightforward manner. The Lotka-Volterra differential equations were then incorporated to show the benefits of sequential optimality on more complex systems. The finalized algorithm in Chapter 4 determines optimal sampling regimes with a stopping criterion, which ultimately improves resource allocation and provides optimal budgets that can assist with efficiency when collecting data.

The univariate model of logistic growth was chosen for this research given its various applications and popularity in ecology. The design study modeled population dynamics and compared simulated annealing with the new approach of sequential optimality. It became clear that there are several ways to design sampling regimes for ecologists. However, sequential optimality proved to be more beneficial when predicting future design points. Given the main goal of predicting optimal designs, the results consistently found that the prediction based criterion outperformed the other criteria when selecting designs. This trend continued with the more complex dynamics. However, regardless of criteria the theoretical process sequentially learned about the systems and provides an advanced statistical method for ecological sampling rather than resorting to convenience.

Sequential optimality was then extended to more complex applications. Implementing the Lotka-Volterra model refined the method by modeling realistic environmental encounters.

Rather than simulating various predator-prey fluctuations, the purpose of these simulations was to further develop sequential optimality by studying the window sizes selected for the algorithm. Exploring larger windows of time represented limited resource availability and smaller windows implied ecologists had ample resources. Thus, results were provided across the various windows which compared designs predicted by the different criteria. Again, the simulations demonstrated that sequential optimality can predict the next optimal design point. However, the dynamics were not always captured in each scenario. Therefore, the algorithm was further developed to execute a stopping criterion.

The purpose of this research is to provide a statistical method that can inform ecologists of the optimal times to collect data. To optimize data collection expenses sequential optimality is finalized with a stopping criterion that can provide the budget size needed to capture the dynamics of a system. Both ecological models were implemented in this process, and again the results favored I optimality criterion. It appeared that the I-optimal designs required the least number of design points to capture the dynamics in each scenario. Though this is impressive, the finalized algorithm can successfully capture population dynamics as intended. Despite criteria, model or window size being explored, the complete algorithm can now inform ecologists of an optimal sampling regime that guarantees capturing the population dynamics.

Though temporal models are the focus of this research, it should be acknowledged that ecologists are interested in studying ecological communities with spatial components. Sampling does involve collecting data from various regions, and geographic information is commonly used in ecology to study shifts in population dynamics (Bascompte and Solé, 1995; Czárán, 1998; Fortin et al., 2014; Malchow et al., 2007). Sequential optimality has potential to further develop by being applied to spatiotemporal models. Optimization of monitoring stations using integer programming has been studied as seen in Hudak et al. (1995) and Lee and Deiningner (1992) which could lead to significant improvements to the sequential optimality method. It is always of interest to advance sampling capabilities by

optimizing sampling efforts (Warrick and Myers, 1987; Stein and Ettema, 2003). Thus, the purpose of this dissertation was achieved by developing a versatile statistical method that can improve sampling design in practice for ecologists.

Bibliography

- Abt, M., Welch, W.J., 1998. Fisher information and maximum-likelihood estimation of covariance parameters in gaussian stochastic processes. *Canadian Journal of Statistics* 26, 127–137.
- Aitchison, J., Brown, J., 1957. *The log-normal distribution*, 1957.
- Albert, C.H., Yoccoz, N.G., Edwards Jr, T.C., Graham, C.H., Zimmermann, N.E., Thuiller, W., 2010. Sampling in ecology and evolution—bridging the gap between theory and practice. *Ecography* 33, 1028–1037.
- Anscombe, F.J., 1961. Bayesian statistics. *The American Statistician* 15, 21–24.
- Atanga, R.E., Boone, E.L., Ghanam, R.A., Stewart-Koster, B., 2020. Optimal sampling regimes for estimating population dynamics. [arXiv:arXiv:2009.11797](https://arxiv.org/abs/2009.11797).
- Atkinson, A., Donev, A., Tobias, R., 2007. *Optimum experimental designs, with SAS*. volume 34. Oxford University Press.
- Bascompte, J., Solé, R.V., 1995. Rethinking complexity: modelling spatiotemporal dynamics in ecology. *Trends in Ecology & Evolution* 10, 361–366.
- Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical transactions of the Royal Society of London* , 370–418.
- Bohachevsky, I.O., Johnson, M.E., Stein, M.L., 1986. Generalized simulated annealing for function optimization. *Technometrics* 28, 209–217.

- Bourdeau, P.F., 1953. A test of random versus systematic ecological sampling. *Ecology* 34, 499–512.
- Bowser, C.J., 1986. Historic data sets: lessons from the past, lessons for the future. *Research data management in the ecological sciences* , 155–179.
- Box, G.E., Tiao, G.C., 1973. *Bayesian inference in statistical analysis*, Addison-Wesley. Reading, MA .
- Brooks, S., Gelman, A., Jones, G., Meng, X.L., 2011. *Handbook of Markov chain Monte Carlo*. CRC Press.
- Bulmer, M., 1974. On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics* , 101–110.
- Casella, G., George, E.I., 1992. Explaining the Gibbs sampler. *The American Statistician* 46, 167–174.
- Caughlan, L., Oakley, K.L., 2001. Cost considerations for long-term ecological monitoring. *Ecological Indicators* 1, 123–134.
- Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: A review. *Statistical Science* , 273–304.
- Chen, R., 1987. The relative efficiency of the sets and the cusum techniques in monitoring the occurrence of a rare event. *Statistics in Medicine* 6, 517–525.
- Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 327–335.
- Clyde, M.A., 2001. Experimental design: A Bayesian perspective. *International Encyclopedia of Social and Behavioral Sciences* .
- Cochran, W.G., 1977. *Sampling techniques*. 3rd ed. J. Wiley & Sons .

- Crawley, M.J., 2009. Natural enemies: the population biology of predators, parasites and diseases. John Wiley & Sons.
- Czárán, T., 1998. Spatiotemporal models of population and community dynamics. volume 21. Springer Science & Business Media.
- Dennis, B., Ponciano, J.M., Taper, M.L., 2010. Replicated sampling increases efficiency in monitoring biological populations. *Ecology* 91, 610–620.
- Eberhardt, L., Siniff, D., 1977. Population dynamics and marine mammal management policies. *Journal of the Fisheries Board of Canada* 34, 183–190.
- Eberhardt, L.L., 1977. Optimal policies for conservation of large mammals, with special reference to marine ecosystems. *Environmental Conservation* 4, 205–212.
- Elfving, G., et al., 1952. Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics* 23, 255–262.
- Ellison, A.M., 2004. Bayesian inference in ecology. *Ecology letters* 7, 509–520.
- Feller, W., 1940. On the logistic law of growth and its empirical verifications in biology. *Acta biotheoretica* 5, 51–66.
- Fisher, R.A., et al., 1960. The design of experiments. *The design of experiments*. .
- Ford, I., Titterton, D., Kitsos, C.P., 1989. Recent advances in nonlinear experimental design. *Technometrics* 31, 49–60x.
- Fortin, M.j., Dale, M.R., Ver Hoef, J.M., 2014. Spatial analysis in ecology. *Wiley StatsRef: Statistics Reference Online* .
- Gamito, S., 1998. Growth models and their use in ecological modelling: an application to a fish population. *Ecological modelling* 113, 83–94.

- Gard, T.C., Hallam, T.G., 1979. Persistence in food webs. *lotka-volterra food chains*. *Bulletin of Mathematical Biology* 41, 877–891.
- Gatabazi, P., Mba, J., Pindza, E., Labuschagne, C., 2019. Grey lotka–volterra models with application to cryptocurrencies adoption. *Chaos, Solitons & Fractals* 122, 47–57.
- Gelfand, A.E., 2000. Gibbs sampling. *Journal of the American statistical Association* 95, 1300–1304.
- Gelfand, A.E., Smith, A.F., 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association* 85, 398–409.
- Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence* , 721–741.
- Gilks, W.R., Richardson, S., Spiegelhalter, D., 1995. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.
- Girolami, M., 2008. Bayesian inference for differential equations. *Theoretical Computer Science* 408, 4–16.
- Goffe, W.L., Ferrier, G.D., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. *Journal of econometrics* 60, 65–99.
- Goos, P., Jones, B., 2011. *Optimal design of experiments: a case study approach*. John Wiley & Sons.
- Green, R.H., 1979. *Sampling design and statistical methods for environmental biologists*. John Wiley & Sons.
- Greig-Smith, P., 1983. *Quantitative plant ecology*. volume 9. Univ of California Press.

- Hale, J.K., Koçak, H., 2012. Dynamics and bifurcations. volume 3. Springer Science & Business Media.
- Hardin, R., Sloane, N., 1993. A new approach to the construction of optimal designs. *Journal of statistical planning and inference* 37, 339–369.
- Harter, H.L., Moore, A.H., 1965. Maximum-likelihood estimation of the parameters of gamma and weibull populations from complete and from censored samples. *Technometrics* 7, 639–643.
- Hastings, W.K., 1970. Monte carlo sampling methods using markov chains and their applications .
- Hedge, J., Lycett, S.J., Rambaut, A., 2013. Real-time characterization of the molecular epidemiology of an influenza pandemic. *Biology letters* 9, 20130331.
- Henderson, P.A., 2001. Ecological methods. e LS .
- Hening, A., Nguyen, D.H., 2018a. Persistence in stochastic lotka–volterra food chains with intraspecific competition. *Bulletin of mathematical biology* 80, 2527–2560.
- Hening, A., Nguyen, D.H., 2018b. Stochastic lotka–volterra food chains. *Journal of mathematical biology* 77, 135–163.
- Heydari, J., Lawless, C., Lydall, D.A., Wilkinson, D.J., 2014. Fast bayesian parameter estimation for stochastic logistic growth models. *Biosystems* 122, 55–72.
- Hilborn, R., 1990. Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. *Canadian Journal of Fisheries and Aquatic Sciences* 47, 635–643.
- Hitchcock, D.B., 2003. A history of the metropolis–hastings algorithm. *The American Statistician* 57, 254–257.

- Hobbs, N.T., Hooten, M.B., 2015. Bayesian models: a statistical primer for ecologists. Princeton University Press.
- Hooten, M.B., Hobbs, N.T., 2015. A guide to bayesian model selection for ecologists. Ecological Monographs 85, 3–28.
- Hotelling, H., 1944. Some improvements in weighing and other experimental techniques. The Annals of Mathematical Statistics 15, 297–306.
- Houlihan, J.E., McKinney, S.T., Anderson, T.M., McGill, B.J., 2017. The priority of prediction in ecological understanding. Oikos 126, 1–7.
- Hsu, F., Nelson, C., Chow, W., 1984. A mathematical model to utilize the logistic function in germination and seedling growth. Journal of Experimental Botany 35, 1629–1640.
- Huang, Z., Liu, S., Bradford, K.J., Huxman, T.E., Venable, D.L., 2016. The contribution of germination functional traits to population dynamics of a desert plant community. Ecology 97, 250–261.
- Hubbert, M.K., et al., 1956. Nuclear energy and the fossil fuel, in: Drilling and production practice, American Petroleum Institute.
- Hudak, P.F., Loaiciga, H.A., Marino, M.A., 1995. Regional-scale ground water quality monitoring via integer programming. Journal of Hydrology 164, 153–170.
- Hutchinson, G.E., 1978. An introduction to population ecology. 504: 51 HUT.
- Johnson, N.L., 1970. L. & kotz, s.(1970). continuous univariate distributions-1.
- Jones, M., Goldstein, M., Jonathan, P., Randell, D., et al., 2018. Bayes linear analysis of risks in sequential optimal design problems. Electronic Journal of Statistics 12, 4002–4031.
- Kennard, M., Arthington, A., Pusey, B., Harch, B., 2005. Are alien fish a reliable indicator of river health? Freshwater Biology 50, 174–193.

- Kiefer, J., 1959. Optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological)* 21, 272–304.
- Kiefer, J., Wolfowitz, J., 1959. Optimum designs in regression problems. *The Annals of Mathematical Statistics* , 271–294.
- Kiefer, J., et al., 1958. On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *The Annals of Mathematical Statistics* 29, 675–699.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *science* 220, 671–680.
- Kutta, W., 1901. Beitrag zur naherungsweise integration totaler differentialgleichungen. *Z. Math. Phys.* 46, 435–453.
- Laarhoven, P.J., Aarts, E.H., 1987. Simulated annealing, in: *Simulated annealing: Theory and applications*. Springer, pp. 7–15.
- Lee, B.H., Deininger, R.A., 1992. Optimal locations of monitoring stations in water distribution system. *Journal of Environmental Engineering* 118, 4–16.
- Leigh, C., Hiep, L.H., Stewart-Koster, B., Vien, D.M., Condon, J., Sang, N.V., Sammut, J., Burford, M.A., 2017. Concurrent rice-shrimp-crab farming systems in the mekong delta: Are conditions (sub) optimal for crop production and survival? *Aquaculture research* 48, 5251–5262.
- Lima, S.L., 1998. Nonlethal effects in the ecology of predator-prey interactions. *Bioscience* 48, 25–34.
- Lin, B.q., Liu, J.h., 2010. Estimating coal production peak and trends of coal imports in china. *Energy Policy* 38, 512–519.
- Lotka, A.J., 1926. *Elements of physical biology*. *Science Progress in the Twentieth Century (1919-1933)* 21, 341–343.

- Lotka, A.J., 1978. The growth of mixed populations: two species competing for a common food supply, in: *The Golden Age of Theoretical Ecology: 1923–1940*. Springer, pp. 274–286.
- Malchow, H., Petrovskii, S.V., Venturino, E., 2007. *Spatiotemporal patterns in ecology and epidemiology: theory, models, and simulation*. Chapman and Hall/CRC.
- Marasco, A., Picucci, A., Romano, A., 2016. Market share dynamics using lotka–volterra models. *Technological forecasting and social change* 105, 49–62.
- Marshall, M.N., 1996. Sampling for qualitative research. *Family practice* 13, 522–526.
- Martino, L., Elvira, V., 2014. Metropolis sampling. *Wiley StatsRef: Statistics Reference Online* , 1–18.
- McAllister, M.K., Peterman, R.M., 1992. Experimental design in the management of fisheries: a review. *North American Journal of Fisheries Management* 12, 1–18.
- McCallum, H., 2008. *Population parameters: estimation for ecological models*. volume 3. John Wiley & Sons.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. *The journal of chemical physics* 21, 1087–1092.
- Mode, N.A., Conquest, L.L., Marker, D.A., 1999. Ranked set sampling for ecological research: accounting for the total costs of sampling. *Environmetrics: The official journal of the International Environmetrics Society* 10, 179–194.
- Montgomery, D.C., 2017. *Design and analysis of experiments*. John wiley & sons.
- Müller, P., Berry, D.A., Grieve, A.P., Smith, M., Krams, M., 2007. Simulation-based sequential bayesian design. *Journal of statistical planning and inference* 137, 3140–3150.

- Murray, J.D., 2002. *Mathematical biology: I. an introduction. interdisciplinary applied mathematics.* Mathematical Biology, Springer .
- Oliver, F., 1964. Methods of estimating the logistic growth function. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 13, 57–66.
- Pagendam, D., Pollett, P., 2009. Optimal sampling and problematic likelihood functions in a simple population model. *Environmental Modeling & Assessment* 14, 759–767.
- Parker, M., Kamenev, A., 2009. Extinction in the lotka-volterra model. *Physical Review E* 80, 021129.
- Pearl, R., 1925. *The biology of population growth.* AA Knopf.
- R Core Team, 2013. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. URL: <http://www.R-project.org/>.
- Reed, L.J., Pearl, R., 1927. On the summation of logistic curves. *Journal of the Royal Statistical Society* 90, 729–746.
- Robert, C., Casella, G., 2013. *Monte Carlo statistical methods.* Springer Science & Business Media.
- Runge, C., 1895. Über die numerische auflösung von differentialgleichungen. *Mathematische Annalen* 46, 167–178.
- Schweiger, A.H., Irl, S.D., Steinbauer, M.J., Dengler, J., Beierkuhnlein, C., 2016. Optimizing sampling approaches along ecological gradients. *Methods in Ecology and Evolution* 7, 463–471.
- Smith, A.F., Roberts, G.O., 1993. Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)* 55, 3–23.

- Stein, A., Ettema, C., 2003. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agriculture, Ecosystems & Environment* 94, 31–47.
- Strogatz, S.H., 2018. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC Press.
- Tsoularis, A., Wallace, J., 2002. Analysis of logistic growth models. *Mathematical biosciences* 179, 21–55.
- Van Rossum, G., Drake Jr, F.L., 1995. *Python tutorial*. volume 620. Centrum voor Wiskunde en Informatica Amsterdam.
- Verdinelli, I., 1992. Advances in bayesian experimental design. *Bayesian Statistics* 4, 467–481.
- Verhulst, P.F., 1838. Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.* 10, 113–126.
- Volterra, V., 1928. Variations and fluctuations of the number of individuals in animal species living together. *ICES Journal of Marine Science* 3, 3–51.
- Wakefield, J., Smith, A., Racine-Poon, A., Gelfand, A.E., 1994. Bayesian analysis of linear and non-linear population models by using the gibbs sampler. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 43, 201–221.
- Wald, A., 1943. On the efficient design of statistical investigations. *The annals of mathematical statistics* 14, 134–140.
- Warrick, A., Myers, D., 1987. Optimization of sampling locations for variogram calculations. *Water Resources Research* 23, 496–500.
- Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. *Analysis and management of animal populations*. Academic Press.

- Williams, P.J., Hooten, M.B., Womble, J.N., Esslinger, G.G., Bower, M.R., 2018. Monitoring dynamic spatio-temporal ecological processes optimally. *Ecology* 99, 524–535.
- Williamson, D., Goldstein, M., 2012. Bayesian policy support for adaptive strategies using computer models for complex physical systems. *Journal of the Operational Research Society* 63, 1021–1033.
- Zhang, G., McAdams, D.A., Shankar, V., Mohammadi Darani, M., 2018a. Technology evolution prediction using lotka–volterra equations. *Journal of Mechanical Design* 140.
- Zhang, J.F., Papanikolaou, N.E., Kypraios, T., Drovandi, C.C., 2018b. Optimal experimental design for predator–prey functional response experiments. *Journal of The Royal Society Interface* 15, 20180186.

VITA

Rebecca Elise Bergee was born on May 31, 1993 in Atlanta, Georgia. She moved multiple times throughout her childhood and attended three high schools ultimately graduating in 2011 from Steel Valley Senior High School, Munhall, Pennsylvania. She received her Bachelor of Science degree in Applied Mathematics with a minor in Statistics in 2014 from University of Pittsburgh at Greensburg, Greensburg, Pennsylvania. While attending college, she met her advisor Dr. Ryad Ghanam who encouraged her to study mathematics and supported her passion for volunteering. After attending a community oriented school and volunteering with Habitat for Humanity, Rebecca chose to serve in the Peace Corps as a junior high school mathematics teacher in Ghana, West Africa in the rural village of Kokofu-Mensase. She met her husband, Prosper Atanga, while in service and was married in 2020 changing her name to Rebecca Elise Atanga. Her teaching experience abroad changed her perspective about obtaining an advanced degree and she consulted Dr. Ghanam to discuss future options. Dr. Ghanam introduced Rebecca to the Doctor of Philosophy in Systems Modeling and Analysis program at Virginia Commonwealth University, Richmond, Virginia. While studying in the interdisciplinary program, Dr. Ghanam introduced Rebecca to her advisor Dr. Edward L. Boone who suggested she work on a research project with ecologists at the Australian Rivers Institute at Griffith University in Australia. Beyond research, Rebecca grew as a statistician by interning at NASA Langley Research Center as well as the Virginia Department of Transportation. She also worked as a statistical consultant in the Department of Statistical Sciences and Operations Research Lab and was encouraged to pursue a Master of Science degree by her advisors. She received a Master of Science in Mathematical Sciences with a concentration in Statistics in 2020 from Virginia Commonwealth University, Richmond, Virginia. Throughout her studies, Rebecca continued her passion for teaching by working as an instructor for the Department of Statistical Sciences and Operations Research and the Department of Applied Mathematics and Mathematics at Virginia Commonwealth University, Richmond, Virginia. A version of Chapter 2 (Optimal Sampling Regimes for Estimating

Population Dynamics) of this dissertation has been published by the Multidisciplinary Digital Publishing Institute-Stats. Rebecca and her co-authors have also submitted a version of Chapter 3 (Optimal Sampling Regimes for Estimating Predator-Prey Dynamics) to the Statistical Methods and Applications Journal. Though not yet known, Rebecca is expected to submit Chapter 4 to a peer-reviewed journal.