Spring 2020

Understanding the Human Gut Microbiota: A Mathematical Approach

Melissa Adrian

UNDERSTANDING THE HUMAN GUT MICROBIOTA: A MATHEMATICAL APPROACH

by

Melissa Adrian

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in the Mathematics

_____

Bruce Ayati
Thesis Mentor

Spring 2020

All requirements for graduation with Honors in the
Mathematics have been completed.

_____

Oguz Durumeric
Mathematics Honors Advisor

UNDERSTANDING THE HUMAN GUT MICROBIOTA: A MATHEMATICAL APPROACH

by

Melissa A. Adrian

A thesis submitted in partial fulfillment
of the requirements for the
Honors Bachelor of Science
degree in Mathematics in the
College of Liberal Arts and Sciences of
The University of Iowa

May 2020

Thesis Supervisor: Professor Bruce P. Ayati

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Bruce Ayati for his mentorship over the past three years. His guidance was not limited to the scope of our project; he has given me invaluable advice for both my life and my career. Thank you for dedicating countless hours to working on this project with me, an undergraduate student.

I would like to extend gratitude to others who have helped me in any way during the process of completing this thesis, such as Dr. Colleen Mitchell, Dr. Ruqiah Muhammad, Kasra Zarei, and Dr. Ashutosh Mangalam. Additionally, thank you to my friends Ross, Nicole, and Kyle for answering any and all of the basic biology and chemistry questions that I had along the way.

I would like to thank the Iowa Center for Research by Undergraduates for the opportunity to present this work at the 2019 Spring Undergraduate Research Festival, and the University of Iowa's Department of Mathematics for providing funding for my poster.

In addition to acknowledging those who have helped me in my academic pursuits, there are many people in my personal life who deserve recognition for their support, such as my close friends Abby, Emily, Claire, Rikki, Ross, Nicole, Kyle, including many others, and, of course, my family.

Thank you, Drew, for your unwavering love, support, and encouragement over the past four years, even from 2,000 miles away. My days have been brighter since they've been with you.

These acknowledgements would be incomplete without recognizing the instrumental role my parents and my brother Patrick have played in my life and in the trajectory of my career, especially during my time as an undergraduate. Without Patrick's initiative of becoming involved in research while at the University of Iowa, I would have never thought to pursue an opportunity like this. Lastly, thank you to my parents for always believing in me, even when I doubted myself.

ABSTRACT

In order to explore the dynamics of the human gut microbiota, we used a system of ordinary differential equations to mathematically model the biomass of three microorganism populations: *Bacteroides thetaiotaomicron*, *Eubacterium rectale*, and *Methanobrevibacter smithii*. Additionally, we modeled the concentrations of relevant nutrients necessary to sustain these populations over time. This system highlights the interactions and the competition among these species in order to further understand their dynamics. These three microorganisms were specifically chosen due to the system's end product, butyrate, which aids in developing the intestinal barrier in the human gut. The basis of the mathematical model assumes the gut acts as a chemostat, with bacteria and nutrients exiting the gut at a rate proportional to the volume of the chemostat, the rate of volumetric flow, and the biomass or concentration of the particular population or nutrient. We performed global sensitivity analysis using Sobol' sensitivities in order to estimate the importance of model parameters and to understand our results.

PUBLIC ABSTRACT

Mathematical modeling represents complex systems in a concise way and can highlight its interesting aspects. Using a system of ordinary differential equations, which is a collection of equations that describe how fast or slow a quantity is changing in a given time frame, we can track the quantities of interest over time. In this case, we track the fluctuations of different populations of microorganisms in the human gut and the nutrients necessary for their survival over time. The mass of these populations depends on a number of factors, such as how fast they are flushed out of the gut, competition with other microorganisms for resources, and how fast they grow. All of these factors are accounted for in the equations of our mathematical model. As we become increasingly aware that the bacteria in the gut impacts on person's overall health, it becomes increasingly important to understand the mechanisms behind their interactions. Mathematical modeling is a quantitative approach to understanding this complex ecosystem, which we implement through a system of ordinary differential equations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The human gut microbiota, which is the collection of microorganisms located in the stomach, large intestines, and small intestines, plays an important role in sustaining overall human health. Each individual's gut composition is unique, and long-term dietary patterns, specifically concerning the types and amounts of carbohydrates, proteins, and fats consumed, affect its composition (David, Maurice, Carmody, et al., 2014; Ji and Nielsen, 2015). The gut microbiota responds rapidly to changes in diet, and this response potentially serves as an explanation for the vast diversity of human diets seen across regions and cultures (David, Maurice, Carmody, et al., 2014). Based on the study conducted in David, Maurice, Carmody, et al., 2014 at Harvard University, rapidly switching between carnivorous and herbivorous diets can cause significant changes to the human gut microbiota, affecting the rates of carbohydrate and amino acid fermentation within the gut (David, Maurice, Carmody, et al., 2014). The ability of the human gut microbiota to alter rapidly based on the available sources of nutrients could be a result of selective pressures during human evolution (David, Maurice, Carmody, et al., 2014). Because animal foods may have been a more volatile source of nutrients, the ability for a human to rapidly switch between animal- and plant-based food sources depending on their hunting success became advantageous for survival (David, Maurice, Carmody, et al., 2014). The flexibility of the human diet serves as an explanation as to how people can thrive on the variety of modern diets seen across regions, cultures, and lifestyles (David, Maurice, Carmody, et al., 2014).

In addition to the need to understand the observed diversity of the human gut microbiota, this ecosystem is especially a focus of study in disease prevention and treatment because the gut is the only place in the body where nutrients are absorbed, and its composition can exacerbate chronic illnesses, such as Crohn's disease. Due to the high variability among people of gut compositions, however, the relation between the gut microbiota composition and its effects on overall health and diseases has not yet been made clear by current research efforts. Consequently, a better understanding of the microbial dynamics may elucidate their effects on diet-related diseases

(David, Maurice, Carmody, et al., 2014).

Motivating Examples

Though the changes in the bacterial composition of the human gut microbiota have been shown to be associated with changes in metabolism and overall health, the underlying interactions among species within this microbiota are not yet well understood (Ji and Nielsen, 2015). To illustrate the impact that the gut microbiota can have on human health and the need to improve our understanding of this ecosystem, consider two motivating examples. Based on an experiment conducted on mice at Cork Cancer Research Center, *Escherichia coli* and *Listeria welshimeri* were shown to independently alter, either positively or negatively, the effects of several chemotherapeutic drugs through enzymatic activity (Lehouritis, Cummins, Stanton, et al., 2015). As a further extension of this research, other studies conducted on both humans and mice have shown that the composition of the gut microbiota can affect the responses to other types of administered drugs by influencing the host's immune system (Guglielmi, 2018; Lehouritis, Cummins, Stanton, et al., 2015). As an optimistic potential further direction of these findings, the composition of a patient's microbiota could be manipulated in such a way that a drug's positive effects are enhanced (Guglielmi, 2018). However, because there exists a delicate equilibrium in the microbiota, altering an individual's gut microbiota might create an increased susceptibility to other unanticipated health problems (Guglielmi, 2018). This uncertainty stems from a collective lack of understanding of interspecies microbial interactions, which necessitates further investigation.

To consider another example, some microorganisms are associated with the initiation and further cultivation of cancer cells, and about 10-20% of human cancers can be linked to these microbes (Guglielmi, 2018; Whisner and Athena Aktipis, 2019). Specifically, *Helicobacter pylori* has been associated with stomach cancer, and a list of nine other carcinogenic microorganisms have been identified by the International Agency for Cancer Research (Whisner and Athena Aktipis, 2019). These bacteria are able to facilitate cancer growth by creating an ideal environment for tumors to grow through various mechanisms, such as causing cells to develop a resistance to

anticancer drugs, inducing inflammation, or through other means (Guglielmi, 2018; Lehouritis, Cummins, Stanton, et al., 2015). Though these identified carcinogenic microorganisms live in the guts of a large proportion of the human population, many individuals never develop the cancers linked to these microorganisms (Whisner and Athena Aktipis, 2019). A question that emerges as a result of this observed behavior is: does the presence of specific surrounding substances or other microorganisms trigger these carcinogenic microorganisms to begin, by some means, facilitating cancer growth? Due to the seemingly vast impact that gut microorganisms can have on human health as illustrated in the two preceding examples, this ecosystem calls for further exploration through various approaches (Shoaie, Karlsson, Mardinoglu, et al., 2013).

Common Methods of Analysis

Common approaches to analyzing the human gut microbiota that provide an enhanced understanding of the diversity and composition of this ecosystem include, but are not limited to, next-generation sequencing, metatranscriptomics, culturomics, and mass spectrometry analyses (Ji and Nielsen, 2015; Kumar, Ji, Zengler, and Nielsen, 2019). However, these approaches give little insight as to how species interact amongst themselves and with the human host (Ji and Nielsen, 2015). Mathematical modeling, on the other hand, attempts to answer such questions and can act as a supplement to the knowledge gained from the aforementioned approaches. To name a few of its key features and uses, mathematical models highlight specific details of interest in the system, provide a representation of reality, and reflect the goals of the research, which in a sense makes these models subjective (Ayati, 2019). Despite the subjective nature of these models when used to understand underlying truths, which might appear to be a weakness of this approach, mathematical models can nonetheless provide evidence to strengthen a hypothesis (Oreskes, Shrader-Frechette, and Belitz, 1994) and offer guiding principles for further study of a phenomenon (Oreskes, Shrader-Frechette, and Belitz, 1994), both of which are of much use to research.

One specific type of mathematical model used to understand biological mechanisms is a genomic-scale metabolic model (GEM). GEMs represent the collection of biochemical reactions

that take place within a microorganism, which forms a metabolic network. These models incorporate genome-level information of a microorganism and provide a representation of the relationship between genotypes and phenotypes. GEMs additionally include information about the enzymes and genes associated with each biochemical reaction, thus providing a connection from gene to protein to reaction to metabolite. Incorporating metabolic data into the GEM model can elicit more specific predictions, such as the gene's effect on a given pathway's functionality and on the metabolic network as a whole. However, one main challenge of implementing GEMs is the lack of information on the strain-level information contained in metagenomic data (Kumar, Ji, Zengler, and Nielsen, 2019).

Agent-based modeling (ABM) is another computational modeling strategy used to simulate unique biological systems and to test the interactions among different parameters, such as the rate of fermentation of specific microorganisms. For the purposes of studying the human gut microbiota, this strategy can also be utilized to further understand the interactions among gut microorganisms. Additionally, ABMs are useful in studying the role of spatial dynamics within a system in tandem with the variations in environmental conditions. Despite this modeling technique's practical use, two of its main drawbacks are the computational intensity required to compute ABMs and the difficulty of determining the effects of the different parameters on the ABM model. This difficulty stems from the fact that ABM models are able to generalize behavior for the whole population, but fail to distinguish the effects of individual bacterium (Kumar, Ji, Zengler, and Nielsen, 2019).

With these common methods of analysis in mind, we chose ordinary differential equation (ODE)-based modeling as the main tool for our analysis, which tracks information about biomass and concentration levels over time rather than genomic information as in the case of GEMs. In this so-called dynamical system approach, we can identify the system's driving parameters and analyze its stability. Dynamical systems models, in the context of the gut microbiota, describe production and consumption trends for specified gut microorganisms and their metabolites over time (Kumar, Ji, Zengler, and Nielsen, 2019). Despite the advantage of having easily interpretable terms in the

model's ODEs, this approach still has some key drawbacks, one of which being the determination of unknown parameters.

In general, once a small-scale model is written and validated through experiments, the model can be expanded to account for additional interactions. However, in the process of expanding the model, the ODE system can become increasingly complex due to the addition of numerous unknown parameters. The parameters used in these models in a biological context include substrate conversion rates, utilization rates, and cellular death rates, which typically can be determined and validated by laboratory experiments. Yet, some gut microorganisms are unable to be cultivated in a laboratory setting, leaving specific information about these species' metabolic activity unknown. Due to the limitations of determining a large number of unknown parameters in more complex ODE models, ODE models are limited to tracking a few species within a community, which is one key drawback (Kumar, Ji, Zengler, and Nielsen, 2019).

Despite the practicality constraints on extending ODE-based modeling to a larger scale, ODE models paired with other types of modeling, such as agent-based modeling, can provide an insightful understanding of the gut microbiota's dynamics and interactions (Kumar, Ji, Zengler, and Nielsen, 2019). As a first step in our modeling efforts, with its limitations in mind, we consider the ODE-based modeling approach for a small-scale system of three abundant microorganisms in the human gut microbiota.

Small-scale Representation of The Human Gut Microbiota

In order to explore this complex ecosystem using the ODE-based modeling approach as a preliminary step, the three microorganisms *Bacteroides thetaiotaomicron*, *Eubacterium rectale*, and *Methanobrevibacter smithii* were chosen to be the focus of our model. These species represent the three main phyla in the human gut: Bacteroidetes, accounting for 17-60% of the total biomass; Firmicutes, 35-80% of the biomass; and Euryarchaeota (Shoaie, Karlsson, Mardinoglu, et al., 2013). The system's main product, butyrate, is of specific interest due to its impact in sustaining human health. Butyrate provides energy to colonocytes, affects overall energy homeostasis, and

inhibits histone deacetylase, which is an enzyme that directly affects colorectal cancer (Shoaie, Karlsson, Mardinoglu, et al., 2013). Along with butyrate, other notable intermediates and products part of this system include acetate, propionate, glutamine, carbon dioxide (CO_2), hydrogen (H_2), and methane (CH_4), which affect health and well-being in various ways. Acetate, propionate, and butyrate, which are short chain fatty acids, are absorbed in the gut's epithelial cells, thus regulating an individual's immune system and metabolism (Shoaie, Karlsson, Mardinoglu, et al., 2013).

Individually, acetate acts as a substrate for cholesterol synthesis and lipogenesis (Shoaie, Karlsson, Mardinoglu, et al., 2013); propionate regulates gluconeogenesis and cholesterol synthesis (Shoaie, Karlsson, Mardinoglu, et al., 2013); glutamine fuels the metabolism and maintains the intestinal barrier (Kim and Kim, 2017); and the gases CO_2, H_2, and CH_4 are products of bacterial fermentation that can cause intestinal discomfort when in excess (Scaldaferri, Nardone, Lopetuso, et al., 2013). As for the microorganisms themselves, *M. smithii* plays a significant role by removing hydrogen gas, which affects bacterial fermentation and energy gathering, and by producing methane gas (Shoaie, Karlsson, Mardinoglu, et al., 2013); *E. rectale* produces butyrate, which is beneficial to the gut's epithelial cells (Shoaie, Karlsson, Mardinoglu, et al., 2013); and *B. thetaiotaomicron* utilizes dietary polysaccharides and indirectly facilitates butyrate production with its outputs (Ji and Nielsen, 2015).

In developing an ODE model to represent this small-scale system, we assume that the human gut acts as a chemostat, which is a novel approach and an assumption not implemented, to our knowledge, in other work. In this thesis, we utilize ODE-based dynamical systems modeling to track the changes in biomass of the three prevalent microorganisms, *B. thetaiotaomicron*, *E. rectale*, and *M. smithii*, as well as their chemical inputs, intermediates, and byproducts, with the goal of providing a better understanding of their interactions within this small-scale system.

GRAPHICAL AND MATHEMATICAL REPRESENTATION OF THE THREE SPECIES

In order to study the interactions among the three microorganisms *B. thetaiotaomicron*, *E. rectale*, and *M. smithii*, we first identified from the available literature the key metabolites of these microorganisms that emphasize their interactions. Upon deciding the metabolites of interest, we developed a graphical representation of this system to serve as a visual guide before creating our ODE-based mathematical model. In this chapter, we present and explain our mathematical formulation of this small-scale system.

Isolation of Key Factors And Graphical Representation Into a Schematic

Through GEMs and analysis of transcriptomics data of the three species *B. thetaiotaomicron*, *E. rectale*, and *M. smithii*, many of the key underlying characteristics of each microorganism and their interactions were revealed. GEMs utilize genomic data to represent cellular metabolism (Ji and Nielsen, 2015), and transcriptomics analysis allows for the study of the functions of the genes that an organism expresses (Lowe, Shirley, Bleackley, Dolan, and Shafee, 2017). Both of these methods of study combined provide insight into the microbial interactions within this subset of species in the human gut microbiome. In order to supplement the knowledge gained from transcriptomic analysis and GEMs, we incorporate this information into our own mathematical interpretation of this small-scale system using ODEs. First, we present information learned through previous investigations of *B. thetaiotaomicron*, *E. rectale*, and *M. smithii* that we utilized in creating our mathematical model.

B. thetaiotaomicron is an abundant bacterial species in the human gut microbiome whose main function is the utilization of polysaccharides (Xu, Bjursell, Himrod, et al., 2003, Ji and Nielsen, 2015). Through polysaccharide degradation, *B. thetaiotaomicron* contributes to the overall ecosystem diversity in the colon, which is its regular environment (Adamberg, Tomson, Vija, et al., 2014). *B. thetaiotaomicron* can survive solely on the uptake of carbon-rich polysaccharides (Martens, Lowe, Chiang, et al., 2011); however, its growth is enhanced in the presence of inorganic

ammonia due to inorganic ammonia's contribution of nitrogen to *B. thetaiotaomicron*'s metabolism (Glass and Hylemon, 1980). Through the utilization of inorganic ammonia, *B. thetaiotaomicron* can synthesize all amino acids that are essential to human health (Ji and Nielsen, 2015), which makes this bacteria a key focus of study.

The two bacterial species *B. thetaiotaomicron* and *E. rectale* exhibit complex interactions due to their changes in gene regulation in the presence of each other. When both bacteria are present in an environment, *B. thetaiotaomicron* up-regulates gene expression for starch utilization and the degradation of specific glycans that *E. rectale* is unable to utilize. Simultaneously, *E. rectale* down-regulates the genes associated with glycan degradation despite the fact that it cannot grow efficiently without a carbohydrate source. Previous research on the interactions of these two species suggests that the presence of *B. thetaiotaomicron* enhances *E. rectale*'s ability to uptake nutrients. Despite the competition for a carbohydrate source between B. thetaiotaomicron and *E. rectale*, these species are still able to coexist in the human gut (Mahowald, Rey, Seedorf, et al., 2009).

M. smithii, which is one of the main methanogenic archaeon in the human gut, improves the productivity of carbohydrate metabolism by utilizing H_2 from *E. rectale* and formate from *B. thetaiotaomicron* to produce methane gas. This process prevents the environment from becoming too saturated with *B. thetaiotaomicon* and *E. rectale*'s by-products, which consequently improves carbohydrate metabolism. Additionally, *M. smithii* removing H_2 in this environment allows for *B. thetaiotaomicron* to generate NAD^+, which is used for glycolysis, a fundamental process in producing cellular energy. Because both *B. thetaiotaomicron* and *M. smithii* benefit one another, these two microorganisms exhibit a mutualistic relationship, further supporting the idea that these two microorganisms can coexist in the gut (Mahowald, Rey, Seedorf, et al., 2009).

To summarize the key exchanges and interactions of the three species *B. thetaiotaomicron*, *E. rectale*, and *M. smithii*: *B. thetaiotaomicron* converts polysaccharides to acetate, which is subsequently utilized by both *E. rectale* and *M. smithii* (Shoaie, Karlsson, Mardinoglu, et al., 2013; Ji and Nielsen, 2015; Mahowald, Rey, Seedorf, et al., 2009); *E. rectale* uptakes polysaccharides,

except in the presence of *B. thetaiotaomicron*; *B. thetaioaomicron* synthesizes amino acids from inorganic ammonia, specifically glutamine, which are then utilized by *E. rectale* (Ji and Nielsen, 2015); the substrates produced from *B. thetaiotaomicron* breaking down polysaccharides is ultimately converted into CO_2 and H_2 (Adamberg, Tomson, Vija, et al., 2014); and *E. rectale* converts acetate to CO_2 and H_2, which is utilized by *M. smithii* (Shoaie, Karlsson, Mardinoglu, et al., 2013; Ji and Nielsen, 2015). Two end products of the system are butyrate, produced by *E. rectale*, and methane, produced by *M. smithii* (Shoaie, Karlsson, Mardinoglu, et al., 2013; Ji and Nielsen, 2015).

Though the metabolic pathway of *B. thetaiotaomicron* converting inorganic ammonia to amino acids enhances the rate at which this bacteria uptakes polysaccharides, we decided to exclude this pathway from our analysis, under the assumption that amino acids will be sufficiently present in the gut over time in order maintain the polysaccharides utilization rate at a mean value. This assumption can only be made for individuals with a healthy gut microbiota. An implementation of this model considering this pathway for individuals with a deficiency in amino acids can be considered in the future. While there are many more pathways to consider involving these microorganisms and their interactions with others in the system, we narrowed our focus to this subsystem, which can be further expanded in the future.

The information known about these three species' interactions was translated into our own graphical representation, shown in Figure 1. This schematic highlights the competition among these species, specifically between *E. rectale* and *M. smithii* and between *B. thetaiotaomicron* and *E. rectale*. Generally, multiple species within a community can grow and compete for a single supply of carbon, which suggests that the genes required to metabolize carbon are present in many members of the community, rather than a single best competitor (Goldford, Lu, Bajie, et al., 2018). A hypothesis presented in Goldford, Lu, Bajie, et al., 2018 regarding this phenomenon suggests that competing species can coexist in the gut microbiome because these microbes tend to secrete substances that are subsequently utilized by other competing microbes, which is the case for *M. smithii* and *E. rectale* (Goldford, Lu, Bajie, et al., 2018). Though these two species compete for

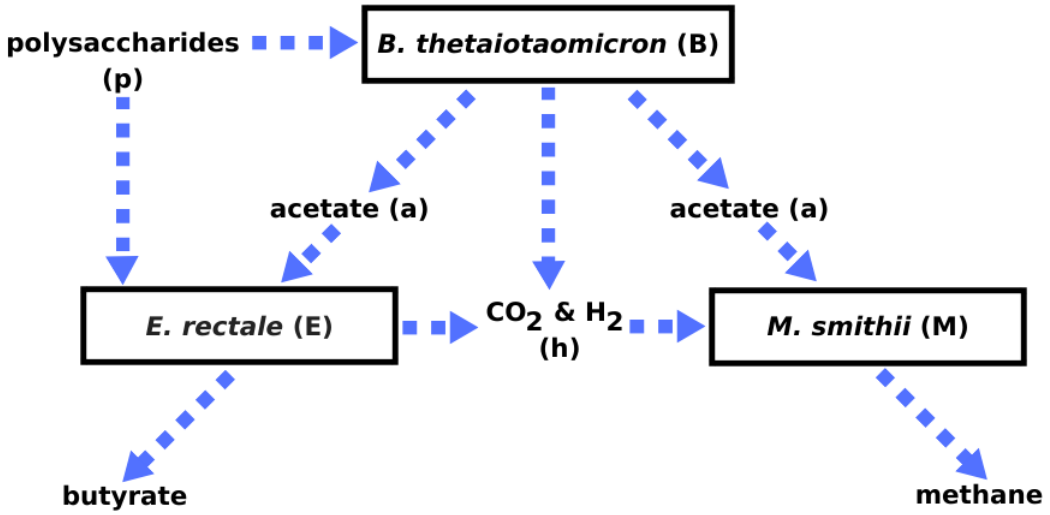

Figure 1: Graphical representation of the interactions of the three species, *B. thetaiotaomicron*, *E. rectale*, and *M. smithii*, and their metabolites. Each microorganism or substance given a letter is included in the mathematical model equations in Section 2.2. This figure was developed from information presented in Ji and Nielsen, 2015, Shoaie, Karlsson, Mardinoglu, et al., 2013, and Adamberg, Tomson, Vija, et al., 2014.

the acetate produced by *B. thetaiotaomicron*, *E. rectale* contributes to *M. smithii*'s metabolism by producing CO_2 and H_2. Although this interaction may seemingly place *M. smithii* at a metabolic advantage over *E. rectale*, both species are able to coexist in the human gut microbiome, providing supporting evidence of the hypothesis presented in Goldford, Lu, Bajie, et al., 2018.

Additionally, Figure 1 shows both *B. thetaiotaomicron* and *E. rectale* uptaking polysaccharides, suggesting a potential competition for this resource; however, *E. rectale* shifts from uptaking polysaccharides to utilizing amino acids, such as glutamine, when *B. thetaiotaomicron* is present. This behavior suggests that *B. thetaiotaomicron* is a better competitor for this resource by some mechanism (Ji and Nielsen, 2015).

Translation of a Graphical Model To a Dynamical System

The key assumption underlying our mathematical model is that the human gut can be thought of as a sort of chemostat, which is illustrated in Figure 2. An actual chemostat is a laboratory device used in the simulation and ecological study of populations, which provides an idealized

representation of naturally occurring phenomena (Smith and Waltman, 2008). Though the conditions of a chemostat are simplified and controlled in a laboratory setting, a chemostat can be useful in the study of population dynamics and the underlying mechanisms of interactions among populations (Smith and Waltman, 2008). The simple chemostat model is a first step in developing an initial theoretical framework, which then can be refine by more complex approaches (Smith and Waltman, 2008).

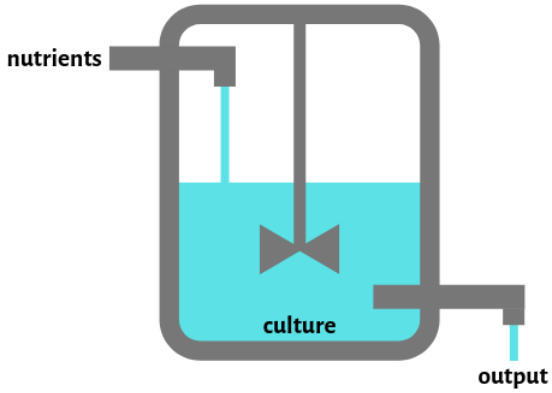

Figure 2: Diagram of a chemostat. A solution of nutrients enters the chemostat at a constant rate and concentration, which is then thoroughly mixed within the vessel in the culture solution. The output of the chemostat is the well-mixed solution, which exits the vessel at the same rate at which the nutrient solution enters the system (Smith and Waltman, 2008). A constant volume of the well-mixed solution is maintained within the chemostat.

In constructing a simple chemostat model, we assume that the contents of the vessel are well-mixed, the rate at which liquid enters the system equals the rate at which the well-stirred contents leave the compartment, and that other significant factors potentially affecting growth, such as temperature, are held constant (Smith and Waltman, 2008). While the baseline assumptions underlying the single chemostat model are biologically unrealistic in general, the inconsistencies with reality that arise with these assumptions can be alleviated by a more complex implementation of this type of model that relaxes these assumptions. Additionally, the general idea of thinking of the human gut as a single chemostat can be extended in future implementations to a multi-compartmental chemostat model, where each compartment represents a connected physiological structure of the human gut, such as the stomach and intestines. However, for the purposes of

evaluating the efficacy of a simplistic model before considering more complex implementations, we retain these assumptions and focus on evaluating a single-chemostat model.

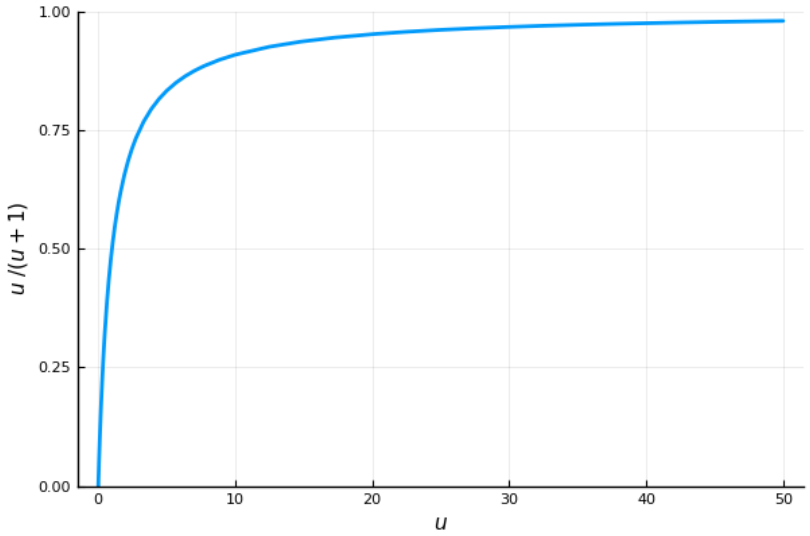

Figure 3: Graph of the value of the Monod form as the concentration of a substance increases. When the values of the concentration are low, this formulation assumes that populations dependent on this substance are not able to reach their maximum growth rate; instead, this population's growth rate is scaled down proportional to the value of the Monod form. As the availability of this substance increases, the value of the Monod form asymptotically approaches 1, which means an adequate concentration of the substance is available for the dependent population to grow at its maximum birth rate.

An additional underlying assumption of our model commonly seen in other modeling literature is that microorganisms grow at a rate following the Monod form

$$\beta_X \left(\frac{u}{u+\gamma} \right) X,$$

where β_X is the maximum birth rate of population X, u is the concentration of the nutrient population on which X's growth depends, γ is the Michaelis-Menten constant, and X is the concentration of the microorganism (Smith and Waltman, 2008). The constants β_X and γ can be experimentally determined, where γ is measured as the additional mass created by an organism divided by the total utilized mass of the nutrient (Smith and Waltman, 2008). Figure 3 provides a graphical representation and explanation of the fractional term $u/(u+\gamma)$ in the Monod form, and this term is subsequently denoted as $\Psi_\gamma(u)$ in our model as specified in equation (3).

Microorganism Equations

$$\frac{dB}{dt} = \beta_B \Psi_{\gamma_p}(p) B - qB \tag{1a}$$

$$\frac{dE}{dt} = \left[\beta_{E_1} \Psi_{\gamma_a}(a) + \beta_{E_2} \left(1 - \Psi_{\gamma_B}(B)\right) \Psi_{\gamma_p}(p)\right] E - qE \tag{1b}$$

$$\frac{dM}{dt} = \left[\beta_{M_1} \Psi_{\gamma_a}(a) + \beta_{M_2} \Psi_{\gamma_h}(h)\right] M - qM \tag{1c}$$

Metabolite Equations

$$\frac{da}{dt} = \beta_a \Psi_{\gamma_p}(p) B - qa - [\mu_{a,E} E + \mu_{a,M} M] a \tag{2a}$$

$$\frac{dh}{dt} = \beta_{h_1} \Psi_{\gamma_a}(a) E + \beta_{h_2} \Psi_{\gamma_p}(p) B - qh - \mu_{h,M} hM \tag{2b}$$

$$\frac{dp}{dt} = \beta_p q (\cos(t) + 1)^3 - qp - [\mu_{p,B} B + \mu_{p,E} E] p \tag{2c}$$

Additional Terms

$$\Psi_\gamma(u) = \frac{u}{u + \gamma} \tag{3}$$

$$q = \frac{V}{Q} \tag{4}$$

The microorganism equations detail the rates of change overtime for the three species *B. thetaiotaomicron* in (1a), *E. rectale* in (1b), and *M. smithii* in (1c), which depend on the availability of necessary nutrients and the rate at which these microorganisms are flushed out of the system. Our model assumes that there is a high enough rate of turnover of fluids in the human gut such that these microorganisms are almost always flushed out of the system before their life expectancy, so death terms are neglected in these three equations. All three microorganism equations are

constructed in the same general format:

$$\Delta_i = P_{ij} - F_i,$$

where Δ_i is the rate of change of biomass for microorganism i, P_{ij} is the rate at which microorganism i proliferates based on the availability of substance j, and F_i is the rate at which microorganism i is flushed out of the gut. In all three microorganism equations, the term F_i is the biomass of the given population multiplied by the constant q. The fixed quantity q is interpreted as the rate at which the contents of the gut leave the system as expressed in equation (4), where V is the volume of the chemostat and Q is the rate of volumetric flow within the chemostat (Smith and Waltman, 2008). The rate at which biomass increases or decreases for a given microbial population, however, depends on the specific set of substances each species metabolizes.

In the case of *B. thetaiotaomicron* in equation (1a), an increase in its biomass depends on the amount of polysaccharides p, as well as its birth rate β_B. *E. rectale* is a microorganism with more complex interactions with *B. thetaiotaomicron*, which led to the intricacy of equation (1b). In order for *E. rectale* to grow in biomass, acetate or polysaccharides need to be present in the ecosystem (Shoaie, Karlsson, Mardinoglu, et al., 2013). We assume that *E. rectale* has different maximum growth rates depending on each nutrient, leading us to split β_E into three different, related constants. Because *E. rectale* shifts to uptaking inorganic ammonia when *B. thetaiotaomicron* is present, we included the term $\left(1 - \Psi_{\gamma_B}(B)\right)$ to reflect this shift. Equation (1c) for *M. smithii* is more simplistic compared to *E. rectale*. *M. smithii* depends on the presence of acetate and the gases CO_2 and H_2, which are both incorporated in the standard Monod form. Again, we assume that *M. smithii* grows at different maximum growth rates in the presence of only one of these metabolites, which lead to the separation of β_M into the constants β_{M_1} and β_{M_2}.

The metabolite equations detail the rates of change over time for the intermediate substances produced and consumed by these three species, where the concentrations are tracked for acetate in equation (2a), CO_2 and H_2 in equation (2b), and polysaccharides in equation (2c). These con-

centrations depend on the rate at which these metabolites are produced by the microorganisms or enter into the system, the rate at which they are flushed out of the system, and the rate at which they are consumed by surrounding microorganisms. All five metabolite equations are constructed in the general format:

$$\Delta_j = P_{ij} - F_j - C_{ij},$$

where Δ_j is the rate of change of metabolite j's concentration, P_{ij} is the rate at which metabolite j is produced by microorganism i or enters the system, F_j is the rate at which metabolite j is flushed out of the chemostat, and C_{ij} is the rate at which the metabolite j is consumed by microorganism i. In each of the metabolite equations, the parameter $\mu_{x,Y}$ is the rate at which substance x is utilized by species Y.

Equations (2a) and (2b) clearly reflect the schematic represented in Figure 1. An increase in acetate depends on the concentrations of *B. thetaiotaomicron* and polysaccharides and an increase in the gases CO_2 and H_2 depends on the concentrations of *B. thetaiotaomicron*, *E. rectale* and acetate. A decrease in the concentration of acetate depends on the rate at which *E. rectale* and *M. smithii* uptake acetate and the total concentration leaving the system, and a decrease in CO_2 and H_2 depends on the rate at which *M. smithii* uptakes these gases and the total concentration of these gases leaving the system.

Polysaccharides enter the human gut through diet, so we accounted for their addition to the gut through a sinusoidal function, $\big(\cos(t)+1\big)^3$. This function attempts to account for the duration of time in between meals through the period of the curve. In addition, this function is defined to be a smooth curve to illustrate the gradual breakdown of food and release of nutrients in the gut. The amplitude of this term is scaled based on the parameters β_p and q.

In our model, many of the parameter values are unknown or cannot be determined experimentally. In order to estimate these parameters mathematically, we searched for data tracking the biomass changes of *B. thetaiotaomicron*, *E. rectale*, and *M smithii* in order fit our model to this data with the goal of estimating these parameters. This system, however, does not account for all possible parameters that could potentially affect the fluctuations in biomass or substrate concentra-

tion. Though a more complex implementation of this model could account for additional species and their interactions with the members of our system, we limited ourselves to the three microorganisms *B. thetaiotaomicron*, *E. rectale*, and *M. smithii*. The number of species and metabolites considered can be expanded in future implementations.

DATA EXTRACTION AND PARAMETER ESTIMATION

In order to estimate our full model parameters, we identified data in the literature that is relevant to our system. From our literature scan, we were able to obtain longitudinal data on an experiment involving *B. thetaiotaomicron* and its substrates, presented in Adamberg, Tomson, Vija, et al., 2014. However, similar longitudinal data on all three species *B. thetaiotaomicron*, *E. rectale*, and *M. smithii* is not openly available in the published literature to our knowledge. Despite the lack of longitudinal data, we were able to extract a single set of data points for the full system including all three species and their relevant substrates from Shoaie, Karlsson, Mardinoglu, et al., 2013. Given these constraints, we present estimates of the parameters in our model to a low degree of precision.

Reduced Model Parameter Estimation

Table 1 presents data found in Adamberg, Tomson, Vija, et al., 2014. The experiment conducted in this paper focused on a subset of our full system by only considering the metabolism of *B. thetaiotaomicron* with the substrates H_2, CO_2, acetate, and polysaccharides. We utilized this data for parameter determination and estimation by condensing our model to only include *B. thetaiotaomicron* and its substrates.

	0 hrs	24 hrs	72 hrs
Polysaccharides (mM)	16.12	11.93	2.95
H_2 (mL)	0	0.0045	0.0051
CO_2 (mL)	0	0.0588	0.154
Acetate (mM)	0	3.296	7.46
Biomass (gDW)	2.34×10^{-5}	1.306×10^{-4}	2.897×10^{-4}

Table 1: Table for experimental results with data for polysaccharides, H_2, CO_2, acetate, and the biomass of *B. thetaiotaomicron* in a medium initially containing with 20 amino acids across distinct time points. Only the bacterial species *B. thetaiotaomicron* was present in the medium. The data contained in this table is originally from Adamberg, Tomson, Vija, et al., 2014.

The subsystem isolating *B. thetaiotaomicron*'s metabolism is illustrated in Figure 4. Based

on this graphical representation, we extracted the reduced system from our full model in equations (1) and (2), producing the reduced system of ODEs in equations (5). Any terms or equations in our full model that are dependent on *E. rectale* and *M. smithii*'s presence were removed in this reduction. Additionally, terms related to chemostatic flow were removed because the experimental data in Table 1 was collected in a closed system, rather than an open system like a chemostat. Lastly, we removed any terms that account for an input of polysaccharides because only a specified initial amount was placed in the experimental device for *B. thetaiotaomicron* to consume over time.

Figure 4: Graphical representation of a subset of *B. thetaiotaomicron*'s metabolism, including the utilization of polysaccharides and the production of acetate, CO_2, and H_2. This diagram is based on information presented in Adamberg, Tomson, Vija, et al., 2014.

$$\frac{dB}{dt} = \beta_B \Psi_{\gamma_p}(p) B \tag{5a}$$
$$\frac{da}{dt} = \beta_a \Psi_{\gamma_p}(p) B \tag{5b}$$
$$\frac{dh}{dt} = \beta_{h_2} \Psi_{\gamma_p}(p) B \tag{5c}$$
$$\frac{dp}{dt} = -\mu_{p,B} B p \tag{5d}$$

By fitting our reduced model to the collected data in Table 1, we estimated the model parameters using the Julia packages `DiffEqParamEstim` and `Optim` with the goal of utilizing these results to inform our full model parameter estimation. We ultimately chose the simulated annealing method to apply to our optimization problem because this method can be used to approximate the global optimum. In an attempt to test the performance of other optimization methods, we also

applied the conjugate gradient descent and the Nelder method to our problem, but found that these methods did not improve our model's fit to the data compared to the simulated annealing method for our specific problem.

In the first step of finding a roughly optimal solution for our reduced model's parameters, we first chose an arbitrary vector of initial values for these parameters. Because we did not have prior knowledge of the expected values of these parameters other than that they are positive, we chose arbitrary values in the hopes that the simulated annealing method would output a solution in the correct direction of the globally optimal solution. Once we chose these arbitrary initial values, we used the `DiffEqParamEstim` and `Optim` packages in order to obtain the optimal results given by the simulated annealing method. The results produced from this constrained optimization given our initial guess of parameter values allowed us to achieve a seemingly local optimum for the system. We utilized this output by adjusting the output parameter values ourselves in order to further reduce the residuals. Using this new set of parameters, we reran the optimization once more in the hopes of producing a better fit to the data.

This approach proved to be useful because the model fit noticeably improved after our alterations. Figures 5, 6, and 7 show the solutions to our system of ODEs using our final set of roughly optimized parameters in Table 2 based on the data in Table 1. Though the output of the simulated annealing algorithm produced many more significant digits for these parameter estimates than is shown in Table 2, we limited these results to one significant digit in order to reflect our uncertainty with the precision of these estimates, especially due to the sparsity of this data.

β_a	β_B	β_{h_2}	γ_p	μ_{pB}
2000	0.08	50	10	200

Table 2: Table of fitted parameter values for equations (5) based on the experimental data in Table 1.

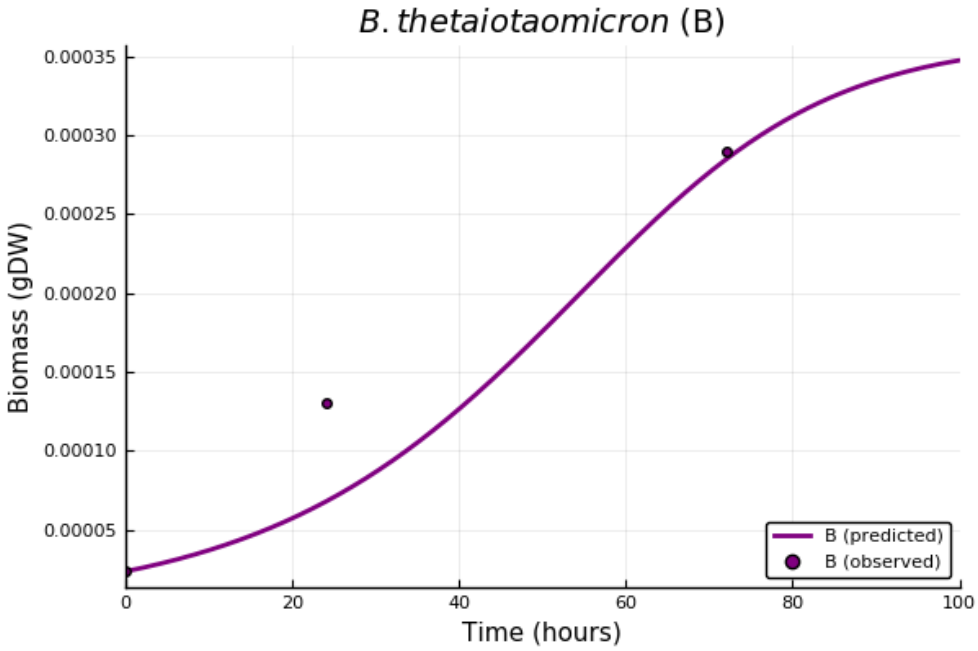

Figure 5: Plot of the solutions to equation (5a) and the observed data for *B. thetaiotaomicron* over the time interval 0 to 72 hours.

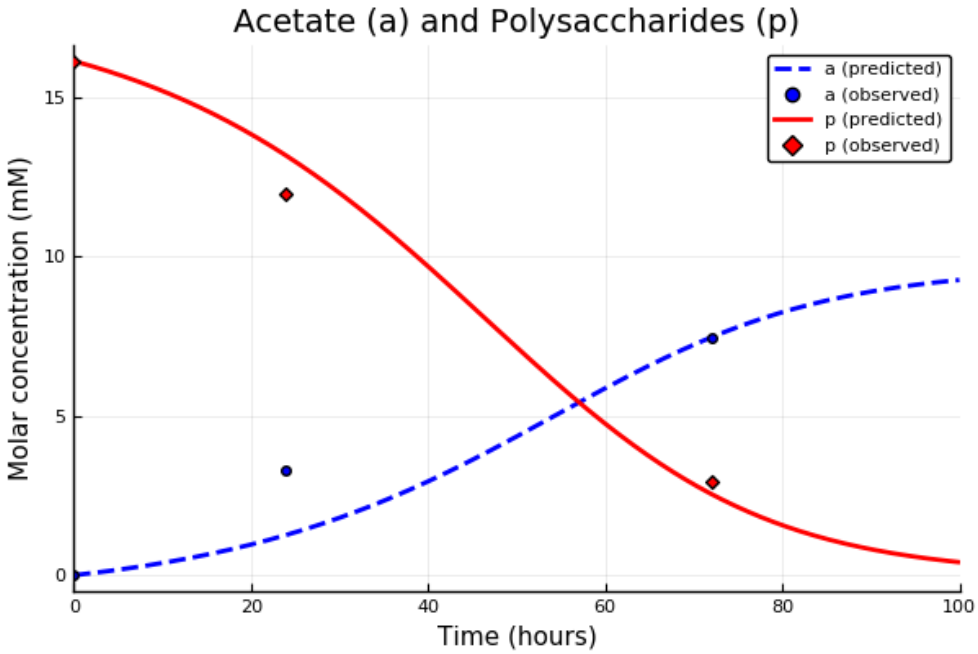

Figure 6: Plot of the solutions to equations (5b) and (5d), as well as the observed data for acetate and polysaccharides over the time interval 0 to 72 hours.

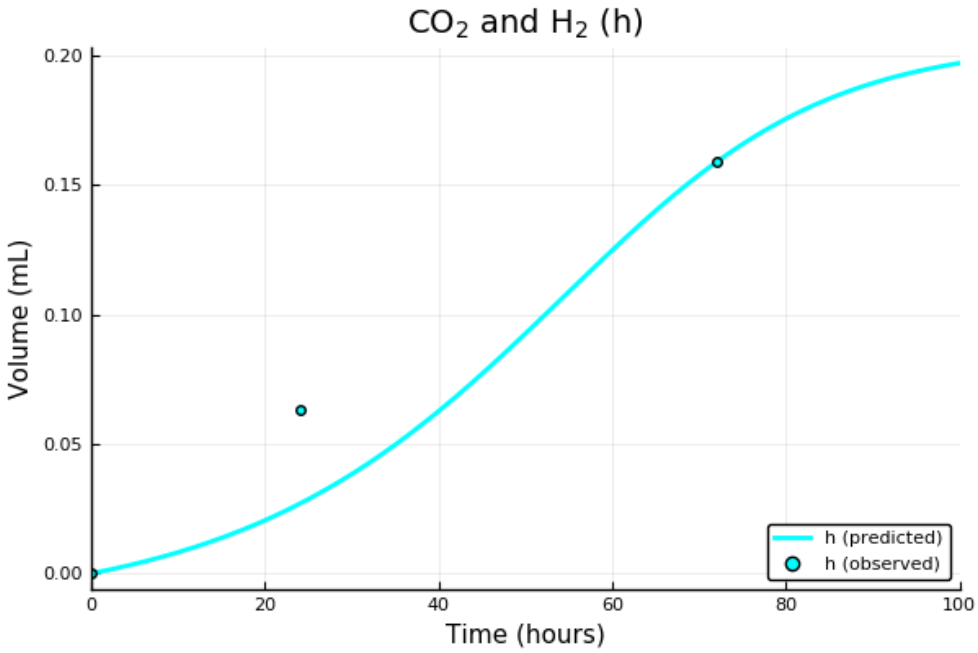

Figure 7: Plot of the solutions to equation (5c) and the observed data for CO_2 and H_2 over the time interval 0 to 72 hours.

Full Model Parameter Estimation

Due to limitations in availability of data for our full model containing the three species *B. thetaiotaomicron*, *M. smithii*, and *E. rectale* simultaneously, we are not able to estimate the remaining model parameters to the same degree of confidence as we were with our reduced system. In our literature review, however, we were able to obtain a single set of endpoint data values for an experiment found in Shoaie, Karlsson, Mardinoglu, et al., 2013 conducted on all three species. Given the lack of available longitudinal data for this biological system, we assume that this data, shown in Table 3, are the center values of the oscillations for each substrate or biomass quantity. One complication that arises from using this data is that the total biomass for all three species was experimentally measured as a single quantity, which is another factor that further contributes to our uncertainty in our full model parameter estimates. To circumvent this limitation, we fit the model parameters to produce numerical solutions with average biomass concentrations that roughly sum to the biomass quantity given in Table 3. The parameter estimates obtained in the previous section on our reduced model, including some manual adjustments, allowed us to provide rough estimates

of the remaining parameters in our full model.

Polysaccharides (μM)	32.06
H_2 (μM)	0
CO_2 (μM)	7.96
Acetate (μM)	9.71
Biomass (gDW)	0.001412

Table 3: Table of experimental data presented in Shoaie, Karlsson, Mardinoglu, et al., 2013 of the microorganisms *B. thetaiotaomicron*, *E. rectale*, and *M. smithii* and their substrates CO_2, H_2, acetate, and polysaccharides. The biomass measurement is a combination of the biomass of the three microorganisms.

Through manually fitting the model parameters to the data shown in Table 3, we obtained rough estimates of our model's parameters, which are shown in Table 4. The general approach we took in estimating these parameters given the lack of longitudinal data was to first utilize the information gained from estimating our reduced model parameters, shown in Table 2. We treated the parameter estimates obtained from the reduced model as initial parameter values that could be adjusted as needed, rather than a fixed quantity. This choice was made because the experiments conducted to produce the data given in Tables 1 and 3 were performed under different experimental conditions, such as the temperature of environment and the medium the culture grew in, which are factors that could potentially impact the estimates of these parameters.

For the remaining parameters in the model which we lacked experimental data to inform our initial choice of parameter values, we assigned arbitrary values that were of the same magnitude as similar parameters that we estimated in the reduced model. After assessing the fit of the numerical solutions to the data using our initial set of parameter estimates, these estimates were iteratively adjusted in order to reduce the error between the center value of the numerical solution's oscillations and the data value given in Table 3. The resulting parameter estimates based on implementing this approach are given in Table 4. In order to reflect our uncertainty in these parameter estimates, we limited these estimates to one significant figure. In specific cases, we ultimately maintained the two significant figures in some parameter estimates due to the resulting improvements in the

model fit.

β_a	1.0×10^7	β_{h_2}	33,000	γ_B	10	μ_{hM}	40
β_B	1.2	β_{M_1}	0.8	γ_h	150	μ_{pB}	200,000
β_{E_1}	0.8	β_{M_2}	0.5	γ_p	400	μ_{pE}	5,000
β_{E_2}	0.6	β_p	10,000	μ_{aE}	25,000	q	0.05
β_{h_1}	150	γ_a	200	μ_{aM}	50,000		

Table 4: Table of fitted parameter values for equations (1) and (2) based on the experimental data in Table 1.

Using the set of parameter estimates in Table 4, the solutions to our full model are given in Figures 8, 9, 10, and 11. These plots show the solutions to our full system of ODEs in equations (1) and (2) from 1,000 to 1,100 hours. This time range was selected in order to allow for a sufficient amount of time to pass in order for the system to converge to an oscillatory steady state. The center values of the oscillations are represented by dotted lines in each plot. In Figures 9, 10, and 11, the data values given in Table 3 are superimposed on their respective plots as dashed lines in order to provide a visual representation of the error between the center of the oscillatory steady state and the observed data. Despite some error in the center value of the ODE solutions compared to the data, the observed data values for acetate, polysaccharides, and the gases CO_2 and H_2 are contained in the oscillation range of the ODEs' numerical solutions, so we conclude that our parameter estimates sufficiently fit the data.

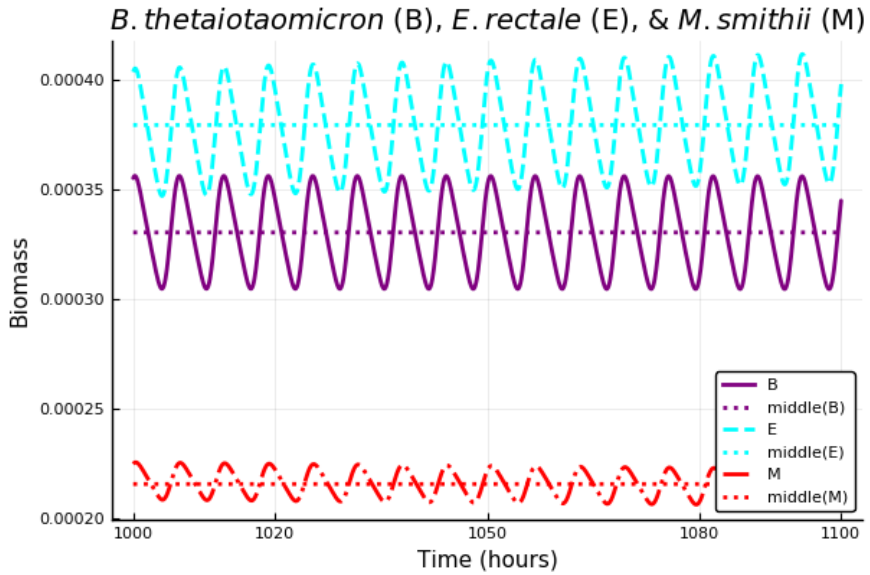

Figure 8: Plot of the solutions to equations (1a), (1b), and (1c) using the observed data in Table 3 over the time interval 1,000 to 1,100 hours. Based on Table 3, the sum of the three species' biomass should be 0.001412 gDW. The middle of the solutions to the ODEs using the parameter estimates in Table 4 for all three species sums to 0.000925 gDW, resulting in an error of -0.000487 gDW.

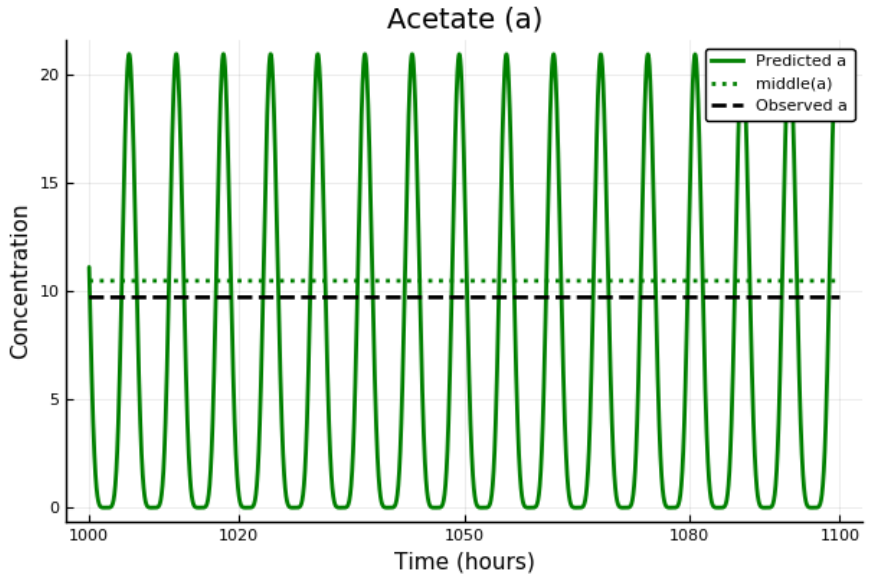

Figure 9: Plot of the solutions to equation (2a) using the observed data in Table 3 over the time interval 1,000 to 1,100 hours. Based on Table 3, the center of acetate's oscillations should be 9.71 μM. The middle concentration of the ODE solutions for acetate is 10.47 μM, resulting in a 0.76 μM error.

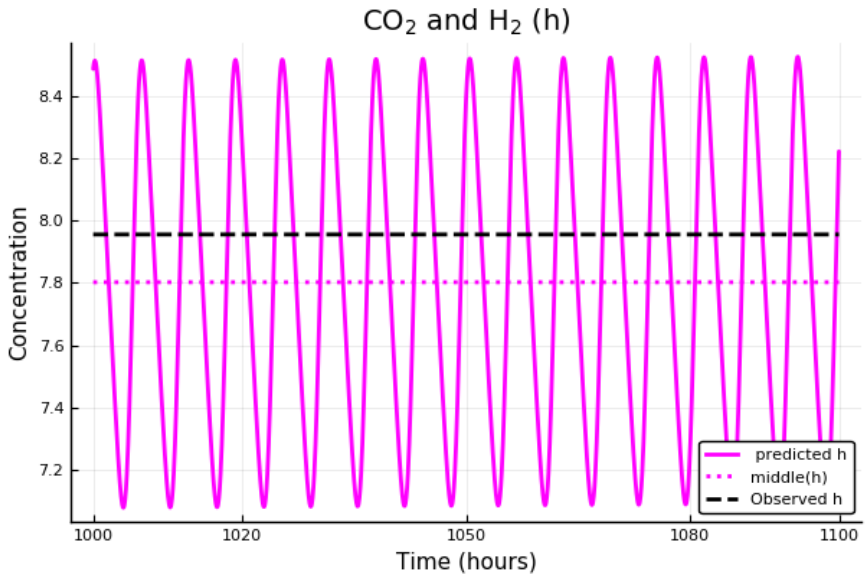

Figure 10: Plot of the solutions to equation (2b) using the observed data in Table 3 over the time interval 1,000 to 1,100 hours. Based on Table 3, the center of CO_2 and H_2's oscillations should be roughly 7.96 μM. The middle concentration of the ODE solutions for CO_2 and H_2 is 7.80 μM, resulting in a -0.16 μM error.

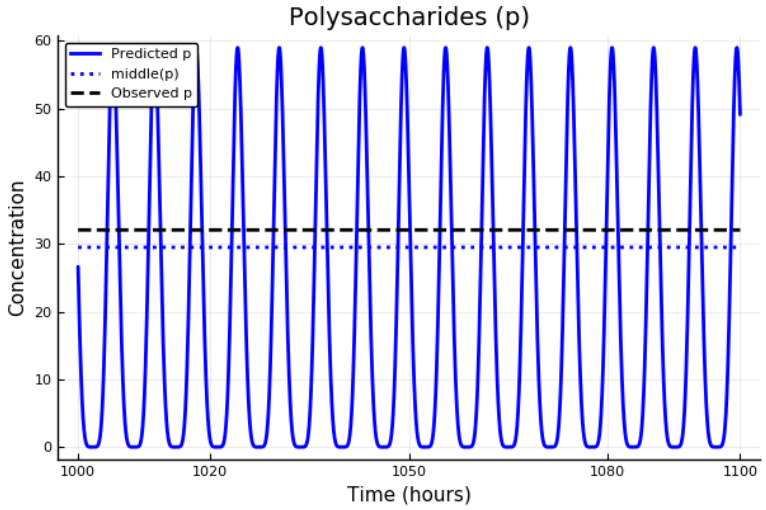

Figure 11: Plot of the solutions to equation (2c) using the observed data in Table 3 over the time interval 1,000 to 1,100 hours. Based on Table 3, the center of polysaccharides' oscillations should be roughly 32.06 μM. The middle concentration of the ODE solutions for polysaccharides is 29.50 μM, resulting in a -2.56 μM error.

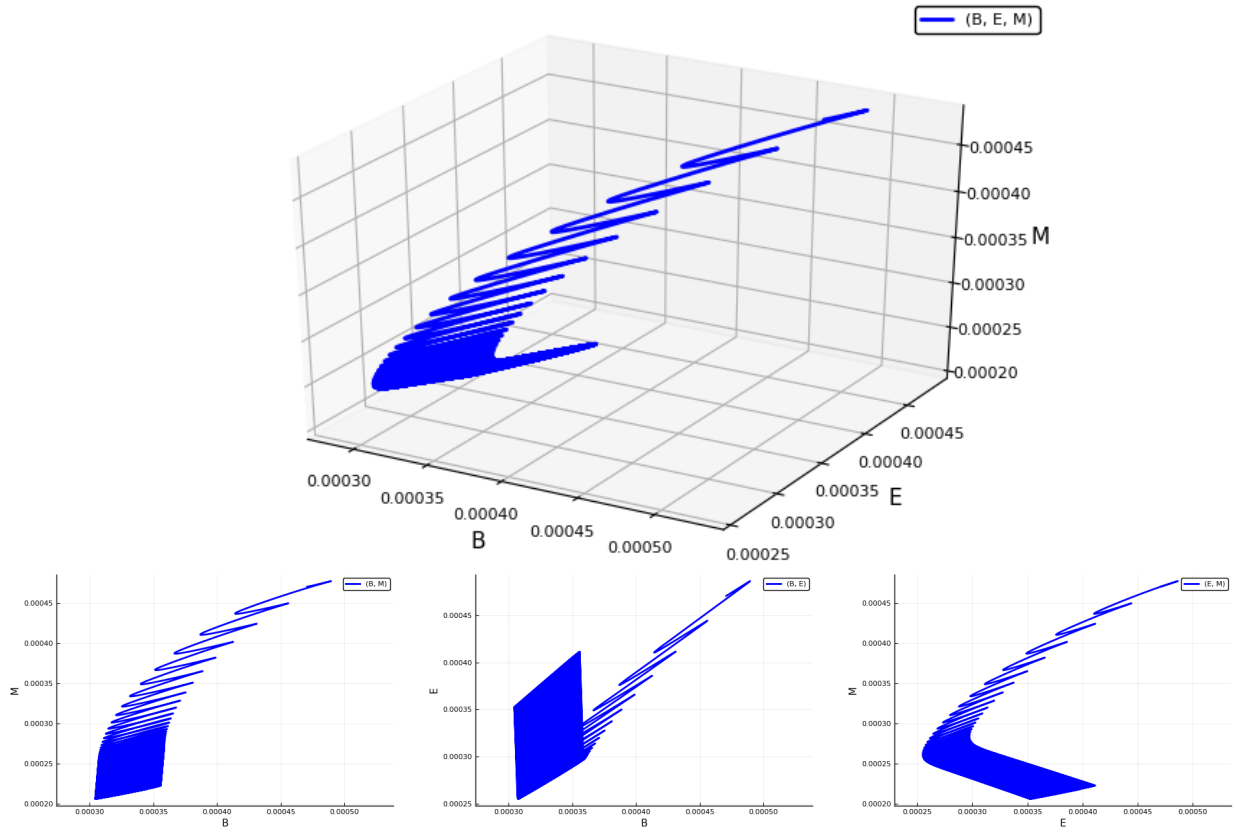

Figure 12: 3D phase plane diagram of the solutions to our full model in equations (1) using the parameter estimates in Table 4. The three bottom plots show the 2D projections of the pairs (B, M), (B, E), and (E, M).

In order to further visualize the convergence of the solutions to our full model's system of ODEs, we produced a 3D phase plane in Figure 12 of the three species *B. thetaiotaomicron* (B), *E. rectale* (E), and *M. smithii* (M). The solutions to the system of ODEs in equations (1) are plotted from 0 to 6,000 hours. By testing various time endpoints and plotting the solutions on a 3D phase plane, we obtained graphical results that showed no changes compared to our plot in Figure 12. With these results, we conclude that the solutions to our ODE system reach a steady state. Reaching a steady state for a system of ODEs with a specific parameter set is an important result to achieve because it illustrates that the solutions to this system are able to maintain plausible biomass or substrate concentrations over time.

SENSITIVITY ANALYSIS

Based on the fitted parameter values and the available data from the literature, we analyzed our model's sensitivity to its parameters. The value of sensitivity analysis in understanding our results is one reason we used baseline parameters with one or two significant digits. In order to identify the parameters in the model that have the greatest effect on the model output, we conducted sensitivity analysis on our reduced and full model by computing the first- and total-order effects based on Sobol' indices.

Sensitivity analysis can be defined as the study of how uncertainty in the model input propagates into uncertainty in the model output (Saltelli, Ratto, Andres, et al., 2008). Once parameters are determined, or estimated along with their errors by one of various methods like simulated annealing or the Levenberg-Marquardt method, sensitivity analysis can be used to identify the driving parameters of the system that contribute to the most variability in the model's output. Overall, sensitivity analysis tests the robustness of the model, identifies if the model relies on weak assumptions, and allows for model simplification (Saltelli, Ratto, Andres, et al., 2008).

Local vs. Global Sensitivity

Local sensitivity analysis explores the changes in model output based on small, incremental changes to the model inputs, centered around a baseline value. Generally, local sensitivity analysis gives a measure of the partial derivatives of the model with respect to each input. This approach is best utilized when there is little uncertainty around model parameters or initial conditions. However, in the context of systems biology and in the case of our mathematical model, the assumption of minimal uncertainty in parameters and initial conditions is not feasible. Due to the vast uncertainty and variability of model parameters relating to biological systems, analyzing the behavior of the model output around a single value in the parameter space is not useful in exploring the model's overall behavior (Marino, Hogue, Ray, and Kirschner, 2009).

Instead, we utilized global sensitivity analysis in order to identify the driving parameters

of our model. In this framework, which utilizes Monte-Carlo (MC) methods, each parameter is assigned a probability density function (pdf) based on *a priori* information known about the parameter. Samples are drawn from the pdfs to evaluate the overall model output. In the case where there is no *a priori* information about the model parameter, a non-informative distribution is assigned to the parameter, commonly a uniform distribution that spans a large range values in the parameter space. On the other hand, if there is some biological knowledge in the literature about the parameter, especially an expected value, a more informative distribution can be assigned to the parameter, such as a normal distribution centered around the expected value. Samples pulled from each parameter's assigned distribution can be generated completely randomly or through quasi-random sampling strategies that aim to explore the entire parameter space, such as Latin hypercube sampling (LHS) and sampling from low-discrepancy sequences (Marino, Hogue, Ray, and Kirschner, 2009). The set of parameter samples based on the specified number of simulations are then used to evaluate the model, each producing an output value. With these computed values for the model output, sensitivity indices can then be computed.

Because our model is nonlinear, implementations of global sensitivity analysis methods for linear relationships like the partial rank correlation coefficient (PRCC), the Pearson correlation coefficient (CC), and standardized regression coefficients (SRC) would not be useful. Consequently, we employed the Sobol' method, a variance-based decomposition method. The extended Fourier amplitude sensitivity test (eFAST) method, which quantifies each parameter's frequency strength from the model input to the output, is also a reliable approach to conducting sensitivity analysis on a nonlinear model (Marino, Hogue, Ray, and Kirschner, 2009).

Sobol' Sensitivities

Sobol' indices are a MC variance-based approach to calculating all first-order and total-effects indices in a model with k parameters. These sensitivity indices are computed based on model evaluations for N simulations. Algorithm 1 outlines the general method of computing both indices, and has a computational cost of $N(k+2)$ runs. The algorithm input N is chosen based on

the number of parameters in the model; more parameters in the model require more model simulations. In Algorithm 1, a matrix of quasi-random numbers is generated from each parameters' pre-specified distribution. In order to improve the sensitivity estimates, the values tested for each parameter in the model are drawn from quasi-random number generators. Because randomly generated numbers tend to cluster together and leave detectable gaps of unsampled space, a stratified sampling approach is recommended to test values that span the entire parameter space. Quasi-random numbers relieve the issue of unsampled gaps in the parameter space by partitioning this space into equal subintervals and randomly sampling points within those subintervals, which are then used in sensitivity calculations. With more than one point randomly sampled from each subinterval, unbiased estimates of the mean and variance can still be obtained (Saltelli, Ratto, Andres, et al., 2008). Though the use of quasi-random numbers from a given distribution is not necessary, we incorporate this approach into our computations.

The formula presented in Saltelli, Ratto, Andres, et al., 2008 defines the first order Sobol' sensitivities as

$$S_i = \frac{V[E(Y|X_i)]}{V(Y)} = \frac{\frac{1}{N}\sum y_A^{(j)} y_{C_i}^{(j)} - \frac{1}{N^2}\sum y_A^{(j)} \sum y_B}{\frac{1}{N}\sum \left(y_A^{(j)}\right)^2 - f_0^2}, \tag{6}$$

where

$$f_0^2 = \left(\frac{1}{N}\sum_{j=1}^{N} y_A^{(j)}\right)^2,$$

and the Sobol' total-effect indices are defined as

$$S_{T_i} = 1 - \frac{V[E(Y|X_{-i})]}{V(Y)} = 1 - \frac{\frac{1}{N}\sum y_B^{(j)} y_{C_i}^{(j)} - f_0^2}{\frac{1}{N}\sum \left(y_A^{(j)}\right)^2 - f_0^2}. \tag{7}$$

In equations (6) and (7), A and B are $N \times k$ matrices of simulated values from each parameter's pre-specified distribution using a low-discrepancy sequence; C_i, where $i = 1, \ldots, k$, is a $N \times k$ matrix where every element of C_i is the matrix B, except column i of C_i is replaced with column i of matrix A; and y_A, y_{C_i}, and y_B are $N \times 1$ vectors of the model outputs produced from the simulated parameter values in matrix A, C_i, and B, respectively (Saltelli, Ratto, Andres, et al.,

2008).

The total effect of parameter X_i, where $i = 1, \ldots, k$, in a model can be interpreted as the sum of its main effect in addition to all higher-order interaction terms. Given a parameter X_i, if $S_{T_i} - S_i \approx 0$, this result suggests that the higher-order effects are not important in explaining the overall variance in the model output. If for a given parameter X_i, the main effect S_i is close to 0, this result is insufficient to conclude that the parameter is non-influential. However, $S_{T_i} = 0$ is a necessary and sufficient condition to deduce that parameter X_i is non-influential, meaning this parameter can be fixed to any value in its distribution and have a negligible effect on the variance of the model's output (Saltelli, Ratto, Andres, et al., 2008). Theoretically, $0 \leq S_i,\ S_{T_i} \leq 1$ for each parameter, and $0 \leq \sum_{i=1}^{k} S_{T_i} \leq 1$. However, due to error in estimation for each sensitivity index in practice, S_i and S_{T_i} can be computed to be slightly negative, and $\sum_{i=1}^{n} S_{T_i}$ can slightly exceed 1. An increased number of simulations N in computing these indices may reduce these errors.

Using Algorithm 1 in Appendix A, we wrote a function in Julia v1.1.1 that computes the first-order and total-order Sobol' indices for a given ODE-based model for each parameter-variable combination, which is included in Appendix C. To compute the first- and total-order Sobol' indices for our reduced model parameters, we specified prior distributions for each parameter from which to draw samples. Using the inverse transform sampling method, parameter samples are initially drawn from a quasi-random number generator on the interval [0,1]. Because the result of this draw is a value between 0 and 1, it can be interpreted as a probability. This probability p is then input into the inverse of the cumulative distribution function (cdf) of the given parameter distribution X in order to produce a value x^* in the parameter distribution's domain. This transformation can be denoted as $F_X^{-1}(p) \equiv x^*$, where F_X^{-1} is inverse of the cdf of the distribution for parameter X. The resulting value x^* is then used as a simulated parameter value for the sensitivity calculation of the Sobol' indices (Devroye, 1986).

First- And Total-order Sobol' Indices For The Reduced Model Parameters

In order to specify the distribution of our model's parameters, we first considered the domain of each parameter. All parameters in our reduced model can reasonably take on positive values, so we restricted our consideration to parameter distributions defined on the positive number line. Due to its positive, continuous domain, we chose the Gamma(k, θ) distribution as the general form of our specified parameter distributions, whose pdf is parameterized as

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}, \quad x > 0,$$

where $k > 0$ is the shape parameter, and $\theta > 0$ is the scale parameter. Using the results of our parameter estimation in Table 2, we defined the mean of the Gamma distributions to be equal to the estimates we obtained from our parameter estimation. Because we obtained one set of parameter estimates for our reduced model, we are not able to obtain an informed variance on our parameter distributions. In order to reflect our minimal knowledge of the true parameter values, we define the variance of these Gamma distributions to be large in order to test a wide range of parameter values in our sensitivity calculations, which are specified in Table 5.

	k	θ	Mean	Variance
β_a	2000	1	2000	2000
β_B	0.01	8	0.08	0.64
β_h	50	1	50	50
γ_p	10	1	10	10
μ_{pB}	200	1	200	200

Table 5: Table of defined parameter distributions for the parameters in our reduced model, including the mean and variance of the distribution.

The first- and total-order results of the sensitivity analysis of the parameters in our reduced model are presented in Tables 6 and 7. Based on the small changes observed from the first-order sensitivities compared to the total-order sensitivities, we can conclude that higher order parameter interactions do not significantly contribute to the variance of the model outputs, so these higher

	B	a	h	p
β_a	0.000	0.003	0.002	0.003
β_B	0.995	0.838	0.572	0.891
β_h	0.000	0.000	0.304	0.000
γ_p	0.000	0.120	0.080	0.064
μ_{pB}	0.000	0.025	0.017	0.034
Total	0.995	0.982	0.974	0.992

Table 6: Table of first-order sensitivity indices based on equation (6) for our reduced model in equations (5) using $N = 2^{18}$ simulations.

	B	a	h	p
β_a	0.003	0.003	0.002	0.002
β_B	1.000	0.856	0.596	0.898
β_h	0.002	0.001	0.318	0.000
γ_p	0.003	0.133	0.092	0.072
μ_{pB}	0.002	0.030	0.021	0.035
Total	1.010	1.012	1.028	1.007

Table 7: Table of total-order sensitivity indices based on equation (7) for our reduced model in equations (5) using $N = 2^{18}$ simulations.

order interactions can safely be ignored. From the sensitivity results, it can be concluded that the parameter β_B is influential on the variance of the model outputs due to β_B's high sensitivity indices for all outputs. An unexpected result based on these indices is that the variable p is more sensitive to changes in the parameter γ_p, the Michaelis-Menten constant in the Monod form, than μ_{pB}, the rate *B. thetaiotaomicron* consumes polysaccharides. Though γ_p explains more variance in p than μ_{pB}, the first- and total-order effects for this parameter are relatively low compared to the main effects of parameter β_B, which appears to be the driving parameter of this subsystem. Overall, the results obtained in Tables 6 and 7 are generally unsurprising and intuitive. With these parameter estimates and Sobol' indices in mind, we then utilized this information in our full model to conduct a sensitivity analysis of all nineteen parameters.

First- And Total-order Sobol' Indices For The Full Model Parameters

In order to specify distributions for each parameter in the full model, we first set all but one parameter to a Gamma(k, θ) distribution since we expect all of these parameters to be strictly positive. The parameter q was not set to a Gamma(k, θ) distribution, but was instead specified as a Beta$(5, 95)$ distribution. Though the support of this distribution restricts the samples to the interval $[0, 1]$, we deemed this option to be more appropriate due to the Gamma(0.025, 2.0) distribution's heavy right-skewness and tendency to sample infinitesimally small, unlikely values.

We implemented a similar approach in specifying each k and θ as in the reduced model; the mean of each parameter's distribution was set to be equal to our estimated value in Table 4. The variance was specified to be equal to two times the estimated parameter value. We chose this value for the variance in order to cover a wide range of sample values for our sensitivity analysis due to a large uncertainty in our parameter estimates. Table 8 shows the values of k, θ, and the mean and variance of each parameter distribution. With these specified parameter distributions, samples were pulled using a low-discrepancy sequence, and these samples were then used to calculate the first- and total-order Sobol' indices for each model parameter and output. The results of our sensitivity analysis are given in Tables 10 and 11 in Appendix B.

Computing the first- and total-order Sobol' sensitivities for our full model using the parameter distributions specified in Table 8 proved to be somewhat computationally intensive due to the amount of time needed to solve our ODE system. In order to mitigate the computational time required to solve our full ODE system for a total of $N = 2^{18}$ simulations, we utilized The University of Iowa's Argon HPC Cluster by submitting our computational jobs in batches. To run our computations concurrently, the simulations were split into twenty-one cases. Each of these twenty-one matrices consisted of 2^{18} rows, where each row is a vector of simulated values for each model parameter. These values were stored in the matrices A, B, $C_1, ..., C_{19}$, which correspond to lines 2 and 5 in the Algorithm 1 included in Appendix A. These matrices of simulated parameter values were then input into our full ODE system, resulting in another set of twenty-one matrices, each with dimensions $2^{18} \times 6$, containing our full ODE model's endpoint solutions. Implementing

	k	θ	Mean	Variance
β_a	5.00×10^6	2.00	1.00×10^7	2.00×10^7
β_B	0.60	2.00	1.20	2.40
β_{E_1}	0.40	2.00	0.80	1.60
β_{E_2}	0.30	2.00	0.60	1.20
β_{h_1}	75.00	2.00	150.00	300.00
β_{h_2}	1.67×10^4	2.00	3.33×10^4	6.67×10^4
β_{M_1}	0.40	2.00	0.80	1.60
β_{M_2}	0.25	2.00	0.50	1.00
β_p	5.00×10^3	2.00	1.00×10^4	2.00×10^4
γ_a	100.00	2.00	200.00	400.00
γ_B	5.00	2.00	10.00	20.00
γ_h	75.00	2.00	150.00	300.00
γ_p	200.00	2.00	400.00	800.00
μ_{aE}	1.25×10^3	2.00	2.50×10^4	5.00×10^4
μ_{aM}	2.50×10^4	2.00	5.00×10^4	1.00×10^5
μ_{hM}	20.00	2.00	40.00	80.00
μ_{pB}	1.00×10^5	2.00	2.00×10^5	4.00×10^5
μ_{pE}	2.50×10^3	2.00	5.00×10^3	1.00×10^4
	α	β	Mean	Variance
q	5.00	95.00	0.05	4.70×10^{-2}

Table 8: Table of specified distributions for the parameters in our full model, including the mean and variance of each distribution. The parameters whose distribution is specified with k and θ follow a Gamma(k,θ) distribution. The parameter q, however, follows a Beta(α,β) distribution.

line 6 of Algorithm 1 in Appendix A results in these $(2^{18}\times 6)$ matrices of model solutions. The total amount of computational time needed in order to produce these solutions was roughly two to five hours, which would have taken between 42 to 105 hours if done in series.

Using our code in Appendix C to produce the first- and total-order Sobol' indices for our full ODE model, we obtained the results seen in Appendix B. Based on Table 10 of the first-order Sobol' indices, β_B is an influential parameter for *B. thetaiotaomicron*'s output variance; β_{E_2} for *E. rectale*'s variance; β_{M_2} for *M. smithii*'s variance; no first-order parameter for acetate's variance; β_B and β_{M2} for CO_2 and H_2's variance; and β_B for polysaccharides' variance. Based on Table 11 of the total-order Sobol' indices, β_B is an influential parameter for *B. thetaiotaomicron*'s output variance; β_{E_2} and β_B for *E. rectale*'s variance; β_{M_2}, q, and β_B for *M. smithii*'s variance; β_B, β_{E_1}, β_{E_2}, β_{M_1},

β_{M_2}, and q for acetate's variance; β_B, β_{E_2}, β_{M_1}, β_{M_2}, and q for CO_2 and H_2's variance; and β_B, β_{E_2}, and q for polysaccharides' variance. Based on the changes in the estimated indices from first-order to total-order indices, there appears to be a higher-order interaction among model parameters for *E. rectale*'s, *M. smithii*'s, acetate's, CO_2 and H_2's, and polysaccharides' output variance. Overall, the two main driving parameters of the ODE model appear to be β_B and q. Because these parameters are identified to be sensitive to input perturbations, careful attention needs to be given to estimating these parameters and reducing the error of their estimates in future efforts.

Some results seen in these tables are unsurprising: the birth rates of the microorganisms affect the variation in their biomass and the concentration of the substances they produce and/or consume. Additionally, the γ parameters are insignificant in affecting the variation in model output, which is to be expected. These unsurprising results provide evidence to support our mathematical model as a hypothesis. On the other hand, some results were unexpected, such as the total-order Sobol' indices results in Table 11 for acetate and polysaccharides summing to be greater than the theoretical maximum sum of 1. In order to investigate whether this discrepancy was due to a failure to sufficiently converge, a total of $N = 2^{19}$ simulations were run, again using the standard algorithm in Appendix C, resulting in a new set of first- and total-order Sobol' indices. These results showed an insignificant change compared to the previous estimates, so we concluded that these results can be attributed to an issue other than a lack of convergence. Based on Iooss and Prieur, 2019, observing Sobol' indices whose sum is greater than 1 is evidence of correlated inputs in the model. From the results in Table 11, there appears to be correlation among parameters when computing the output variance of acetate and polysaccharides due to their large total-order sum.

CONCLUSIONS AND FURTHER DIRECTIONS

The human gut microbiota has received much consideration in the recent published literature, reflecting its growing importance in understanding human health. Our specific method of study for this ecosystem was to mathematically model the population dynamics of a small subset of microorganisms commonly contained in the human gut using a system of ordinary differential equations. This method of analysis supplements other common approaches, such as GEMs and ABMs, which do not consider information at the strain-level as our model does (Kumar, Ji, Zengler, and Nielsen, 2019). The combination of these approaches provides insight into the dynamics of the human gut microbiota, driving further research directions. Based on the information complied about these three species from the available literature, we created a deterministic mathematical model in order to help organize, clarify, and concatenate existing knowledge. To this end, we suggest and illustrate that a mathematical representation similar to those of chemostats is a natural way to capture the inflow and outflow in the gut.

The three species chosen to be the focus of the model, *B. thetaiotaomicron*, *E. rectale*, and *M. smithii*, play an important role in polysaccharide degradation and the production of buyrate, which both aid in the human gut's ability to absorb nutrients through the epithelial cells (Shoaie, Karlsson, Mardinoglu, et al., 2013). The system of the three microbial species has been considered in previous works, such as Shoaie, Karlsson, Mardinoglu, et al., 2013 and Ji and Nielsen, 2015, which largely informed our knowledge of this system. Based on information complied on these species from the available literature, we created a system of ODEs that tracks the population changes over time in order to obtain a model with a parameter set that achieves a nontrivial, oscillatory steady state. By creating a mathematical model based on the interactions of these species, we have analyzed their interactions and identified aspects of this system that should be further explored through empirical investigation.

Due to the limited availability of data for a more rigorous parameter estimation, our sensitivity analysis took on additional importance. Through the results of the sensitivity analysis using

first- and total-order Sobol' indices, we more narrowly identified specific links in the microbial food web that would be fruitful targets for additional empirical work. Specifically, we identified the parameters β_B, β_{E_1}, β_{E_2}, β_{M_1}, β_{M_2}, and q as being largely significant in contributing to the variance of the model output, including higher-order interactions among these parameters. With these results, we suggest that estimates of these significant parameters be obtained through laboratory experimentation in order to capture these values to a higher degree of precision and accuracy.

These significant β parameters correspond to the growth rates of the three microorganisms when supported on a medium containing specific nutrients. Experiments should focus on cultivating these microorganisms in isolation in germ-free mice, as done previously in the experiment found in Shoaie, Karlsson, Mardinoglu, et al., 2013, with only one nutrient. Despite the fact that these microorganisms are able to be cultured in isolation of other microorganisms, they may not be able to be sustained on one single nutrient, but rather require the presence of additional substrates. In this case, nonlinear effects from these secondary nutrients would factor in to the resulting estimates of the growth parameters. Table 9 provides a general outline of the the specific nutrient and microorganism necessary to estimate each parameter. For example, β_B can be experimentally estimated by cultivating *B. thetaiotaomicron* in a medium containing only polysaccharides. The remaining parameters given in Table 9 can be estimated with similar experiments. Additionally, experiments approximating the rate of flow of the digestive system's fluids would largely inform the estimate for the true value of q.

Parameter	Microorganism	Nutrient
β_B	*B. thetaiotaomicron*	polysaccharides
β_{E_1}	*E. rectale*	acetate
β_{E_2}	*E. rectale*	polysaccharides
β_{M_1}	*M. smithii*	acetate
β_{M_2}	*M. smithii*	CO_2 and H_2

Table 9: Table of microorganism growth rate parameters suggested to be estimated experimentally.

Due to the possibility of correlation among model parameters, variance-based sensitivity analyses specifically for correlated parameters should be explored and applied to this system, such

as the methods discussed in Iooss and Prieur, 2019 and Rabitz, 2010. In regards to specifying the prior distributions on parameters, further analyses should include testing differing prior parameter distributions in order to compute the Sobol' indices, especially to explore the amount of variability in results based on the chosen parameter distributions (Saltelli, Ratto, Andres, et al., 2008).

As another further extension of the work compiled in this thesis, our main suggestion is to collect further longitudinal data on this biological system, including all three species *B. thetaiotaomicron*, *E. rectale*, and *M. smithii* and its relevant substrates. With this experimental data, more precise estimates of the model parameters can be achieved if the previously suggested experiments prove to be infeasible or too costly. With this approach, particular attention can be paid to estimating the parameters that were identified as sensitive by the Sobol' indices.

Additionally, further extensions of our model may include relaxing some of the simplistic and potentially biologically unrealistic assumptions of a general chemostat model. Specifically, the rate of volumetric flow through the chemostat can be generalized to reflect aspects of the natural flow of the gut, such as periodically restricted flow, and the framework of a single vessel representation can be extended to consider additional compartments. Aside from the assumptions of a simple chemostat model, future implementations of a similar model can account for absorption rate of substrates within the gut, which is especially important to consider for substances like amino acids. With these potential improvements of our baseline model, additional aspects about the dynamics of this biological system can be uncovered, and these improvements to our model could fuel further research directions related to this system.

REFERENCES

Adamberg, S., Tomson, K., Vija, H., Puurand, M., Kabanova, N., & Visnapuu, T. (2014). Degradation of fructans and production of propionic acid by Bacteroides thetaiotaomicron are enhanced by the shortage of amino acids. *Frontiers in Nutrition*, *1*(December), 1–10. doi:10.3389/fnut.2014.00021

Ayati, B. P. (2019). Considerations for modeling Proteus mirabilis swarming. In *Proteus mirabilis. methods in molecular biology* (pp. 285–296). Humana, New York, NY: Pearson.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., ... Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559–563. doi:10.1038/nature12820.Diet

Devroye, L. (1986). General principles in random variate generation. In *Non-uniform random variate generation* (Chap. 2, pp. 27–82). Springer.

Glass, T. L. & Hylemon, P. B. (1980). Characterization of a pyridine nucleotide-nonspecific glutamate dehydrogenase from Bacteroides thetaiotaomicron. *Journal of Bacteriology*, *141*(3), 1320–1330.

Goldford, J. E., Lu, N., Bajie, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., ... Sanchez, A. (2018). Emergent simplicity in microbial community assembly. *Science*, *474*(361), 469–474.

Guglielmi, G. (2018). How gut microbes are joining the fight against cancer. *Nature*, *557*(7706), 482–484. doi:10.1038/d41586-018-05208-8

Iooss, B. & Prieur, C. (2019). Shapley effects for sensitivity analysis with correlated inputs: Comparisons with sobol' indices, numerical estimation and applications. *International Journal for Uncertainty Quantification*, *9*(5), 493–514. doi:10.1615/Int.J.UncertaintyQuantification.2019028372. arXiv: 1707.01334

Ji, B. & Nielsen, J. (2015). From next-generation sequencing to systematic modeling of the gut microbiome. *Frontiers in Genetics*, *6*(219), 1–9. doi:10.3389/fgene.2015.00219

Kim, M. H. & Kim, H. (2017). The roles of glutamine in the intestine and its implication in intestinal diseases. *International Journal of Molecular Sciences*, *18*(5). doi:10.3390/ijms18051051

Kumar, M., Ji, B., Zengler, K., & Nielsen, J. (2019). Modelling approaches for studying the microbiome. *Nature Microbiology*, *4*(8), 1253–1267. doi:10.1038/s41564-019-0491-9

Lehouritis, P., Cummins, J., Stanton, M., Murphy, C. T., McCarthy, F. O., Reid, G., ... Tangney, M. (2015). Local bacteria affect the efficacy of chemotherapeutic drugs. *Scientific Reports*, *5*, 1–12. doi:10.1038/srep14554

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS Computational Biology*, *13*(5), 1–23. doi:10.1371/journal.pcbi.1005457

Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., ... Gordon, J. I. (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 5859–5864.

Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2009). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Microbiology*, *254*(1), 178–196. doi:10.1016/j.jtbi.2008.04.011.A

Martens, E. C., Lowe, E. C., Chiang, H., Pudlo, N. A., Wu, M., McNulty, N. P., ... Gordon, J. I. (2011). Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biology*, *9*(12). doi:10.1371/journal.pbio.1001221

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*(5147), 641–646. doi:10.1126/science.263.5147.641

Rabitz, H. (2010). Sixth International Conference on Sensitivity Analysis of Model Output Global Sensitivity Analysis for Systems with Independent and / or Correlated Inputs. *Soins Aides-soignantes*, *2*(6), 7587–7589. doi:10.1016/j.sbspro.2010.05.131

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... Tarantola, S. (2008). In *Global sensitivity analysis. the primer* (Chap. 1-6, pp. i–xi). John Wiley Sons,

Ltd. doi:10.1002/9780470725184.fmatter. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470725184.fmatter

Scaldaferri, F., Nardone, O., Lopetuso, L. R., Petito, V., Bibbò, S., Laterza, L., … Gasbarrini, A. (2013). Intestinal gas production and gastrointestinal symptoms: from pathogenesis to clinical implication. *European review for medical and pharmacological sciences*, *17*(Suppl 2), 2–10.

Shoaie, S., Karlsson, F., Mardinoglu, A., Nookaew, I., Bordel, S., & Nielsen, J. (2013). Understanding the interactions between bacteria in the human gut through metabolic modeling. *Scientific Reports*, *3*, 1–10. doi:10.1038/srep02532

Smith, H. L. & Waltman, P. (2008). *The Theory of The Chemostat: Dynamics of Microbial Competition*. Cambridge University Press.

Whisner, C. M. & Athena Aktipis, C. (2019). The role of the microbiome in cancer initiation and progression: how microbes and cancer cells utilize excess energy and promote one another's growth. *Current Nutrition Reports*, *8*(1), 42–51. doi:10.1007/s13668-019-0257-2

Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., … Gordon, J. I. (2003). A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science*, *299*(5615), 2074–2076. doi:10.1210/jcem-10-10-1361

APPENDIX A: ALGORITHM FOR FIRST- AND TOTAL-EFFECT SOBOL' INDICES FOR MODEL PARAMETERS

Algorithm 1: First- and total-effect Sobol' indices for model parameters

Result: S: $k \times R$ matrix of Sobol' indices for every parameter k and for each result R.

Input: f: model with R outputs; N: number of simulations; k: number of parameters; idx: 1 for first-order effects, 0 for total-order effects.

$M \leftarrow$ Generate an $N \times 2k$ matrix using a quasi-random number generator from each parameter's underlying distribution;

$A \leftarrow M[N, 1:k]$; $B \leftarrow M[N, (k+1):2k]$;

$S \leftarrow$ initialize a $k \times R$ matrix;

for $i = 1, \ldots, k$ **do**

    $C_i \leftarrow B$, except replace the ith column with the ith column of A;

    $y_A \leftarrow f(A)$; $y_B \leftarrow f(B)$; $y_{C_i} \leftarrow f(C_i)$;

    if $idx == 1$ **then**

        $S \leftarrow$ Compute first-order Sobol' indices based on equation 6;

    else

        $S \leftarrow$ Compute total-order Sobol' indices based on equation 7;

    end

end

Return S

APPENDIX B: TABLES OF FIRST- AND TOTAL-ORDER SOBOL' INDICES FOR MODEL PARAMETERS

	B	E	M	a	h	p
β_a	0.00	0.00	0.00	0.00	0.00	0.00
β_B	0.98	0.02	0.01	0.02	0.37	0.13
β_{E_1}	0.00	0.00	0.00	0.04	0.00	0.00
β_{E_2}	0.00	0.82	0.01	0.02	0.10	0.03
β_{h_1}	0.00	0.00	0.00	0.00	0.00	0.00
β_{h_2}	0.00	0.00	0.00	0.00	0.00	0.00
β_{M_1}	0.00	0.00	0.04	0.04	0.06	0.00
β_{M_2}	0.00	0.00	0.65	0.02	0.17	0.00
β_p	0.00	0.00	0.00	0.00	0.00	0.00
γ_a	0.00	0.00	0.00	0.00	0.00	0.00
γ_B	0.00	0.00	0.00	0.00	0.00	0.00
γ_h	0.00	0.00	0.00	0.00	0.00	0.00
γ_p	0.00	0.00	0.00	0.00	0.01	0.00
μ_{aE}	0.00	0.00	0.00	0.00	0.00	0.00
μ_{aM}	0.00	0.00	0.00	0.00	0.00	0.00
μ_{hM}	0.00	0.00	0.00	0.00	0.00	0.00
μ_{pB}	0.00	0.00	0.00	0.00	0.00	0.00
μ_{pE}	0.00	0.00	0.00	0.00	0.00	0.00
q	0.00	0.01	0.02	0.07	0.02	0.09
Total	0.99	0.85	0.74	0.23	0.75	0.24

Table 10: Table of model parameters' estimated first-order Sobol' indices for each output, B, E, M, a, h, and p, using $N = 2^{18}$ simulations. These indices were calculated based on equation (6).

	B	E	M	a	h	p
β_a	0.00	0.00	0.00	0.00	0.00	0.00
β_B	0.99	0.13	0.11	0.33	0.56	0.80
β_{E_1}	0.00	0.01	0.00	0.56	0.01	0.04
β_{E_2}	0.01	0.97	0.04	0.22	0.18	0.49
β_{h_1}	0.00	0.00	0.00	0.00	0.00	0.00
β_{h_2}	0.00	0.00	0.00	0.00	0.00	0.00
β_{M_1}	0.00	0.00	0.05	0.58	0.10	0.01
β_{M_2}	0.00	0.00	0.88	0.19	0.25	0.00
β_p	0.00	0.00	0.00	0.00	0.00	0.00
γ_a	0.00	0.00	0.00	0.00	0.00	0.00
γ_B	0.00	0.00	0.00	0.00	0.00	0.00
γ_h	0.00	0.00	0.01	0.00	0.00	0.00
γ_p	0.01	0.01	0.00	0.00	0.01	0.00
μ_{aE}	0.00	0.00	0.00	0.00	0.00	0.00
μ_{aM}	0.00	0.00	0.00	0.00	0.00	0.00
μ_{hM}	0.00	0.00	0.04	0.01	0.00	0.00
μ_{pB}	0.00	0.00	0.00	0.00	0.00	0.00
μ_{pE}	0.00	0.00	0.00	0.00	0.00	0.00
q	0.01	0.05	0.11	0.62	0.17	0.72
Total	1.02	1.17	1.25	2.48	1.29	2.09

Table 11: Table of model parameters' estimated total-order Sobol' indices for each output, B, E, M, a, h, and p, using $N = 2^{18}$ simulations. These indices were calculated based on equation (7).

APPENDIX C: JULIA CODE FOR FIRST- AND TOTAL-ORDER SOBOL' INDICES

```
"""
sobol(f, N, u0, tspan, t, k, θ, index = "first-order")

Creates a matrix of Sobol' sensitivities where each row contains
the sensitivity indices for a parameter across different model
outputs.

 f - system of ODEs
 N - number of simulations (can vary from hundreds to thousands)
 u0 - vector of initial conditions for ODE
 tspan - time span
 k - vector of scale parameters from Gamma(k_i, θ_i) for all
     model parameters
 θ - vector of shape parameters from Gamma(k_i, θ_i) for all
     model parameters
 index - default is the first order index, but can take on the
         other value "total-order"
 t - specified time in tspan to evaluate each model variable
"""

using Sobol
using Distributions
function sobol(f, N, u0, tspan, k, θ, index = "first-order",
                t = 5)
    K = length(k)
```

```
    vars = length(u0)
# quasi-randomly selected points
    S = zeros(N,2*K)
    s = SobolSeq(2*K)
    θ = [θ;θ]
    k = [k;k]

    S = zeros(N,2*K)
    for sim in 1:N
        S[sim,:] = next!(s)'
        for parameter in 1:2*K
            S[sim,parameter] = quantile(Gamma(k[parameter],
                                θ[parameter]), S[sim,parameter])
        end
    end

    A = S[:, 1:K] # accounts for half of the random sample
    B = S[:, (K+1):2*K] # accounts for second half of the
                        # random sample

    sums = zeros(4,vars)
    ya = solveODE(f, u0, vars, tspan, t, A, N)
    yb = solveODE(f, u0, vars, tspan, t, B, N)
    sumC = zeros(K,vars) # (parameter, variable)

    for p in 1:K # iterates through changing one column of C at
                 # a time (parameters)
```

```
    C = Ci(p,A,B)
    yc = solveODE(f, u0, vars, tspan, t, C, N)

    for v in 1:vars
        if index == "first-order"
            sumC[p,v] = yc[:,v]'*ya[:,v]/N
        else
            sumC[p,v] = yc[:,v]'*yb[:,v]/N
        end
    end
end

sens = zeros(K,vars) # (parameter, variable)


for v in 1:vars
    sums[1,v] = ya[:,v]'*ya[:,v]/N
                # sum(ya^2)/N
    sums[2,v] = (sum(ya[:,v])/N)^2 # [sum(ya)/N]^2
    sums[3,v] = sums[1, v] - sums[2,v]
                # denominator for sensitivity calculation
    sums[4,v] = sum(yb[:,v])*sum(ya[:,v])/N.^2
                # sum(ya)*sum(yb)/N^2

    for p in 1:K
        if index == "first-order"
```

```
                sens[p,v] = (sumC[p,v] - sums[4,v])/sums[3,v]
            else
                sens[p,v] = 1 - (sumC[p,v] - sums[2,v])/sums[3,v]
            end
        end
    end
    return sens # matrix of sensitivity indices
end
```

```
"""
Ci(i, A, B)

Creates a new matrix C by replacing the ith column of matrix B
with the ith column of matrix A.

 i – column number to be changed in matrix B
 A – matrix containing half of the pseudo–random samples
 B – matrix containing the other half of the pseudo–random
     samples
"""

function Ci(i, A, B)
    C = zeros(size(B))
    for j in 1:size(A)[2]
        if j == i
            C[:,i] = A[:,i]
        else
            C[:,j] = B[:,j]
        end
    end
    return C
end
```

```
"""
solveODE(f, u0, vars, tspan, t, M, N)

Solves the system of ODEs f and creates a matrix of model
evaluations for each simulation and output variable. Each row
corresponds to model solutions at a given time t for one
simulation.

 f – system of ODEs
 u0 – vector of initial conditions for the ODEs
 vars – number of output variables
 tspan – time span
 t – specified time in tspan to evaluate each model variable
 M – matrix with rows as parameter values and columns as
      simulations
 N – number of simulations (can vary from hundreds to thousands)
"""

function solveODE(f, u0, vars, tspan, t, M, N)
    y = zeros(N,vars) # (simulation, variable)
    for sim in 1:N
        p = M[sim,:]
        prob = ODEProblem(f,u0,tspan,p)
        sol = solve(prob)
        for variable in 1:vars
            y[sim,variable] = sol.u[t][variable]
        end
```

```
    end
    return y # matrix of model outputs
end
```