FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

FLOUR POWER:

MODELING THE GROWTH AND DECAY OF SOURDOUGH STARTERS

By

JONAH KELEMAN SMITH

A Thesis submitted to the
Department of Mathematics
in partial fulfillment of the requirements for graduation with
Honors in the Major

2023

Jonah Keleman Smith defended this thesis on April 27, 2023.
The members of the supervisory committee were:

Nicholas Cogan
Professor Directing Thesis

Monica Hurdal
Committee Member

Nancy Rogers
Committee Member

Nora Underwood
Committee Member

To the Yeastie Boys (March 23, 2020 - present)

ACKNOWLEDGMENTS

This project would not have been possible without immense and generous help from the following people. My deepest thanks go to Dr. Michael Gänzle (University of Alberta), who so kindly answered a series of questions about yeast symbiosis and clarified my understanding of bacteriofungal interactions in sourdough; to Jennifer Anderson, who as the math department undergraduate advisor encouraged me to seek out this and other exceptional opportunities and to incorporate a love for baking into my studies; and of course to Dr. Nick Cogan, who guided me through eighteen incredible months of research and learning, invited me to participate in graduate student lab meetings, and never once doubted my ability to bring this project to life.

An extra special thanks goes to anyone who has ever helped maintain or distribute open-source software, casually or full-time. Without the accessibility afforded by Jupyter Notebooks, Github, Python, LaTeX, etc., there would be less fun and meaningful research in the mathematics and arts.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

The following symbols and abbreviations are used consistently throughout the document to prevent ambiguity in meaning.

C.	*Candida*
cfu	colony forming unit
Fl.	*Fructilactobacillus*
g	gram
h	hour
K.	*Kazachstania*
LAB	lactic acid bacteria
Lb.	*Lactobacillus*
$\mathbb{N}$	natural numbers (not including 0)
mol	mole
$\mathbb{R}$	real numbers
$\mathbb{R}_{\geq 0}$	non-negative real numbers
$\mathbb{R}_{+}$	positive real numbers
S.	*Secale*
Sc.	*Saccharomyces*
T.	*Triticum*

ABSTRACT

Sourdough fermentation is a complex biochemical process governed by interactions at the microscopic scale between lactic acid bacteria (LAB) and yeasts. The most commonly encountered species of each of these organisms, *Fructilactobacillus sanfranciscensis* (formerly *Lactobacillus sanfranciscensis*) and *Kazachstania humilis* (formerly *Candida milleri*) exhibit several characteristics of symbiosis that give sourdough starters their leavening power. We seek to model the interactions between these two forces as they consume, produce, and partition different resources found in wheat flour. Using a combination of Gompertz and Michaelis-Menten models for population kinetics, we create a system of non-linear ordinary differential equations to map the continuous growth of sourdough between starter refreshment and a discrete process by which we can simulate back-slopping. The combined model accurately captures many qualities of sourdough fermentation and enables further research and optimization of the model.

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 Motivation for Study

Bread is a staple food in almost every part of the world, having been baked and broken by humanity for millennia [1]. The simplest recipe calls for the baker to mix flour, water, salt, and yeast together, let the mixture ferment until it has developed a desirable structure and taste, and bake it at a high temperature until the loaf reaches optimal aroma and color. Sourdough breads are especially popular for their superior anti-spoilage properties and improved taste over non-sourdough baked goods [2]. Because sourdough bread does not require use of commercial yeast, many amateurs and home bakers have become interested in using sourdough as a more cost effective baking technique. It is for these reasons that the study of sourdough is of incredible biological and economic importance.

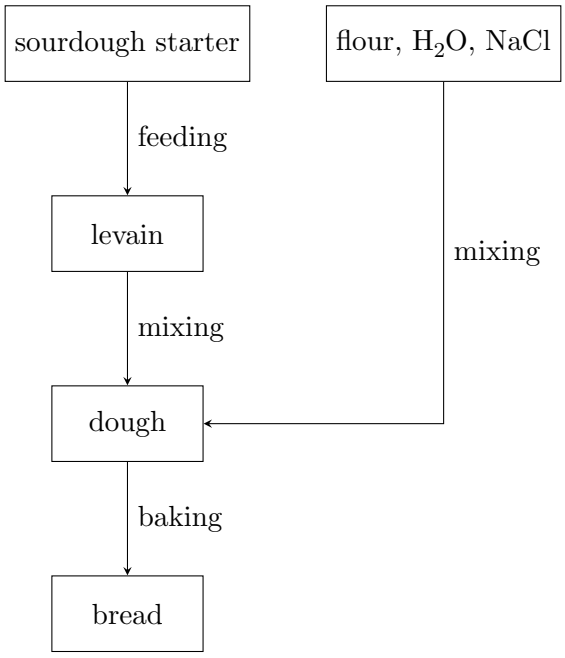

Figure 1.1: A typical traditional process for baking sourdough bread, from starter and raw ingredients to final product. Levain is occasionally omitted [3], [4].

One of the biggest challenges faced by bakers at every skill level is the conundrum of finding an appropriate way to maintain the health of a sourdough starter. Many cookbooks and guides suggest using apples, milk, and other non-flour ingredients to kick-start the fermentation process by introducing wild yeasts found in or on the surfaces of these foods [5, p. 214]. Others prefer to cultivate a more "pure" starter, using only distilled water and wheat (*T. aestivum*) or rye (*S. cereale*) flours. In any case, a sourdough starter is almost always a back-slopped fermented food, meaning a portion of fully fermented starter is used to begin the fermentation process for the next generation or iteration of the starter [3]. Doing this takes bacteria and yeasts which are already adapted for the particular flour and environment into a fresher source of food, ensuring continuous population growth and a constant source of ready-to-use starter for the baker. The portion of starter which is used to begin the next iteration of fermentation is typically referred to as the "carryover" or "inoculum." In general, we will refer to the ratio by mass of inoculum to gross starter as an "inoculation ratio," representing it with p. The ratio by mass of water which constitutes a starter is also a generally important figure, typically referred to as "hydration."[1] This quantity will be represented with the symbol w.

We are interested in understanding how the type of flour used, the times spent between back-slopping, and the inoculation ratio can contribute to varied growth rates and patterns in sourdough cultures. We will seek answers to these questions by constructing and analyzing mathematical models designed to predict sourdough growth.

1.2 Scope of Inquiry

The world of sourdough is vast and diverse. By varying the type of flour used, ratio by mass of water content, and ambient fermentation temperature, bakers can create diverse collections of sourdough starters suited for different needs [6]. For the purposes of this model, we will only examine starters fed with identical types and amounts of flour and water from day to day, fermented at consistent ambient temperatures. Sourdoughs can also be segregated by the technology used to apply them to a baked good, in one of three types. Type I sourdoughs are characterized by daily or otherwise regularly spaced back-slopping, as constant refreshment of the microbiome is necessary to prevent spoilage. While sourdoughs of Types II and III are economically important products for

[1] Technically speaking, bakers use the term "hydration" to refer to the ratio by mass of water to flour in a recipe or starter, not the ratio of water to total mass. As one could imagine, this often results in ratios over 100%. To avoid this confusion, we ignore baker's percentages of weight-to-flour weight and instead focus on weight-to-total weight ratios.

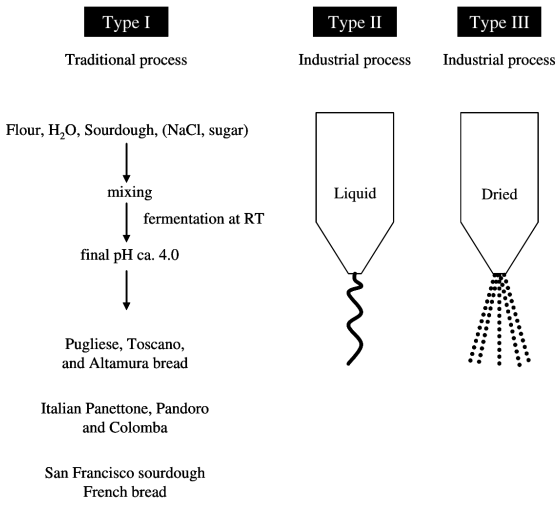

Figure 1.2: The three types of sourdough. Type I sourdough is usually fermented at ambient or lower temperatures ($< 30°$C), with back-slopping occurring at fixed intervals of 12-24 h. Types II and III refresh every 2-5 days, and often ferment at higher temperatures ($> 30°$C) [3].

their industrial applications, they do not contain enough yeast to leaven dough on their own and must be supplemented with additional *Sc. cerevisiae* to have any leavening effect [3]. Therefore, we will limit the scope of this thesis to studying Type I sourdoughs.

1.3 Previous Research

A cornerstone of the modern study of demography is Gompertz's law, one of the earliest applications of the exponential growth and decay to describe human populations. Initially used by actuaries and statisticians to model the rise in mortality rates as a function of age [7], Gompertz's law has ironically found its way into the biological sciences as a simple model to describe how populations grow exponentially over time. Most modern studies on both mortality and population dynamics will draw in one form or another from the derivations that Benjamin Gompertz made almost two centuries ago.

Building upon Gompertz's writings is another of the fundamental pieces of literature on quantifying growth of microorganisms: Nobel Laureate Jacques Monod's 1949 treatise *The Growth of Bacterial Cultures*. Monod [8] separates the growth cycle of bacterial cultures into six distinct steps: an initial lag phase, an exponential growth phase with a positive growth rate, a constant growth phase, another exponential growth phase with a negative growth rate, a stagnant plateau phase with no growth, and then a final phase of decay with largely constant rate. This model also

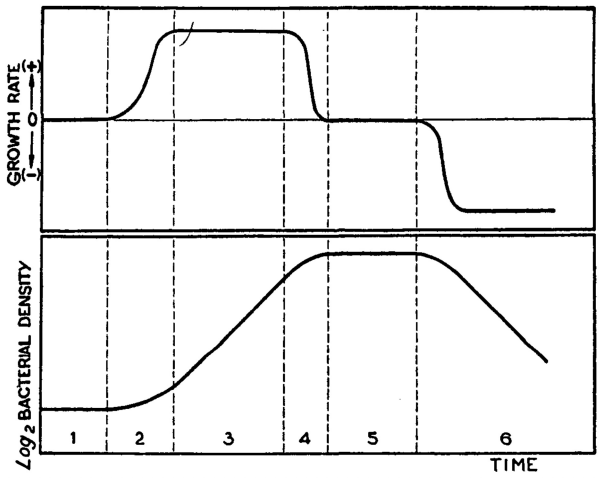

Figure 1.3: A graph of the phases of growth from Monod. The top graph shows the growth rate and the bottom graph shows bacterial density on a log-2 scale [8].

ascribes exponential models to bacterial growth and serves as an inspiration to most future studies, including this one. A key difference between Gompertz's and Monod's work is the separation of stages of growth and decay. It is more difficult to fit a continuous exponential function which satisfies the requirements at each of the six stages of growth, but combining ideas from both models can prove successful.

Based on Gompertz's and Monod's framework, several studies in the latter half of the twentieth century have tried to pin down the mechanics of sourdough. These studies analyze the growth and decay of starters from a lens that relies heavily on the writings of these two individuals, building a specialized LAB fermentation model from a generic one.

In 1992 and 1995, Zweitering and Wijztes collaborated to examine the effects of water activity, temperature, and pH on different *Listeria* and LAB cultures. This study suggests that the simple Gompertz model can be combined with modified forms of a Belehradek-type (square root) model for comparing temperature and pH to bacterial growth [9]–[11]. While we will assume temperature and pH to be relatively unimportant variables later in the thesis, this study shows the adaptability and usefulness of a Gompertz-based model to describe population dynamics in food microbiology.

In 2005, Neysens and De Vuyst create a system of ordinary differential equations to describe cell population growth, separate maltose and fructose consumption, lactic and acetic acid production, and bacteriocin levels. The duo once again used a modified logistic growth equation with a modified Belehradek-type model for parameter fitting the maximum growth rate [2]. The models described in this paper incorporate more variables than we are interested in studying, but a focus on metabolism

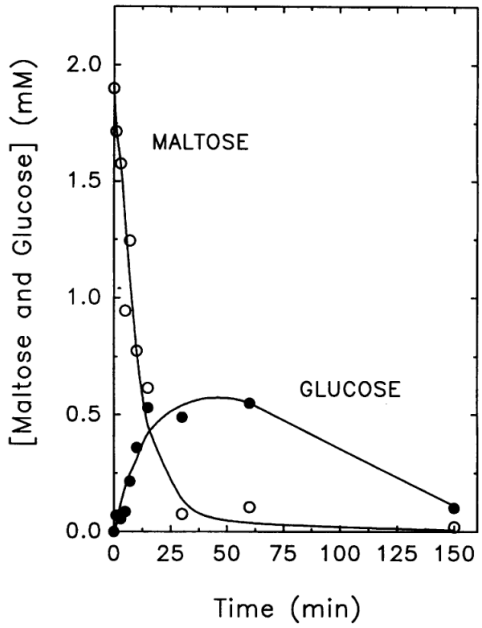

Figure 1.4: After consuming maltose, *Fl. sanfranciscensis* isolated in a dish will produce excess glucose as waste. Approximately half of all generated glucose is excreted as waste [13]. Notice how the maltose curve seems to follow an exponential decay.

and sugar consumption inspires much of the model derived in the second chapter of this thesis. Inclusion of an exponential model as a baseline indicates once again that Gompertz is an appropriate point from which to begin.

1.4 Sourdough Kinetics

Lactic acid fermentation is a biological process involving two main groups of microorganisms, lactic acid bacteria (LAB) and yeasts. The two species interact in several ways: sharing components on their food webs, producing favorable environments for growth, and preventing competing species from colonizing the sourdough culture. *Fl. plantarum*, for instance, is able to produce bacteriocins and antifungal substances which ward off other hostile species of bacteria and yeast [1].

There are many reasons we might consider the cohabitation of LAB and yeasts to be a symbiosis. While the propensity for LAB to create bacterioricins and acidic environments which just so happen to favor yeast growth are great choices, perhaps the most interesting factor is the interconnectedness of the two food webs [12].

The most prevalent species of LAB in sourdough is the strain *Fl. sanfranciscensis*, named for San Francisco's sourdough scene. This LAB prefers to consume maltose and other disaccharides,

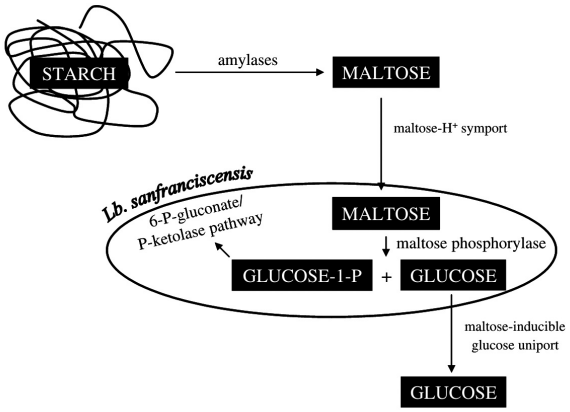

Figure 1.5: *Fl. sanfranciscensis* prefers to consume maltose, formed from tough starches in wheat flour. One can think of this consumption as a separating of the two glucose molecules joined to create a maltose molecule [3].

excreting glucose as a byproduct of this metabolism [3], [14]. Maltose is found in large quantities in grain products, explaining the bacteria's adaptation to flour media [15], [16]. As much as this heterofermentative LAB loves to eat maltose, no other organisms in the sourdough jar are able to process this disaccharide. The waste product of this consumption, glucose, just happens to be exactly what yeast love eating to reproduce and grow [13]. It is because of this system where one organism excretes a waste product for the other that we can think of their dynamic as a symbiosis.

CHAPTER 2

BACK-SLOPPING MODELS

We seek to accurately model the growth of a sourdough starter before incorporation into the final dough. For this, we will need to combine a continuous model of molecular growth kinetics with a discrete model of the back-slopping process. Back-slopping is usually done by taking a small portion of the previous day's fully fermented starter and combining it with a mixture of raw flour and distilled water. This new mixture is left to ferment for some time, with the process repeated at regularly scheduled time intervals.

The amount of old starter used to create the new starter, the ratio of fresh flour to fresh water, the temperature at which the mixture ferments, the type of flour used to refresh the starter, and the time spent between feedings all vary between a diverse array of starters and bakers. This model will incorporate variables of time (t), concentration of LAB in cfu (B), concentration of yeast in cfu (X), concentration of glucose in grams of glucose per kilogram of flour or g kg^{-1} (G), and concentration of disaccharides[1] in g kg^{-1} (D). We will restrict all five variables (four state variables and one time variable) to $\mathbb{R}_{\geq 0}$.

Initially we will consider parameters of p and w to represent carryover and hydration respectively. However, we will need to add more parameters as the model becomes more complex. Because p and w represent proportions of mass to total mass, they must obey the following limitations:

$$p \in (0, 1) \tag{2.1}$$

$$w \in (0, 1 - p). \tag{2.2}$$

We would like to create an abstraction of the process using differential equations to simulate the continuous growth a starter experiences between feedings and a recurrence relation to simulate the discrete process of feeding a starter new flour and discarding the excess. We will begin by assuming the amounts of glucose and disaccharides in a given gram of raw (un-fermented) flour are consistent from batch to batch, that all mixtures are thoroughly mixed, and that yeast and bacteria populations from the environment are consistent from feeding to feeding. We also have initial

[1] We will from here on treat sucrose and maltose, the two most commonly encountered disaccharides in wheat flour, as one entity.

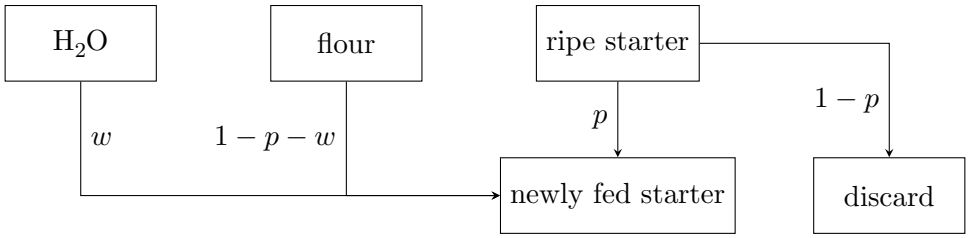

Figure 2.1: The back-slopping process using an inoculation ratio of p and hydration ratio of w. Note that these are not equivalent to "baker's percentages." Here we assume a constant size of 1 is maintained for the starter between generations.

conditions to use as building blocks for model construction: $B(0) = B_0$, the initial concentration of bacteria; $G(0) = G_0$, the initial concentration of glucose; $X(0) = X_0$, the initial concentration of yeast; and $D(0) = D_0$, the initial concentration of disaccharides.

2.1 Model Derivation

The simplest model for bacterial growth as outlined by Monod serves as a basis for the model we derive for sourdough growth. A purely exponential model for nutrient consumption and organism growth takes the form of

$$\frac{dB}{dt} = \alpha_B B \tag{2.3}$$
$$\frac{dD}{dt} = -\alpha_B D \tag{2.4}$$
$$\frac{dG}{dt} = \alpha_B B - \alpha_X X \tag{2.5}$$
$$\frac{dX}{dt} = \alpha_X X. \tag{2.6}$$

We must assume for the sake of non-triviality that both growth terms $\alpha \in \mathbb{R}_+$. This crude model gets some things right, such as the relationship between glucose production from LAB and preferential glucose consumption by yeasts, but it fails to accurately model the metabolism of microorganisms in a thoroughly mixed solution. For this to successfully describe a sourdough culture, we need to incorporate some model of molecular kinetics.

2.1.1 Michaelis-Menten Kinetics

The Michaelis-Menten model is an appropriate choice for approximating molecular kinetics where molecules of substrate and receptors exist at a one-to-one ratio in a solution. The Michaelis-Menten model is what is known as a "limiting rate" law, showing the velocity of a reaction as

it approaches its maximum asymptotic rate. A useful characteristic of this type of model is that at low concentrations of nutrients, the reaction speed is mostly proportional to the concentration and grows rapidly, while at high concentrations the reaction speed slows down as all receptors are "busy" and saturated with nutrients. For this reason, the Michaelis-Menten model is often also said to have a property of "saturation." [17]

The equation takes two parameters, K_{max} and K_n, which represent the maximum reaction speed and the amount of substrate needed to reach half of that speed, respectively. For some microbial concentration C and nutrient concentration N, we can represent this relationship generally with

$$\frac{dC}{dt} = \frac{K_{max}N}{K_n + N}. \tag{2.7}$$

When applied to the specific case of our model, we will multiply each term showing consumption with an adaptation of the Michaelis-Menten model. As each metabolic process is unique, we will assign K_D as our "halfway" parameter governing disaccharide consumption and K_G as the parameter governing glucose consumption. We will also absorb the K_{max} term for each relationship into its neighboring α term to avoid redundancy. Our model now shows that LAB and yeast populations can be governed by the relationships

$$\frac{dB}{dt} = \alpha_B B \frac{D}{K_D + D} \tag{2.8}$$

$$\frac{dD}{dt} = -\alpha_B B \frac{D}{K_D + D} \tag{2.9}$$

$$\frac{dG}{dt} = \alpha_B B \frac{D}{K_D + D} - \alpha_X X \frac{G}{K_G + G} \tag{2.10}$$

$$\frac{dX}{dt} = \alpha_X X \frac{G}{K_G + G}. \tag{2.11}$$

To avoid a situation where the denominator of any fraction is zero, we will restrict each parameter $K \in \mathbb{R}_+$. This new implementation works relatively well for approximating transient solutions in the short term but fails to account for two crucial elements. The first is LAB and yeast mortality. Leave a sourdough starter on the counter for a week without feeding and refreshing, and aggressive species of bacteria and fungi will overtake the friendly ones. Another consideration we must make are the ratios of nutrient to organism we can expect to be converted. We will acknowledge these two factors with new parameters to strengthen our model.

2.1.2 Decay and Yield Coefficients

Decay. With no decay term, there exist some non-trivial steady states of our model. In other words, the model suggests that an unfed starter with the right amount of flour and certain parameter values could exist in perpetual homeostasis after an arbitrarily long time. Let us look at the concentration of LAB, with our model as it currently is (i.e., no decay term). Without new nutrients added to the sourdough starter, we must assume disaccharide concentrations decrease and approach 0 as time progresses, that is

$$\lim_{t\to\infty} D(t) = 0.$$

A steady state occurs when the first time derivative of motion equals 0. Our LAB equation has a steady state when

$$\frac{dB}{dt} = \alpha_B B \frac{D}{K_D + D} \tag{2.12}$$

$$0 = \alpha_B B \frac{D}{K_D + D} \tag{2.13}$$

$$\alpha_B = 0 \quad \text{or} \quad B = 0 \quad \text{or} \quad D = 0. \tag{2.14}$$

We have already determined that $\alpha_B \neq 0$. Therefore, stability can only occur when either B or D are equal to 0. But these are technically mutually inclusive steady states, so there exist some steady states where $D = 0$ and $B \neq 0$. This implies that in a sourdough culture with a lack of nutrients, LAB concentration could be any positive real number at a steady state.

This has no basis in reality, so we need to implement some microbial decay term. Taking a page from Gompertz's book, we model population mortality with a simple exponential term γ. We will restrict $\gamma \in \mathbb{R}_+$ to avoid triviality and unbounded growth. Our new equation for LAB population dynamics looks like

$$\frac{dB}{dt} = \alpha_B B \frac{D}{K_D + D} - \gamma_B B.$$

It is easy to calculate our set of steady states by once again setting the left hand side to 0:

$$\frac{dB}{dt} = \alpha_B B \frac{D}{K_D + D} - \gamma_B B \tag{2.15}$$

$$0 = B\left[\alpha_B \frac{D}{K_D + D} - \gamma_B\right] \tag{2.16}$$

$$B = 0 \quad \text{or} \quad D = \frac{\gamma_B K_D}{\alpha_B - \gamma B}. \tag{2.17}$$

But this second steady state is not feasible, since at least one of K_D or γ_B would need to be 0. Therefore, the only steady state with a decay term is when $B = 0$, which matches our intuition. We can show the same process for our yeast term and revise our system of equations:

$$\frac{dB}{dt} = \alpha_B B \frac{D}{K_D + D} - \gamma_B B \tag{2.18}$$

$$\frac{dD}{dt} = -\alpha_B B \frac{D}{K_D + D} \tag{2.19}$$

$$\frac{dG}{dt} = \alpha_B B \frac{D}{K_D + D} - \alpha_X X \frac{G}{K_G + G} \tag{2.20}$$

$$\frac{dX}{dt} = \alpha_X X \frac{G}{K_G + G} - \gamma_X X. \tag{2.21}$$

Yield. The last addition we must make to our system of differential equations is a yield coefficient. This parameter describes the ratio of biomass units created to nutrient units consumed and can be thought of as an efficiency measure. There should be one such parameter for each interaction between an organism and its nutrition, so we will divide each growth term by a unique yield coefficient. It is important to also note that the yield term corrects issues between unit equivalencies in the previous iterations of this model. The equations for the changes in disaccharides and glucose have had incorrect units up to now. We now have:

$$\frac{dB}{dt} = \alpha_B B \frac{D}{K_D + D} - \gamma_B B \tag{2.22}$$

$$\frac{dD}{dt} = -\frac{\alpha_B}{Y_D} B \frac{D}{K_D + D} \tag{2.23}$$

$$\frac{dG}{dt} = \frac{\alpha_B}{Y_D} B \frac{D}{K_D + D} - \frac{\alpha_X}{Y_G} X \frac{G}{K_G + G} \tag{2.24}$$

$$\frac{dX}{dt} = \alpha_X X \frac{G}{K_G + G} - \gamma_X X, \tag{2.25}$$

where each Y is a yield term.

2.1.3 Back-Slopping

Of paramount importance to the applicability of this model is the ability to simulate a back-slopping event, colloquially referred to as a feeding. We will think of the model so far as governing the kinetics of exactly one time interval between feedings given the initial conditions presented at .

To make exactly 1 gram of new starter, the baker will take p grams of ripe sourdough and combine it with w grams of water and $1 - w - p$ grams of flour. This process occurs every $t = h$ hours. We can think of the back-slopping model as a series of difference equations where initial conditions for cycle $n + 1$ are scaled down final conditions from the end of cycle n:

$$B_{i+1}(0) = pB_i(h) \tag{2.26}$$

$$D_{i+1}(0) = pD_i(h) \tag{2.27}$$

$$G_{i+1}(0) = pG_i(h) \tag{2.28}$$

$$X_{i+1}(0) = pX_i(h) \tag{2.29}$$

However, the above equations are only abstractions for taking old starter and placing it in a new jar. We need need to appropriately account for the action of placing fresh flour in the jar and feeding the starter. This flour addition will add to the nutrient density, so we will increment $D_{i+1}(0)$ and $G_{i+1}(0)$ by their representative initial conditions. Doing so also introduces new amounts of ambient bacteria and yeast from the raw flour into the mixture, so we similarly add proportionate amounts of those initial conditions to the counts of LAB and yeast. Now we have a more realistic model for feeding:

$$B_{i+1}(0) = pB_i(h) + (1-p-w)B_0(0) \tag{2.30}$$

$$D_{i+1}(0) = pD_i(h) + (1-p-w)D_0(0) \tag{2.31}$$

$$G_{i+1}(0) = pG_i(h) + (1-p-w)G_0(0) \tag{2.32}$$

$$X_{i+1}(0) = pX_i(h) + (1-p-w)X_0(0) \tag{2.33}$$

For each of these four equations, it can be helpful to think of the first addend as representing nutrient or microbe addition "from the starter" and the second term as the addition "from the flour." Since in this model we are only feeding our starter with distilled water[2], it is not possible for nutrients or organisms to enter the starter through water addition and there is no need to include a third "water" term in each of these difference equations.

[2]Many bakers choose to only feed their starters with distilled water as this reduces the chances of disrupting the pH balance.

2.1.4 Full Model

Our final model combines the quartets of differential equations and difference equations to model growth and back-slopping:

$$\frac{dB_i}{dt} = \alpha_B B_i \frac{D_i}{K_D + D_i} - \gamma_B B_i \tag{2.34}$$

$$\frac{dD_i}{dt} = -\frac{\alpha_B}{Y_D} B_i \frac{D_i}{K_D + D_i} \tag{2.35}$$

$$\frac{dG_i}{dt} = \frac{\alpha_B}{Y_D} B_i \frac{D_i}{K_D + D_i} - \frac{\alpha_X}{Y_G} X_i \frac{G_i}{K_G + G_i} \tag{2.36}$$

$$\frac{dX_i}{dt} = \alpha_X X_i \frac{G_i}{K_G + G_i} - \gamma_X X_i \tag{2.37}$$

$$B_{i+1}(0) = pB_i(h) + (1 - p - w)B_0(0) \tag{2.38}$$

$$D_{i+1}(0) = pD_i(h) + (1 - p - w)D_0(0) \tag{2.39}$$

$$G_{i+1}(0) = pG_i(h) + (1 - p - w)G_0(0) \tag{2.40}$$

$$X_{i+1}(0) = pX_i(h) + (1 - p - w)X_0(0). \tag{2.41}$$

When describing the first four equations of the model, we will simplify notation by dropping the subscript denoting generation, i.e., "$B_i(t)$" will simply be referred to as "$B(t)$". The following tables give a range of values for each parameter and variable listed above.[3]

Parameter List			
Parameter	Range	Units	Biological Significance
α_B	$\mathbb{R}_+$	h^{-1}	LAB growth
α_X	$\mathbb{R}_+$	h^{-1}	yeast growth
Y_D	$[1, \infty)$	g (cfu kg)$^{-1}$	disaccharide yield
Y_G	$[1, \infty)$	g (cfu kg)$^{-1}$	glucose yield
γ_B	$\mathbb{R}_+$	h^{-1}	LAB decay
γ_X	$\mathbb{R}_+$	h^{-1}	yeast decay
K_D	$\mathbb{R}_+$	g kg^{-1}	Michaelis-Menten for disaccharide consumption
K_G	$\mathbb{R}_+$	g kg^{-1}	Michaelis-Menten for glucose consumption
p	$(0, 1)$	1	inoculation/carryover
w	$(0, 1 - p)$	1	hydration
h	$\{12, 24\}$	h	cycle length

Table 2.1: Symbols and acceptable values for each parameter in the system of equations.

[3]Though technically a dimensionless time variable, we only use the i variable to iterate through loops in the python program designed to solve this series of ODE's. For this reason, i is not considered "dimensionless" the same way we would consider the ratios p and w dimensionless.

Variable List			
Variable	Range	Units	Biological Significance
t	$\mathbb{R}_{\geq 0}$	h	time since last feeding
B	$\mathbb{R}_{\geq 0}$	cfu	LAB density
D	$\mathbb{R}_{\geq 0}$	g kg^{-1}	disaccharide concentration
G	$\mathbb{R}_{\geq 0}$	g kg^{-1}	glucose concentration
X	$\mathbb{R}_{\geq 0}$	cfu	yeast density
i	$\mathbb{N}$	n/a	number of feedings performed on starter

Table 2.2: Symbols and acceptable values for each variable in the system of equations.

2.2 Stability Analysis

One of the most useful tools used for understanding model behavior is stability analysis. By setting the first time derivatives for concentrations and densities to 0, we can understand whether and when steady states form in the system. We have already examined stability stemming from the first equation, finding that the only biologically significant steady state occurs when LAB dies out completely. Indeed, if we set the equation for yeast concentration, $\frac{dX}{dt} = 0$, we find the same case to be true; the only plausible steady state is the trivial one, where both yeast and glucose counts linger at 0. Repeating this process with the equation for glucose consumption and production, $\frac{dG}{dt} = 0$, we reach a similar result. We can safely conclude that the only stability is found in annihilation of resources.

2.3 Building a Computational Model

To properly simulate this model, we will need to create a program that solves the series of differential equations and difference equations. Python is well-suited for this task, boasting several useful free packages that can handle large data sets and compute numerical integration. In this program, we use the `numpy`, `scipy`, and `matplotlib` packages for handling vectors, integrating differential equations, and graphing solutions respectively.

Once all differential equations have been coded in, we define and initialize some parameter values needed to create solutions across multiple back-slops (i.e., number of cycles, lengths of

cycles, and hydration and inoculation ratios). Next we loop through the number of days we would like to model, using empirical amounts of flour carbohydrate levels and LAB/yeast concentrations to initialize the model on the first day. Every consecutive day's "initial conditions" are scaled down counts of the previous day's final conditions in accordance with the iterative portions of the model, with an added "ambient condition" proportional to the concentrations of nutrients and organisms supplied by raw flour. After as many loops through the model as the number of cycles we want to model, we can graph the resulting curves against time to examine stability and predict growth with different parameters.

Table 2.3: Initial values and parameters used to generate Figures 2.2 and 2.3

Table 2.4: Initial values

Variable	Value
B	100
D	1.544
G	0.114
X	1

Table 2.5: Parameters

Parameter	Value
α_B	0.1
α_X	1
Y_D	1
Y_G	1
γ_B	0.01
γ_X	0.01
K_D	1
K_G	1
p	0.3
w	0.3
h	12

The initial conditions for nutrient concentrations in g/kg come from studies on sugars in wheat flours published in the journal *Food Chemistry* [15], [16], but the initial conditions for microorganism density were slightly more difficult to find. Generally speaking, the ratio of LAB to yeast in a sourdough starter is 100:1 with yeast numbering about 1.0×10^6 cfu [1], so we can initialize our model with `B0 = 100` and `X0 = 1`. From there we are able to adjust other parameters within the model to either fit data or create interesting scenarios.

Here, we have executed the outlined code and generated a graph showing the populations of LAB (green) and yeast (red) as well as the concentrations of disaccharides (blue) and glucose (black). Notice the similarities between the curves for disaccharides and glucose in almost every iteration of this graph and the curves for maltose and glucose in Fig. 1.4. As LAB feed on the maltose-rich flour, they produce more glucose and cause a brief hump in the concentration of that nutrient.

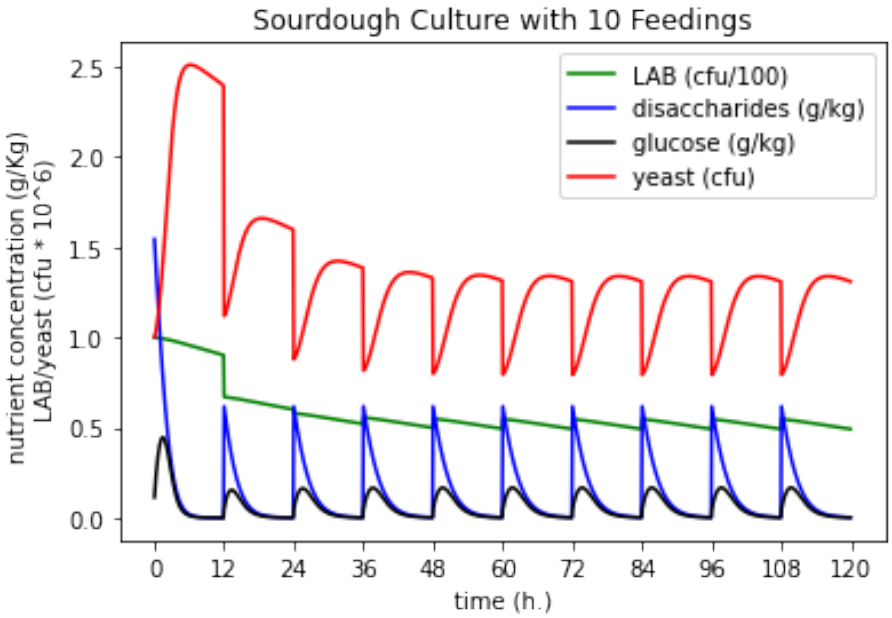

Figure 2.2: Over the course of ten days, the concentrations of LAB and yeast level out to predictable levels. Note that the concentrations of D and G bear a striking similarity to exponential decay of maltose and gradual rise of glucose in the presence of LAB from Figure 1.4.

One interesting and unintentionally appropriate feature of this system is a tendency towards stability over several generations. Many bakers who begin a sourdough culture will describe several initial cycles of unusually abundant growth, followed by a dampening period that leads into constant and predictable growth. That phenomenon is reflected very well here, with what appears to be a total of four feedings needed to bring the starter to a sort of equilibrium.

It is worth noting that the term "equilibrium" as used to describe the system over the course of several cycles is not at all related to the idea of equilibrium in a steady state for a continuous model. In fact, the continuous model is still unstable at each cycle, with the only force preventing collapse being the repetitive back-slopping process. This equilibrium stems instead from the idea that, over time, the final concentrations of all nutrients and organisms is consistent from cycle to cycle.

Also worth examining is the relationship between the cycle length h and the average concentrations of LAB and yeast after sufficiently many cycles, say 100 back-slops. By looking at how mean organism concentration changes with respect to the amount of time between feedings, we can try and find what feeding schedule optimizes sourdough growth and ensures a baker always has access to a viable starter.

To solve this problem, we simulate many generations of feeding until a quasi-steady state is reached. We then integrate our LAB and yeast functions over the course of the last computed cycle and divide by the cycle length, imagining the resulting value as a function of that cycle length. For our LAB optimization, we want to create a program that computes and graphs

$$f(h) = \lim_{n\to\infty} \frac{\int_0^h B_n(t)\mathrm{d}t}{h}.$$

Doing so yields the following graph, showing that with the set of parameters and initial values that created Figure 2.2, a cycle length of somewhere between ten and eleven hours will yield the maximum average yeast population, while the LAB population monotonically decreases as cycle length increases.

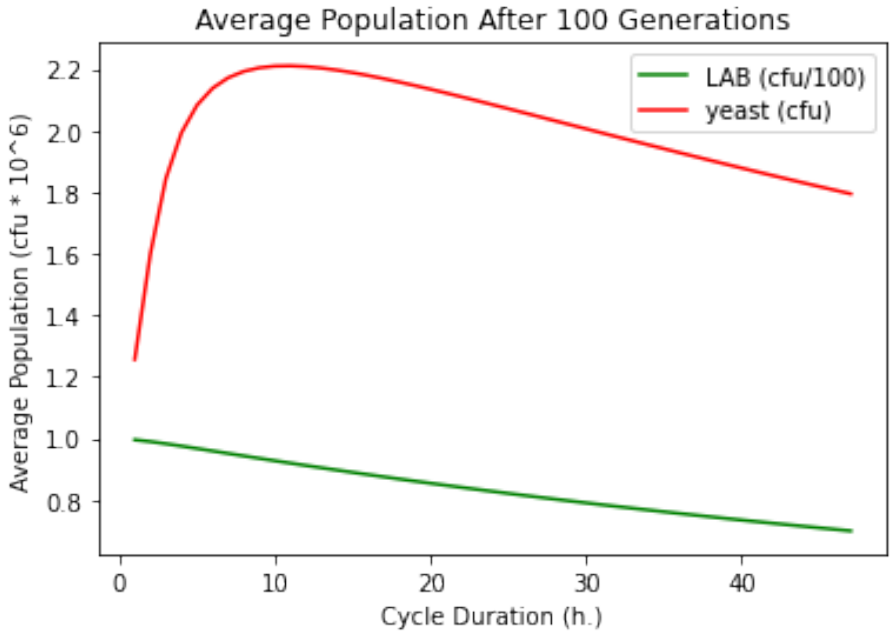

Figure 2.3: With the set of parameters used to generate Figure 2.2, we can find the cycle length which maximizes yeast and LAB concentration. In this particular case, average yeast concentration is maximized when a cycle lasts between ten and eleven hours, while average LAB concentration always decreases as cycle length increases.

For a full code listing, consult Appendix A - Source Code.

CHAPTER 3

DISCUSSION AND CONCLUSIONS

The model developed in this project is unique among most current models in that it is contextualized from the point of view of a baker who feeds a sourdough from generation to generation. This choice of approach was heavily influenced by the author's personal experiences with sourdough cultures and bread baking, and presents a new tool for professionals and amateurs alike to learn more about their own starters. Creating a model from this point of view has led to a prioritization of broader aspects of fermentation, like nutrient concentrations and metabolism pathways, over finer intricacies of the underlying biology.

3.1 Limits of the Model

By not focusing as much on the more intricate details of lactic acid fermentation, the model inevitably omits some key biological details that other models do not. In particular, we do not examine effects of temperature or pH fluctuations on population counts. By declining to include these variables, our model lacks some utility and could fail to more accurately predict sourdough growth in different environments with varying ambient temperature and water quality. Though pH variability has some interesting properties, growth rate as a function of temperature can largely be ignored in this specific context if we consider the fact that many sourdoughs ferment at room temperature, left in jars on counters and work benches in bakeries and kitchens. Our model would fail to accurately predict the behavior of a sourdough starter that moves between a refrigerator and a room-temperature area.

It is good to argue that a model born of a coarser-grained approach has some benefits in applicability and portability. Since the only variables needed to fit this model (carbohydrate make-up of flour and concentrations of LAB and yeast) are mostly easy to estimate or measure in a laboratory setting, this model can be adapted and results replicated for any number of starters in any number of environments.

Much research has already been done on the effects of temperature on bacterial and fungal growth. One step we can take to optimize this model in the future is the incorporation of temperature controls in the growth rates α_B and α_X. In fact, assuming a constant temperature over the

entire growth period would most likely yield a constant or barely significant α for each biological term in the series of equations [18]. It is most probable that incorporating such temperature terms would only be statistically significant for scenarios where temperature changes drastically over the course of the fermentation period.

3.2 Uncertainty Quantification (UQ)

The lack of in vitro and in vivo experimentation in this work makes it difficult to ascertain contextually appropriate values for parameters and initial values. We were able to find from previous research alone that the ratio of LAB to yeast is generally 100:1, and that the initial values of maltose and glucose in a kilogram of well-mixed wheat flours is typically 1.544 g and 0.114 g, respectively. Despite not knowing the "true" values of parameters, we could find smaller ranges through uncertainty quantification and sensitivity analysis. These newer processes can narrow the scope of uncertainty in model evaluation through Monte Carlo simulations.

There exist several packages in Python that aide mathematicians and modelers in performing this genre of analysis on their own creations. One fairly notable one is the `uncertainpy` package[19], a toolbox created for neuroscientists to evaluate model strength for systems with far more parameters than exist in this paper. An exciting next step in the course of this investigation would be to implement this package's tools with the model we have created to determine acceptable ranges for the parameters according to principles of UQ.

3.3 Optimization

As useful as UQ is in determining "goodness of fit" for parameters, there is no more economically effective tool at the mathematician's disposal than optimization. A goal of this thesis that has not yet been brought to fruition is the implementation of an optimization algorithm to iterate over flour types and hydration/inoculation ratios to minimize the number of days needed to reach a consistent sourdough starter. Bakers are very adept at changing these three variables, and it would be fascinating to see what effects these changes could have on resource management; water is inexpensive compared to flour, and if higher hydration were to yield similar results, businesses and bakers would be able to cut marginal costs while still delivering excellent bread.

3.4 Closing Thoughts

This model is a fascinating tool to use for modeling sourdough growth, particularly for its ability to iterate over several generations and show the slow process of approaching consistency in fermentation behavior. Despite some clear disadvantages to other contemporary models, the focus on metabolism and food chains has led to a whole world of opportunities for future research and collaboration with biologists and food scientists.

There are other organisms and groups that exhibit similar mutualistic behaviors to whom this model could just as easily apply in its current state, and there are other ways not explored in this model that make the romance between LAB and yeasts a symbiotic one. By no means is this model a fully comprehensive approach to lactic acid fermentation; rather it is a foundation for future examinations of the interactions which take place in every mouthful of bread.

It is crucial in future research to explore these ideas of human interaction with food fully. Food safety is a critical part of ecology which impacts the entire world, and by understanding the intricacies of humanity's storied relationship to grains and fermentation we can be better equipped to more compassionately interact with the food that sustains us.

APPENDIX A

SOURCE CODE

A.1 Back-slopping Program

```
#import necessary python packages

import numpy as np
from scipy.integrate import solve_ivp
from matplotlib import pyplot

pyplot.style.use('tableau-colorblind10')

#define ode system

def f1(t, Y, params): #dB/dt | bacteria growth
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return (alpha_b * B * (D / (k_d + D)) - (gamma_b * B))

def f2(t, Y, params): #dD/dt| disaccharide consumption
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return (-alpha_b / yield_d) * B * (D / (k_d + D))

def f3(t, Y, params): #dG/dt | glucose production
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return (alpha_b / yield_d) * B * (D / (k_d + D)) - (alpha_x / yield_g) * X * (G / (k_g + G)

def f4(t, Y, params): #dX/dt | yeast production
```

```
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return alpha_x * X * (G / (k_g + G)) - (gamma_x * X)

def rhs(t, Y, params):
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return f1(t, Y, params), f2(t, Y, params), f3(t, Y, params), f4(t, Y, params)

#ODE solver

#<cycle_length> = time between feedings (in hours)
#<subdivisions> = size of partition for integration
#<hydration> = hydration ratio; refers to "w" in text
#<carryover> = inoculation/carryover ratio; refers to "p" in text
#<cycles> = number of cycles iterated through, with feedings between them

cycle_length = 12
subdivisions = 120
hydration = 0.5
carryover = 0.3
cycles = 3

for day in range(cycles):
    tspan = np.linspace(cycle_length * day,
                        (cycle_length * day) + cycle_length - (cycle_length/subdivisions),
                        subdivisions)

    params = np.array([.01, #alpha_b
                       1, #alpha_x
                       1, #k_d
                       1, #k_g
```

```
                    .01, #gamma_b
                    .01, #gamma_x
                    1, #yield_d
                    1  #yield_g
                   ])
if day == 0:
    #feeding the model initial conditions found from literature
    B0 = 100
    D0 = 1.544
    G0 = 0.114
    X0 = 1
    yp = solve_ivp(lambda t, Y: rhs(t, Y, params),
           [tspan[0], tspan[-1]],
           [B0, D0, G0, X0],
           method = 'RK45',
           t_eval = tspan)
    B_curve = yp.y[0]
    D_curve = yp.y[1]
    G_curve = yp.y[2]
    X_curve = yp.y[3]
    timespan = yp.t
else:
    Bn = carryover * yp.y[0][-1] + (1 - carryover) * B0
    Dn = carryover * yp.y[1][-1] + (1 - carryover) * D0
    Gn = carryover * yp.y[2][-1] + (1 - carryover) * G0
    Xn = carryover * yp.y[3][-1] + (1 - carryover) * X0
    yp = solve_ivp(lambda t, Y: rhs(t, Y, params),
           [tspan[0], tspan[-1]],
           [Bn, Dn, Gn, Xn],
           method = 'RK45',
           t_eval = tspan)
    B_curve = np.concatenate([B_curve, yp.y[0]])
```

```
        D_curve = np.concatenate([D_curve, yp.y[1]])
        G_curve = np.concatenate([G_curve, yp.y[2]])
        X_curve = np.concatenate([X_curve, yp.y[3]])

        timespan = np.concatenate([timespan, yp.t])

#graphing solutions

plt.figure(1)

plt.title("Sourdough Culture with " + str(cycles) + " Feedings")

plt.plot(timespan, B_curve/100, 'green')
plt.plot(timespan, D_curve, 'blue')
plt.plot(timespan, G_curve, 'black')
plt.plot(timespan, X_curve, 'red')
plt.legend(['LAB (cfu/100)',
               'disaccharides (g/kg)',
               'glucose (g/kg)',
               'yeast (cfu)'])
plt.xlabel('time (h.)')
plt.xticks(np.arange(0, (cycle_length * cycles) + 1, cycle_length))
plt.ylabel('Amounts')
plt.ylabel('nutrient concentration (g/Kg) \n LAB/yeast (cfu * 10^6)')
```

A.2 Cycle Length Optimization Program

```
#import necessary python packages

import numpy as np
from scipy.integrate import solve_ivp
import matplotlib.pyplot as plt
plt.style.use('tableau-colorblind10')
```

```
#define ode system

def f1(t, Y, params): #dB/dt | bacteria growth
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return (alpha_b * B * (D / (k_d + D)) - (gamma_b * B))

def f2(t, Y, params): #dD/dt| disaccharide consumption
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return (-alpha_b / yield_d) * B * (D / (k_d + D))

def f3(t, Y, params): #dG/dt | glucose production
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return (alpha_b / yield_d) * B * (D / (k_d + D)) - (alpha_x / yield_g) * X * (G / (k_g + G)

def f4(t, Y, params): #dX/dt | yeast production
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return alpha_x * X * (G / (k_g + G)) - (gamma_x * X)

def rhs(t, Y, params):
    B, D, G, X = Y
    alpha_b, alpha_x, k_d, k_g, gamma_b, gamma_x, yield_d, yield_g = params
    return f1(t, Y, params), f2(t, Y, params), f3(t, Y, params), f4(t, Y, params)

#define backslop function:
#take imputs of cycle_length and variable type
#"b" for LAB, "x" for yeast
```

```
def backslop(cycle_length, variable):
    subdivisions = 120 # dt
    cycles = 100 # number of feedings
    hydration = 0.3
    carryover = 0.3
    params = np.array([0.01, #alpha_b
                       1, #alpha_x
                       1, #k_d
                       1, #k_g
                       .01, #gamma_b
                       .01, #gamma_x
                       1, #yield_d
                       1  #yield_g
                      ])

    for day in range(cycles):
        tspan = np.linspace(cycle_length * day,
                            (cycle_length * day) + cycle_length - (cycle_length/subdivisions),
                            subdivisions)
        if day == 0:
            B0 = 100
            D0 = 1.544
            G0 = 0.114
            X0 = 1
            yp = solve_ivp(lambda t, Y: rhs(t, Y, params),
                   [tspan[0], tspan[-1]],
                   [B0, D0, G0, X0],
                   method = 'RK45',
                   t_eval = tspan)
            B_curve = yp.y[0]
            D_curve = yp.y[1]
            G_curve = yp.y[2]
```

```
            X_curve = yp.y[3]
            timespan = yp.t
        else:
            Bn = carryover * yp.y[0][-1] + (1 - carryover) * B0
            Dn = carryover * yp.y[1][-1] + (1 - carryover) * D0
            Gn = carryover * yp.y[2][-1] + (1 - carryover) * G0
            Xn = carryover * yp.y[3][-1] + (1 - carryover) * X0
            yp = solve_ivp(lambda t, Y: rhs(t, Y, params),
                   [tspan[0], tspan[-1]],
                   [Bn, Dn, Gn, Xn],
                   method = 'RK45',
                   t_eval = tspan)
            B_curve = np.concatenate([B_curve, yp.y[0]])
            D_curve = np.concatenate([D_curve, yp.y[1]])
            G_curve = np.concatenate([G_curve, yp.y[2]])
            X_curve = np.concatenate([X_curve, yp.y[3]])

            timespan = np.concatenate([timespan, yp.t])

    if variable == "b":
        return np.average(B_curve[-subdivisions:])
    elif variable == "x":
        return np.average(X_curve[-subdivisions:])

def b_avg(h):
    return backslop(h, "b")

def x_avg(h):
    return backslop(h, "x")

#create arrays of function values to be graphed
```

```
h_points = []
b_points = []
x_points = []

for h in range(1, 48):
    h_points.append(h)
    b_points.append(b_avg(h)/100)
    x_points.append(x_avg(h))

#display graphs

plt.figure(1)

plt.title("Average Population After 100 Generations")
plt.plot(h_points, b_points, 'green')
    #showing scaled down LAB counts to enhance readability
plt.plot(h_points, x_points, 'red')

plt.legend(['LAB (cfu/100)',
               'yeast (cfu)'])
plt.xlabel('Cycle Duration (h.)')
plt.ylabel('Average Population (cfu * 10^6)')
```

REFERENCES

[1] G. Ventimiglia, A. Alfonzo, P. Galluzzo, *et al.*, "Codominance of Lactobacillus plantarum and obligate heterofermentative lactic acid bacteria during sourdough fermentation," *Food Microbiology*, vol. 51, pp. 57–68, 2015, ISSN: 0740-0020. DOI: 10.1016/j.fm.2015.04.011.

[2] P. Neysens and L. De Vuyst, "Kinetics and modelling of sourdough lactic acid bacteria," *Trends in Food Science & Technology*, vol. 16, no. 1, pp. 95–103, 2005, ISSN: 0924-2244. DOI: 10.1016/j.tifs.2004.02.016.

[3] A. Corsetti and L. Settanni, "Lactobacilli in sourdough fermentation," *Food Research International*, vol. 40, no. 5, pp. 539–558, 2007, ISSN: 09639969. DOI: 10.1016/j.foodres.2006.11.001.

[4] J. Hamelman, *Bread: A Baker's Book of Techniques and Recipes*, 2nd Edition. Wiley, 2012, ISBN: 1-118-13271-8.

[5] P. Hollywood, *How to Bake*. Bloomsbury Publishing Plc, 2012, ISBN: 9781408835562.

[6] N. Panikov, *Microbial Growth Kinetics*. London: Chapman & Hall, 1995, ISBN: 0-412-56630-3.

[7] B. Gompertz, "On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. F. R. S.," *Philosophical Transactions of the Royal Society*, vol. 115, pp. 513–583, 1825, ISSN: 2053-9223. DOI: 10.1098/rstl.1825.0026.

[8] J. Monod, "The Growth of Bacterial Cultures," *Annual Review of Microbiology*, vol. 3, no. 1, pp. 371–394, 1949. DOI: 10.1146/annurev.mi.03.100149.002103.

[9] T. Ross, "Belehradek-type models," *Journal of Industrial Microbiology*, vol. 12, pp. 180–189, 1993. DOI: https://doi.org/10.1007/BF01584188.

[10] T. J. Wijtzes, T. Rob, M. Zwietering, and P. McClure, "Modelling bacterial growth of Listeria monocytogenes as a function of water activity, pH and temperature," *International Journal of Food Microbiology*, vol. 18, no. 2, pp. 139–149, 1993.

[11] T. J. Wijtzes, J. de Wit, K. van 't Riet, J. H. J. Huis in 't Veld, and M. H. Zwietering, "Modelling Bacterial Growth of Lactobacillus curvatus as a Function of Acidity and Temperature," *Applied and Environmental Microbiology*, vol. 61, no. 7, pp. 2533–2539, 1995. DOI: 10.1128/aem.61.7.2533-2539.1995.

[12] A. Digaitiene, Å. S. Hansen, G. Juodeikiene, D. Eidukonyte, and J. Josephsen, "Lactic acid bacteria isolated from rye sourdoughs produce bacteriocin-like inhibitory substances active against bacillus subtilis and fungi," *Journal of Applied Microbiology*, vol. 112, no. 4, pp. 732–742, 2012. DOI: 10.1111/j.1365-2672.2012.05249.x.

[13] H. Neubauer, E. Glaasker, W. P. Hammes, B. Poolman, and W. N. Konings, "Mechanism of maltose uptake and glucose excretion in lactobacillus sanfrancisco," *Journal of Bacteriology*, vol. 176, 10 1994. DOI: 10.1128/jb.176.10.3007-3012.1994.

[14] M. G. Gänzle, "Enzymatic and bacterial conversions during sourdough fermentation," en, *Food Microbiology*, vol. 37, pp. 2–10, 2014, ISSN: 07400020. DOI: 10.1016/j.fm.2013.04.007.

[15] N. Struyf, J. Laurent, B. Lefevere, J. Verspreet, K. J. Verstrepen, and C. M. Courtin, "Establishing the relative importance of damaged starch and fructan as sources of fermentable sugars in wheat flour and whole meal bread dough fermentations," *Food Chemistry*, vol. 218, pp. 89–98, 2017, ISSN: 0308-8146. DOI: 10.1016/j.foodchem.2016.09.004.

[16] J. Pico, M. M. Martinez, M. T. Martin, and M. Gomez, "Quantification of sugars in wheat flours with an HPAEC-PAD method," *Food Chemistry*, vol. 173, pp. 674–681, 2015. DOI: 10.1016/j.foodchem.2014.10.103.

[17] L. Edelstein-Keshet, "Models for Molecular Events," in *Mathematical Models in Biology*, 1988, pp. 271 –310, ISBN: 978-0-89871-554-5.

[18] M. G. Gänzle, M. Ehmann, and W. P. Hammes, "Modeling of Growth of Lactobacillus sanfranciscensis and Candida milleri in Response to Process Parameters of Sourdough Fermentation," *Applied and Environmental Microbiology*, vol. 64, no. 7, pp. 2616–2623, 1998. DOI: 10.1128/AEM.64.7.2616-2623.1998.

[19] S. Tennøe, G. Halnes, and G. T. Einevoll, "Uncertainpy: A python toolbox for uncertainty quantification and sensitivity analysis in computational neuroscience," *Frontiers in Neuroinformatics*, vol. 12, 2018. DOI: 10.3389/fninf.2018.00049.