5-2013

Students Rise to the Challenge of Modeling Yeast Growth Despite Sour Hiccups from Imperfect Data

Alicia Caldwell
Utah State University

STUDENTS RISE TO THE CHALLENGE OF MODELING YEAST GROWTH DESPITE SOUR HICCUPS FROM IMPERFECT DATA

by

Alicia Caldwell

A report submitted in partial fulfillment of the requirements for the degree of

MASTER OF MATHEMATICS

in

Mathematics

Approved:

Dr. James A. Powell Major Professor	Dr. Brynja R. Kohler Committee Member
Dr. Luis Gordillo Committee Member	Dr. Nghiem Nguyen Committee Member

UTAH STATE UNIVERSITY
Logan, UT
2013

Students Rise to the Challenge of Modeling Yeast Growth Despite Sour Hiccups from Imperfect Data

Alicia Caldwell

ABSTRACT

Students rarely receive the opportunity to experience a learning activity involving mathematical modeling. This paper describes a lab in which students in an Applied Mathematics in Biology course observe the growth of S*accharomyces cerevisiae,* a yeast strain, in differing sugar concentrations for use in learning modeling. They parameterized the logistic equation and an alternate model, which they themselves constructed, based on the data they collected. I participated in this lab as a student in 2012 then observed and reviewed the work of other students involved. I found that students gained a deeper understanding of limiting factors and the role of parameter values in a model. Creative approaches were applied and problem solving skills refined as students exposed themselves to the modeling process. Student results and methodology are discussed.

1 INTRODUCTION

Explaining data through models and interpreting parameters are essential for any student of science. Mathematics students rarely receive the opportunity to be involved in collecting their own data when learning differential equations and models. Biology and ecology students learn data collection and statistical analysis to describe data but rarely gain an understanding of the tenor of a model and its associated parameters, much less how to determine them. Zbiek and Conner (2006) explain that mathematical modeling activities can help:

- prepare students to work professionally,
- motivate students to study mathematics through real-world applicability, and
- provide students with opportunities to integrate mathematics with other areas of curriculum.

A difficult barrier students must cross is learning that past solutions to problems are not always the finest solutions. Powell et al. (2012) assert that modeling allows students to construct and criticize models of their own and helps them "quickly over the notion that there is a single 'correct' model." Instructors must provide an atmosphere that promotes creativity with students so they feel comfortable learning outside traditional mathematical teaching methods i.e. learning ideas and concepts through lecture then completing homework problems that typically have one correct answer, to solidify these concepts. Students also tend to pattern the work of their instructor causing misgivings when confronted with a problem that did not come with a step-by-step handbook. With the proper instruction and supplemental labs for hands-on learning, students can grow in creativity.

Moshchkovich (2004) discusses how a person learning mathematics connects pieces of knowledge then adds new and corrects old pieces of knowledge. When confronted with a problem without an exact solution or specific method for reaching a conclusion, a student must make these

connections which force him or her, as the modeler, to play a leading role in the process. The assumptions pertaining to the data and the tools used for graphing and parametrization influence the modelers' learning. In addition, the modeler's awareness of these assumptions permit them to see mathematical structure and success with their work (Zbiek and Conner 2006). As the modeler becomes involved in the construction process, mathematical understanding is critical for the model to be successful. This process creates a situation better targeted for deeper understanding, not only about the situation being modeled, but about the actual application of mathematics taking place.

Students who are given the opportunity to formulate their own questions and design a model with the instructor's guidance become personally responsible for their learning. Biembengut and Hein (2010) state, "learning becomes richer, considering that the student does not just learn mathematics inserted in the context of another area of knowledge, but also has his critical and creative senses stirred." This learning method also provides students with a better understanding of mathematical concepts and training to read, interpret, formulate and solve specific situation problems (Biembengut and Hein 2010).

Placing responsibility of learning directly on the student gives them the opportunity to grow and strengthen their capabilities. Zbiek and Connor (2006) stated that "modeling work provides not only *motivation* to learn mathematics but also *opportunities* to learn mathematics." The goal of this particular approach is to provide the kind of environment where students are accountable for their own development. Learning not only the basics of a model, but how its creation came to be, allows students to gain skills in putting together their own models with the knowledge they already possess, coupled with what they are currently learning.

Despite all the research that modeling opportunities promote richer learning, students are not

exposed to experiences that give them the opportunities described above. Many students taking calculus, linear algebra, and differential equations courses are not pure mathematicians, but will apply mathematics in their field, such as engineering, physics and biology. These students are not provided experiences that give them an opportunity to apply mathematics in a context that relates to the real world. To address this lack of genuine modeling experiences, this paper will discuss a university lab to present such an experience where students concentrated on the modeling of yeast.

Yeast is a common experimental organism because it is unicellular and grows on simple media, giving investigators control over its environmental parameters. The first glimpse of the nature of yeast came from Van Leeuwenhoek in 1680. The role of yeast in alcoholic fermentation was first recognized in 1835 by Cagnaird-Latour (Horst Feldmann 2010); later, Louis Pasteur correlated fermentation with yeast metabolism in his work *Études sur la biére* in 1876 (Hornsey 2007). Most notably, Georgy Gause experimented with yeast and published *The Struggle for Existence* where he describes his findings, now called the competitive exclusion principle or Gause's Law. All of this research led to the 1930 recognition that yeast represents "an ideal system to investigate cell architecture and fundamental cellular mechanisms." Since then, yeast has been a common experimental choice since cellular functions are largely conserved from yeast to mammals (Feldmann 2010).

In this paper, we present a lab based on the population growth of yeast in which students collected their own data then constructed and parameterized models. From this exercise students gained a better understanding of parameters, population dynamics, limiting factors and capturing natural phenomena using differential equation models. The lab was launched with a brief discussion of yeast physiology and the logistic equation, which provided a springboard for students to develop their own models. To illustrate learning outcomes we will present two sets of data collected by students, one

with unusual results due to experimental error and the other with more common results. Three novel models engineered by students working with the imperfect data from 2012 will also be presented. Students successfully told a story about the yeast through their models and parameterized them in such a way that captured the data, illustrating how students gained a deeper understanding of modeling and parameterization when allowed to interact with real data. My experiences as a student participant of the lab and as an observer are also given and discussed.

2 LAUNCHING THE LAB

The experiences of this lab are those of the instructor and students in an Applied Mathematics in Biology (AMB) course designed specifically at Utah State University for students to learn the art of modeling. The data for this paper was collected in 2010 and 2012. Undergraduate and graduate students in mathematics and statistics, biology, ecology, and physics participated in the lab. Each participant was exposed to calculus and differential equations at some point in their education, although their familiarity with those subjects varied, each was able to understand enough to create models.

To introduce the lab, a brief explanation and background of the subject was provided by the instructor. First, an overview of yeast and sugar dynamics was given. Review of the logistic equation, how it is derived and why it is used for yeast, was discussed next. The instructor briefly touched on some techniques of differential equations such as separation of variables, derivation of conserved quantities, fixed points and one dimension stability analysis. Once the students possessed the necessary tools, the instructor assisted them in the lab, providing aid that encouraged them to do their own thinking.

2.1 Introductory Information

2.1.1 Lab Instructions and Expectations

The task at hand was for students to be a part of the experimental process by measuring sugar concentrations, counting yeast cells and collecting data. Once the students had collected the data their goal was to fit the logistic model (the null model) to the experimental data and find parameters for the model to test with the validation data. Students, in groups, then attempted to create at least one alternate model that improved correspondence with the validation data when compared to the logistic model. The challenge given was open-ended which allowed students to explore different paths and ideas for models. Each student turned in a lab write-up in traditional form (introduction, methods, results, and discussion and conclusion). Specific components were included in the report:

- Explanation of methods for data collection
- Definition of alternate model for growth as well as the null model
- Explanation of methods for parameter estimation
- Comparison of the models with the validation data
- Discussion of the results

Each student was asked to include in their lab write-up figures and tables that illustrate the results in the paper. Once the objectives were presented, the instructor allowed himself to be readily available for students to answer questions and help lead them in the right direction when at a crossroads.

2.1.2 Introduction of Yeast

Beer making can be traced back 5,000 years from documents and inscriptions found in Egyptian tombs. In ancient times, brewing beer was an art or craft having an absence of scientific know-how (Reed and Tilak 1991). Early brewers spent most of their trial-and-error struggles dealing with ideal temperatures

and yeast recycling (Unger 2004). It was not until later brewers discovered that good beer depends highly on the fermentation of sugar. More studies showed the ideal use of yeast as an experimental system and now various industries are based on the use of yeast (Reed and Tilak 1991). Yeast is often used for population growth experiments since it grows rapidly and resources necessary can be readily found.

Gause shares his experience with yeast and other organisms in *The Struggle For Existence.* He states that "for every species or race there is a maximal number of individuals which can never be surpassed". This maximal number is hard to determine and depends on environmental conditions. After running experiments with yeast Gause found, even though multiplication of organisms is potentially unlimited, limitations are introduced by external forces. Coming out of these experiences Gause discovered that "all conditions of cultivation ought to be so arranged that the growth depends distinctly on only one limiting factor" (Gause 1934).

The yeast species used in this class was S*accharomyces cerevisiae*, a single celled fungi, one of the strains on which Gause experimented. It was discovered in 1837 in malt in connection to beer making (Feldmann 2010). Like any other living organism, yeast needs energy to grow and multiply. Yeast obtain energy from carbohydrates, such as sugar, and in order for a yeast population to increase a sufficient supply of resource is necessary; these carbohydrates are a limiting factor. Yeast enzymes break down the carbohydrates under anaerobic or aerobic conditions. If oxygen is available the yeast will break down carbohydrates using aerobic respiration which captures energy in the form of Adenosine Triphosphate (ATP) at a rate of 36 ATP per sugar molecule (Voet and Voet 2004). This is the energy the yeast cells use to repair and reproduce. In the absence of oxygen yeast ferments. The fermentation process breaks down sugars to alcohol, carbon dioxide, and ATP as follows

$$\underbrace{C_6H_{12}O_6}_{glucose} \rightarrow \overbrace{2CH_3CH_2OH}^{alcohol} + \underbrace{2CO_2}_{carbon\ dioxide} + \overbrace{energy}^{ATP} .$$

However, fermentation is not as efficient as aerobic respiration, generating only 2 ATP per sugar molecule (Voet and Voet 2004). Gause (1934) also found that with the accumulation of alcohol in the medium, which corresponds to the amount of sugar consumed, growth of yeast ceases, but alcohol continues to accumulate. Once this happens, yeast cells continue to bud actively but daughter cells die upon separation from the mother cell.

2.1.3 Logistic Model

The logistic model captures yeast interaction with sugar and is the standard model for yeast growth. The curve was discovered by Verhulst (1838) on the idea that growth rate of a population is determined by the relation between potential rate of increase and environmental resistance (Gause 1934).

For the yeast lab the logistic model is used as a null model by which students can assess the performance of their own models. In the AMB class, following Verhulst and paving the way for more complex models from the students, the model was presented as two differential equations, one for the resource and another for population growth. These were written as a single, nonlinear equation after eliminating the resource variable. In doing this the students saw how the different variables affect one another, especially how the sugar, or the limiting factor, is incorporated in the model. This also provided them with a way to start developing their own models by thinking about each variable of interest solely. Lastly, it helped them gain an understanding of parameters and where they come from.

Let Y represent the concentration of yeast cells, measured in cells per liter $\left(\frac{C}{L}\right)$, in a

population and *S* the glucose in grams per liter $\left(\frac{g}{L}\right)$. Based on the assumption that the glucose is used for the yeast population to grow, let ξ be defined as the rate of growth per sugar concentration in liters per gram per time $\left(\frac{L}{g \cdot t}\right)$, and η the amount of sugar needed to produce new yeast in grams per cell $\left(\frac{g}{C}\right)$, then the logistic model in *t,* time, is

$$\frac{dY}{dt} = \xi S Y \quad , \tag{1}$$

$$\frac{dS}{dt} = -\eta \xi S Y \quad . \tag{2}$$

See Table 1 for parameter reference. The right hand side of these equations differ only by a constant, giving the following relationship between the derivatives,

$$\eta \frac{dY}{dt} = -\frac{dS}{dt} \quad , \tag{3}$$

which can be integrated.

$$\eta \int_0^t \dot{Y}\, dt = -\int_0^t \dot{S}\, dt \quad . \tag{4}$$

Using the fundamental theorem of calculus (4) becomes,

$$\eta (Y - Y_0) = S_0 - S \quad , \tag{5}$$

where Y_0 is the initial density of yeast and S_0 is the initial concentration of sugar. Solving for S gives

$$S = S_0 - \eta (Y - Y_0) \quad . \tag{6}$$

Using (6) and (1) it follows that

$$\frac{dY}{dt} = \xi\ Y(S_0 + \eta\, Y_0 - \eta\, Y) \quad , \tag{8}$$

and factoring gives

$$\frac{dY}{dt} = \xi\,\eta\left(\frac{1}{\eta}S_0 + Y_0\right) Y\left[1 - \frac{Y_0}{\left(\frac{1}{\eta}S_0 + Y_0\right)}\right] \quad . \tag{9}$$

Let $r = \xi(S_0 + \eta\, Y_0)$ and $K = \frac{1}{\eta}S_0 + Y_0$,

$$\frac{dY}{dt} = rY\left(1 - \frac{Y}{K}\right) \quad , \tag{10}$$

where *r* is the intrinsic growth rate and *K* is the carrying capacity, the equilibrium density of yeast that can be supported by the environment. This is the logistic equation for yeast growth, in which carrying capacity reflects the initial amount of sugar present.

Table 1: Variables and Parameters with descriptions for Logistic Model		
Variable	Description	Units
Y	Yeast concentration	$\frac{C}{L}$
S	Sugar concentration	$\frac{g}{L}$
Parameter	Description	Units
η	Amount of sugar needed to produce new yeast	$\frac{g}{C}$
ξ	Rate of population growth	$\frac{L}{g \cdot t}$
r	Intrinsic growth rate	$\frac{1}{t}$
K	Carrying capacity	$\frac{C}{L}$

2.1.4 Methods for Parametrization

AMB students were introduced to parameterization via minimization of sum squared error. Let *SSE* be the sum of the squared residuals,

$$SSE(data, \theta) = \sum_{i,j} (y_i(t_j) - y_{pred}(t_j, \theta))^2 \quad , \tag{12}$$

where y_i represents the observed data values for replicate i at time t_j and y_{pred} are the values predicted by the model at time t_j with parameter values given in vector $\boldsymbol{\theta}$. The least squares method finds the set of parameters that minimizes the sum of the squared errors. This can be done using programs like Matlab or Maple. Example code in Matlab will be given in Appendix C for reference.

For some of the more involved models that students cannot find an explicit solution, some kind of program, such as Matlab, was needed to numerically solve and fit these models. Appendix B discusses parameterization of the logistic model using linear regression.

2.2 Running the Lab

2.2.1 Time Frame

The AMB classes were held twice a week for two hours. The introduction to yeast and the logistic equation, as well as discussion on how to execute the lab, was done in one class period. The experimental set up was done in one class period as well and the students took turns coming in every two hours for the next 50 hours following the set up to collect the data. The following class period after data collection, the students discussed the lab results and began model construction. Three hours was also given to the students to work in the computer lab on fitting the models to the allowed data. The lab reports were expected to be written and turned in a week after time given for model construction.

2.2.2 Lab Execution

Exact instructions to lab set up and the experimental process can be found in Appendix A. There were three experimental groups and a control group, each having three replicates. The experimental groups included a group with a high amount of sugar added to the medium, another with a low amount added, and the final was somewhere in-between. The final experimental group was used as the validation data for the student models. This means that data for this group was collected with all the others but was not used to fit the models. Students participated in the experimental process by measuring out initial weights of sugar, counting yeast at different dilutions, preparing each replicate and collecting and recording data throughout the defined time.

2.2.3 Data Description

Data comes from 2010 and 2012 classes in the AMB course at Utah State University. In 2010, the three treatments groups had 50 g/L, 20 g/L and 2 g/L of added glucose. There was also a control group and the validation data had 10 g/L of added glucose. The 2012 data had two treatment groups, 20 g/L and 5 g/L of added glucose, a control and the validation data had 10 g/L of added glucose. At least, that was the plan. One student made a small mathematical error that resulted in the high treatment group receiving 0.2 g/L of added glucose, 0.05 g/L for the low, and 0.1 g/L for the validation.

The 2010 experiment produced more typical data as the yeast with more available sugar grew to higher concentrations than those with less and the control group gained fewer than 250 cells in 150 ml of medium, which was the volume of each flask, over the 35 hours. The yeast having 50 g/L added grew to populations concentrations just below 10,000 cells with an approximate starting value of 400 cells. The yeast with 20 g/L of added glucose grew to population concentrations between 4,000 and 5,000 cells. This can be seen in Figure 1.

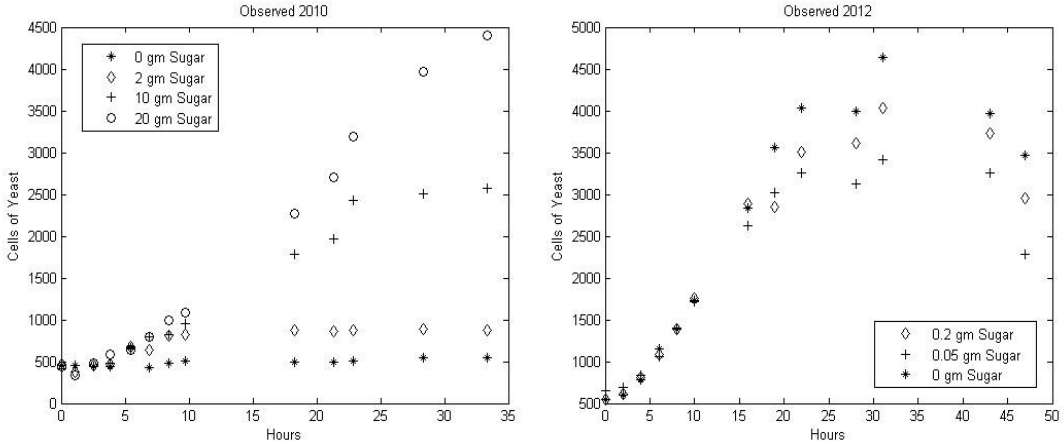


Figure 1: Left) Observed data 2010 with four experimental groups, Control, 2 gm of sugar added, 10 gm of sugar added, 20 gm sugar added. Right) Observed data 2012 with three experimental groups, Control, 0.05 gm sugar added (low), 0.2 gm sugar added (high)

The 2012 data produced surprising results, a hiccup, as the control group grew to higher populations than those of the treatment groups. Ironically, the treatment groups were consistent within themselves, meaning the treatment group having 0.2 gm/L added grew to higher yeast concentrations than that of the validation group having 0.1 gm/L of added glucose, which were higher than those of the treatment group with 0.05 gm/L of added glucose. Yeast concentrations hit a maximum then begin to decrease between 30 and 35 hours (Figure 1). After experimentation in 2012, the instructor informed the students of the small amount of dextrose present in the growth medium. This happened every year the experiment was run but did not affect the results as it was a very small amount. However, with the experimental error in 2012, the amount no longer seemed that small compared to what was added.

3 STUDENT OUTCOMES

The 2012 class was split into two groups, each group coming up with alternate models and parametrization methods. Responding to the challenge of the 2012 data, students developed creative approaches, illustrating that problematic experiments provide mathematical opportunities. Three

models created by the students are shared and explained in this section as well as how these models compare with the null model and the validation data. Methods utilized by the students to parameterize the null model and their novel models are given, illustrating students' ability to construct a model to represent data and how that process helped students gain a deeper understanding of parameter values and limiting factors.

3.1 Student Parametrization of Logistic Model

There are two parameters in the logistic model: the growth rate r and the carrying capacity K. Both groups applied the least squares method to find these parameters. Recall each treatment group had three replicates. Group I used this method to find parameters for each replicate then averaged those parameters to work for that specific treatment group. Group II created a function in Matlab that would fit all the replicates together to give parameter values for each treatment group. This gave each group a growth rate and carrying capacity for the control, high, and low treatment groups.

Each treatment group all yielded roughly the same r, $r = 0.2$, in the process explained above so both groups used this r value for the validation data having 0.1 g/L amount of sugar added. Trying the different r values around 0.2, the students decided they didn't make much difference in how the model looked with the data so they used $r = 0.2$. To find parameter K for the validation data, Group I took the difference between K for the high amount of sugar added ($K = 4062.5$) from the low amount of sugar added ($K = 3588.5$). They multiplied this difference by one third as the initial sugar concentration for the validation data is one third of the way between the low and high initial concentrations. They added this result to the low K value which gave a carrying capacity for the validation data of $K = 3275.9$.

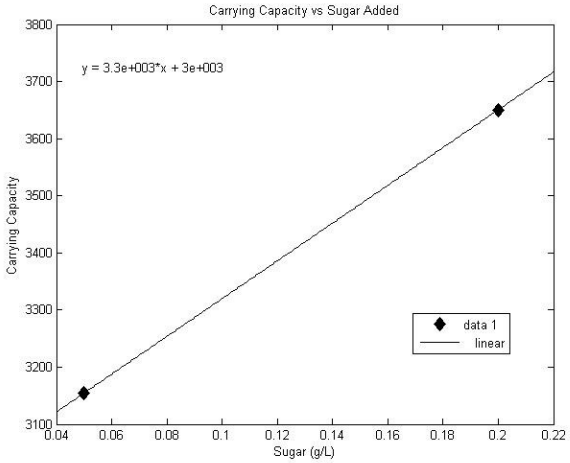


Figure 2: Graph used to find carrying capacity by Group II in parameter estimation.

The students in Group II plotted initial sugar added against the carrying capacity using the parameters they found for high, low and control groups, as seen in Figure 2. The fit a line to the data and found their *K* value to be *K = 3330*, not too far off from Group I. Ironically, the students had been shown that the carrying capacity is a linear function of initial sugar, $K = \frac{1}{\eta} S_0 + Y_0$, when the logistic equation was presented in class, they just did not make the connection. They later discovered this upon talking to the instructor who was happy to see students discovering these ideas for themselves. This equation could have been used to help find an *r* value, as $r = \xi\,(S_0 + \eta\, Y_0)$, rather than settling on something that seemed to look good.

3.2 Student Models and Results

3.2.1 Model Creation

Once the data was collected the instructor set aside a class period to brainstorm models. Printouts of the plotted data were available to analyze and get creative juices flowing. Students worked with the

same group as with the logistic model for their authentic model. As each group contained students with different educational backgrounds, different ideas were presented as to why the control group grew to higher concentrations than the experimental groups having the extra added sugar. Also, there was a question as to why the yeast started to "die off" in the last couple hours. Once students finalized their ideas they created models by figuring out what affects the change in yeast density over time, sugar density over time, and any other important factors introduced in their model.

One idea to capture the death of the yeast population in later hours was to incorporate into the logistic equation the effect of alcohol on the growth rate. Two ideas were presented to explain the high growth of the control data. One was that another strain of yeast was introduced into the system with the sugar causing competition in the experimental groups, and two, another type of sugar was introduced into the medium that would cause slower growth rates for the yeast due to competition. Elaboration of these ideas are continued in their respective model sections below.

3.2.2 Alcohol Death Model

Possessing a solid foundation in biology, the students in Group I created the Alcohol Death Model. These students designed this model to take into account death of yeast due to alcohol produced during fermentation as well as the decrease in growth as a result of the decline in sugar concentrations. Using the knowledge from the logistic equation provided in class, the students added another equation to the system targeting alcohol. The participants decided that the alcohol will grow jointly with the yeast and sugar concentrations and also hinder the growth of yeast upon reaching high concentrations. The alcohol death model was comprised of

$$\frac{dY}{dt} = \gamma\, SY - \beta\, AY \quad , \tag{13}$$

$$\frac{dS}{dt} = -\kappa\, \gamma\, SY \quad , \tag{14}$$

$$\frac{dA}{dt} = \alpha\, SY \quad , \tag{15}$$

where A represents the alcohol concentration, γ represents the rate of growth per sugar concentration, β represents the rate of yeast death due to alcohol, κ represents the amount of glucose needed to produce new yeast, and α the growth rate of alcohol, see Table 2. The sugar contributes to the growth of yeast with the same relationship as explained in the logistic model, but now the death of yeast due to alcohol is taken into account. As seen is Equation 15 the alcohol concentration will grow as the yeast population grows, but Equation 13 shows that as the alcohol concentration grows it slows the growth rate of yeast. This rate can become negative if there is no longer sugar available; this gives the potential for the yeast concentrations to drop as time moves on.

Table 2: Variables and Parameter values with descriptions for Alcohol Death Model			
Variable	Description	Units	
Y	Yeast concentration	$\frac{C}{L}$	
S	Glucose concentration	$\frac{g}{L}$	
A	Alcohol concentration	$\frac{L}{L}$	
Parameters	Description	Units	Parameter Values
α	Increase in alcohol	$\frac{L^2}{g \cdot C \cdot t}$	5.051×10^{-2}
β	Rate of yeast death due to alcohol	$\frac{1}{t}$	1.708×10^{-5}
γ	Uptake of glucose by yeast	$\frac{L}{g \cdot t}$	1.139×10^{-1}
κ	Amount of glucose needed to produce new yeast	$\frac{g}{C}$	1.976×10^{-4}

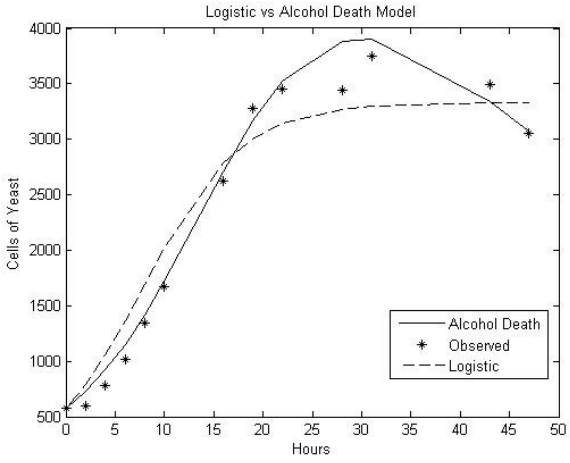


Figure 3: Predicted Logistic and Alcohol Death Model with validation data from 2012. Logistic R^2 value is .949 and Alcohol Death R^2 value is 0.957.

The students chose to reduce the model to two differential equations. They started by multiplying Equation 15 by $\gamma\kappa$ and Equation 14 by α giving

$$\alpha \frac{dS}{dt} = -\alpha \kappa \gamma SY \quad , \tag{16}$$

$$\kappa \gamma \frac{dA}{dt} = \alpha \kappa \gamma SY \quad . \tag{17}$$

Adding (16) and (17) together results in

$$\alpha \frac{dS}{dt} + \kappa \gamma \frac{dA}{dt} = 0 \quad . \tag{18}$$

Integrating (18)

$$\alpha S + \kappa \gamma A = \alpha S_0 + \kappa \gamma A_0 \quad , \tag{19}$$

where $A_0 = 0$ and B is a constant picked up from integration.

Solving (19) for A gives

$$A = \frac{\alpha (S_0 - S)}{\kappa \gamma} \quad . \tag{20}$$

Now substitute (20) into (13) and the system becomes

$$\frac{dY}{dt} = \gamma SY - \alpha^2 \beta Y \left(\frac{S_0 - S}{\kappa \gamma} \right) \quad , \tag{21}$$

$$\frac{dS}{dt} = -\kappa \gamma SY \quad . \tag{22}$$

Rather than settling at a carrying capacity, the yeast will reach a higher population as long as sugar is available. Increasing alcohol and depletion of glucose creates negative growth conditions.

Parameter estimates for this model came from minimizing the sum of the error squared as stated above. To find parameters for the Alcohol Death Model, Group I decided to fit all the treatment groups

separately, giving different parameter values for each treatment group. To utilize them with the validation data, Group I took an average of the different values from each treatment group. The average initial sugar concentration between the high and low was 0.125 g/L which is slightly higher than the validation at 0.1 g/L but close enough that the students decided averaging high and low parameter values should work for the model.

This model gave a different shape, see Figure 3. The model takes into account the death in yeast due to high alcohol concentrations. Looking at Figure 3 the first thought may be that the data looks to follow a logistic pattern, but a closer look will reveal that the yeast concentration begin to decrease in the last hours in some of the data sets. The Alcohol Death Model captures the decrease in population for the higher time records due to increase in alcohol concentration. This group was able to qualitatively capture that phenomenon, although R^2 only improved from 0.949 to 0.950.

3.2.3 Two Sugars Model

The control population received no added sugar; therefore, a low carrying capacity is expected in relation to the populations with added sugar. One student model stemmed from fact that the control group reached higher populations than that of the other treatment groups and taking into account that there was consistent growth within the treatment groups. Having learned that the growth medium contained dextrose after the experimental process, Group II entertained the idea that having dextrose and glucose present (despite the fact that they are essentially the same) slowed the process of sugar breakdown, delaying the cultivation of new yeast.

While creating the model they hypothesized that without the added resource, the system should act logistic, and used that model as a template. They also decided that higher concentrations of glucose should slow the growth of yeast due to the alternate resource (dextrose). This rate increases as the

available amount of glucose depletes. To achieve this, they added a diminishing reaction term, $\left(\frac{aD}{b+G}\right)$, to the change in yeast concentration and another equation for the second resource.

Let Y represents the yeast, D represent an alternate resource, and G the added glucose, then the two sugars model can be written as as follows:

$$\frac{dY}{dt}=\left(\frac{aD}{b+G}\right)Y+zGY \quad , \tag{23}$$

$$\frac{dD}{dt}=-\left(\frac{caD}{b+G}\right)Y \quad , \tag{24}$$

$$\frac{dG}{dt}=-qzGY \quad , \tag{25}$$

where a is the rate of growth per alternate resource concentration, b is the half saturation term, c is the amount of alternate resource needed to produce new yeast, q is the amount of glucose needed to produce new yeast, and z is the rate of growth per glucose concentration, refer to Table 3. If the glucose is not added, $G=0,$ then the yeast behaves like the logistic model, feeding only on the alternate resource, as seen from Equation 23. However, analyzing Equation 24, more added glucose, G, slows the growth of yeast from dextrose.

Unlike Group I, Group II fit all of the data together giving parameter values that work for each treatment group; building a function in Matlab that requires the initial resources and yeast concentrations allowed them to accomplish this. They thought this would provide more accurate parameter values than fitting individually and averaging (See coding example in Appendix C). In order to use the least squares method for parameterization initial guesses are needed. To find reasonable guesses it helps to note that z and q to act much like η and ξ in the logistic equation.

The written reports from Group II discussed the results of their work with the Two Sugars Model in predicting the growth of the validation yeast. The Two Sugars Model does a decent job predicting yeast growth producing a slightly better R^2 value than the logistic (Figure 4). However, the shape of the models resemble each other. One student in Group I expressed his thoughts on why this might be, saying "there is not difference between the growth of *S. cervisiae* on dextrose versus glucose." The amount of glucose needed to produce new yeast is about $3\times10^{-4},$ which is the same value for found for the logistic equation. So, in this model the yeast are growing at the same rate from the glucose as in the logistic equation. The dextrose is depleting much faster than the glucose causing that source to be used up quickly leaving the yeast only the glucose for energy. Since it grows in this model from the glucose much like that of the logistic, the end result looks logistic. Another student thought the model as a small refinement to the logistic.

Table 3: Variable and Parameter values with descriptions for the Two Sugars Model			
Variable	Description	Units	
Y	Yeast concentration	$\frac{C}{L}$	
G	Glucose concentration	$\frac{g}{L}$	
D	Alternate resource	$\frac{g}{L}$	
Parameter	Description	Units	Parameter Value
a	Rate of growth per alternate resource concentration	$\frac{L}{g \cdot t}$	2.174×10^{4}
b	Half saturation rate	$\frac{g}{L}$	9.403×10^{2}
c	Amount of alternate resource needed to produce new yeast	$\frac{g}{C}$	5.091×10^{-3}
q	Amount of glucose needed to produce new yeast	$\frac{g}{C}$	3.308×10^{-4}
z	Rate of growth per glucose concentration	$\frac{L}{g \cdot t}$	1.483×10^{-1}
D_0	Initial amount of Dextrose	$\frac{C}{L}$	1.023

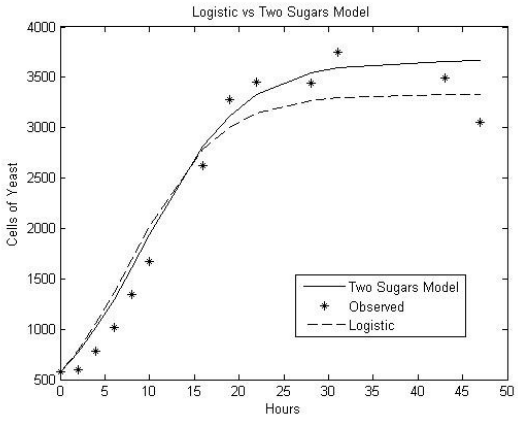


Figure 4: Predicted Logistic and Two Sugars Model with validation data from 2012. Logistic R^2 value is 0.949 and Two Sugars R^2 values is 0.950

3.2.4 Wild Yeast Model

Upon further inspection of the collected data, Group II, the group who came up with the Two Sugars Model, thought maybe another "wild" type of yeast could have been introduced with the sugar. Then two types of yeast were depleting the food source causing the yeast to not have the resources needed to grow at high rates. The control group, growing to higher yeast populations, did not receive the extra competition from the yeast introduced with the sugar allowing that population to grow. Their model needed two types of yeast using up the sugar and an equation for concentration growth for each type of yeast. To capture this thought the group came up with the following model:

$$\frac{dY_i}{dt} = \mu S Y_i \tag{26}$$

$$\frac{dY_w}{dt} = \nu S Y_w \tag{27}$$

$$\frac{dS}{dt} = -\left(\tau\mu S Y_i + \varphi\nu S Y_w\right) \quad . \tag{28}$$

In this model, Y_i represents the original yeast introduced by the experimenters, Y_w represents the yeast introduced with the sugar, S, and μ and ν represent the rate of growth per sugar concentration for the respective yeast, and lastly, τ and φ the amount of sugar needed to produce new respective yeast.

To find parameters for their model, this group used the same method of that with the two sugars model. Results from the Wild Yeast Model also produced decent fits to the data. Inspection of Figure 5 shows the model still follows a logistic pattern but predicts the data fairly well. Analyzing the growth of the two yeasts and depletion of sugar over time using the model and said parameters, one student realized the wild yeast does not grow to reach high concentrations. The student determined that

another yeast strain must not have been added in with the sugar as originally thought.

While finding parameter values, Group II found that some value outputs were not consistent. The inconsistencies were in the initial amount of wild yeast and ν, the growth rate per sugar concentration of that yeast. The Matlab function converged to two sets of parameter values, one with a large initial amount of sugar and and a smaller ν, and the other vice a versa. After discussing why this might be with their group members and the instructor they found that the two parameters always appear as a pair, see Equation 27. Upon learning this they non-dimensionlized the initial amount of wild yeast and let ν pick up the difference. So, using 1 as a starting population value the group was able to find a growth rate for wild yeast.

Again, the shape of the model looks logistic. The students in Group II realized that ν was really small and that the wild yeast did not seem to be growing much. To determine how much the wild yeast actually grew over the time frame, the students ran their model with their parameters and the correct initial information for the validation data and they found the wild yeast only grew to 3.92×10^{-10} percent of their starting population. They group members decided that due to the lack of growth of the wild yeast, there must not have been wild yeast introduced into the system.

Table 4: Variables and Parameter Values with descriptions for the Wild Sugar Model			
Variable	Description	Units	
Y_i	Yeast introduced by experimenters	$\frac{C}{L}$	
Y_w	Wild yeast, Introduced with the sugar	$\frac{C}{L}$	
S	Sugar concentration	$\frac{g}{L}$	
Parameter	Description	Units	Parameter Value
μ	Rate of population growth for Y_i	$\frac{L}{g \cdot t}$	1.486×10^{-1}
ν	Rate of population growth for Y_w	$\frac{L}{g \cdot t}$	3.161×10^{-9}
φ	Amount of sugar needed to produce more Y_w	$\frac{g}{C}$	5.961×10^{-5}
τ	Amount of sugar needed to produce more Y_i	$\frac{g}{C}$	3.310×10^{-4}

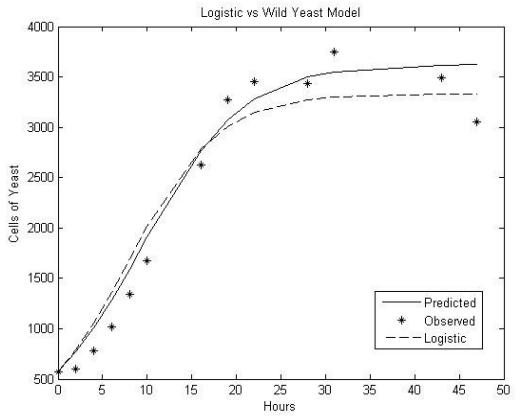


Figure 5: Predicted Logistic and Wild Yeast Model with the mean of the validation data from 2012. Logistic R^2 value is .949 and Wild Yeast R^2 value is .95.

3.2 Student Results

As mentioned above, each student was required to write a report. Each group was able to successfully implement a model and find parameters to produce a curve that predicted the validation data well. Also, the different groups used their own methods for parameter estimation and discussed how their alternate models predicted the validation data. Group I was pleased to see their model actually display the decrease in yeast concentrations. That was their goal to begin with and seeing the results was rewarding. Group II learned quite a bit about finding parameters that work for the two models they presented. At the end of the day, they too were satisfied they had found models which successfully predicted the validation data.

3.2.2 Student Responses

A biology student in Group I explained how before doing the lab she knew "a ton about yeast and their population dynamics but it was neat to actually see it and model it". A biology student in Group II stated that even though he knew about population dynamics, he gained a better understanding of limiting substrates and inhibitory factors. Learning to become creative in putting a model together seemed to be difficult for most the students. A participant with more math experience shared how the lab helped to solidify what she knew about the logistic equation, but the most valuable information was the meaning behind the parameter values and their role in explaining population dynamics.

I realized through much work and puzzled thoughts, completion of the lab provided much satisfaction. Being part of a group who created a model implemented into real data provided another dimension of understanding to mathematics. Learning and gaining experience on how to think through a problem which does not have an exact or expected solution pushed me out of a comfort zone. I gained confidence in the experimental process realizing that all directions of thinking provided some

insight to help me achieve the overall goal of the lab. Though all the errors and success made from beginning to end, the exploration process was enlightening and helped fuel confidence for future work.

4 DISCUSSION AND CONCLUSION

The yeast lab led students to improve understanding of population dynamics with a limiting factor as well as model creation and parameterization. The participants observed the dynamics of yeast in solutions of differing sugar concentrations. They were expected to parameterize both the logistic model and another model they produced themselves. Three innovative models, the Alcohol Death, the Two Sugars and Wild Yeast models were put together and parameterized by students in 2012. These participants reviewed their results from this exercise and discussed their findings and learning. The three models successfully mirrored yeast growth and displayed student understanding of the task. Their methods and thought process for building and parameterizing these models demonstrate their improved understanding of population dynamics and their ability to construct a model from a possible underlying story. The students became agents for their own learning and understanding forcing them to connect pieces of knowledge they already possessed with knowledge they were currently learning and explore different model options.

The students also had their own ideas of what they were interested in learning from an authentic modeling experience. Throughout various discussions with students in the AMB course, I found three main goals biology and ecology students were looking to achieve from modeling exercises. The first being a basic knowledge of mathematical models to permit understanding in readings; second, learning how to create models for their data to tell a story; lastly, learning computer programing skills in order to reconcile models and data. Students of mathematics wish for skills in executing a lab, knowing what

to measure and how to best maximize data collection for use with modeling, and experience with mathematics in real world applications. Completion of this lab allowed students to gain some of the skills they mentioned. After the lab students reflected on the things that they learned.

One student noted that he will no longer skip the model sections in the scientific papers he reads. Before this experience, these models were confusing and getting lost proved easy. The skills formulated during this lab afforded him the confidence and knowledge to tackle these sections providing him an outlet for further understanding in his subject. Learning how to ask questions to promote thought and progress in problem solving is another skill a student learned from this lab. These are examples of how these skills can be expanded beyond modeling.

Exposure to mathematical modeling is an essential tool for all students of science, not just those in math. A math biology student taking part in this process stated, "I felt like I was starting to learn how to create a mathematical model for a physical phenomenon, which was very exciting to me. It was very interesting to see the traditional models used and where they came from and to be able to analyze their faults myself."

Given the opportunity to have these experiences students will excel and accept challenges presented to them. They will gain skills that will benefit their learning throughout the rest of their educational studies and throughout their careers. The construction and execution of the lab pushes students to stretch their knowledge and creativity. This freedom will allow students an opportunity to work and learn without limits. This is evident with the students in Group II who found two different paths to pursue to model the data, one resulting in the Two Sugars Model and one with the Wild Yeast Model.

Biembengut and Hein (2010) state that teaching mathematical modeling, although time

consuming for students and instructors, provides both a better chance of success, "becoming one of the chief agents for change." A form of this lab was also done with students in a differential equations course with less experienced students on a shorter time scale. Appendix B contains information on how the lab was presented as well as the students results. Applying this lab in any of its forms will allow instructors to enhance their teaching and give students the tools they need to succeed throughout the rest of their scientific education and careers.

REFERENCES

Biembengut, MS, and N Hein. 2010. Mathematical Modeling: Implications for Teaching. Pages 481-90 *in* R. Lesh, P.L. Galbraith, C.R. Haines, A. Hurford, editors. Modeling Students' Mathematical Modeling Competencies: ICTMA 13. Springer, New York.

Feldmann, H. 2010. Yeast: Molecular and Cell Biology. Pages 1-2, 4. Wiley-VCH, Weinheim.

Fix, GJ. 1999. Principles of Brewing Science: A Study of Serious Brewing Issues. Brewers Publications, Boulder, CO.

Fraenkel, DG. 2011. Yeast Intermediary Metabolism. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Gause, GF. 1934. The Struggle for Existence. Williams and Wilkins, Baltimore, MD.

Hornsey, IS 2007. The Chemistry and Biology of Winemaking. Page 114. Royal Society of Chemistry, Cambridge.

Moschkovich, JN. 2004, Appropriating mathematical practices: A case study of learning to use and explore functions through interaction with a tutor. Educational Studies in Mathematics: 49–80.

Powell, JA, BR Kohler, JW Haefner, and J Bodily. 2012. Carrying BioMath Education in a Leaky Bucket. Bulletin of Mathematical Biology 74.9: 2232-264.

Reed, G, and TW Nagodawithana. 1991. Yeast Technology. Pages 89-90, xi. Van Nostrand Reinhold, New York.

Unger, RW. 2004. Beer in the Middle Ages and the Renaissance. Page 152. University of

Pennsylvania, Philadelphia.

Verhulst, PF. 1838. Notice sur la loi que la population suit dans son accroissement. Correspondance Mathématique et Physique 10: 113–121.

Voet, DJ, and JG Voet. 2004. Biochemistry. J. Wiley & Sons, Hoboken, NJ.

Zbiek, RM, and A Conner. 2006. Beyond Motivation: Exploring Mathematcial Modeling as a Context for Deepening Students' Understandings of Curricular Mathematics. Educational Studies in Mathematics 63.1: 89-112.

APPENDICES

APPENDIX A

Lab Execution and Data Collection

Growth in yeast was calculated through yeast densities. These densities were measured through absorbency of light in a spectrophotometer, counting cells, or even the volume of balloons. To set up the experiment 100 ml Yeast extract - Peptone –Dextrose (YDP) solution, without the dextrose, or some other medium in which yeast can grow, was distributed into 250 ml Ehrlenmeyer flasks. Dextrose is a sugar and to allow sugar concentrations to be altered and controlled it needs to be absent from the medium. The experiment required treatment groups which have added glucose, and a control group, each having at least three replicates. Three of the flasks, already containing YDP, received 20 g/L of glucose to each, another three received 5 g/L of glucose each and 1 g/L each to yet another three flasks. Three flasks were left without any added glucose to be the control group. Then, 10 g/L of glucose was added to a final three flasks to use as the validation data. The validation set of data is the data used to test the models on how well they predict the yeast growth. Each flask was clearly labeled with the treatment type and replicate number.

Using the stock culture, a preliminary count was done, replicated by three observers, to determine the density in the culture. This allowed for calculating the proper volume to bring the densities in each flask to desired counts. The students aimed for starting population densities around 30 cells/microliters in a target volume of 150 ml. Yeast and YPD were added to each flask until all contain a volume of 150 ml. The flasks were covered with saran wrap to keep as much oxygen from interfering with growth. Again, replicated counts were made and initial conditions recorded. There was fifteen flasks in total. You are not constrained to this set up although this is template is ample.

In this paper, yeast densities were determined using a spectrophotometer which observes how much 600nm of light is absorbed and the cell density absorbing it. In order to use a spectrophotometer this relationship needs to be established. To do this yeast counts were done under a microscope using hemacytometers, first with the with the mixed stock culture, than with at least four different dilutions, such as 1:20 (ratio of yeast solution to YDP) 1:10, 1:5 and 1:1, whose levels fall in the 10-90% absorbance range. To achieve the best results two or three different students did a count for each dilution. The counts were then plotted against the dilutions and a curve was fit, which is called the calibration curve. Due to the fact yeast can grow to high densities dilutions were used to stay within that 10-90% range for accuracy and the calibration curve was used to standardize the readings.

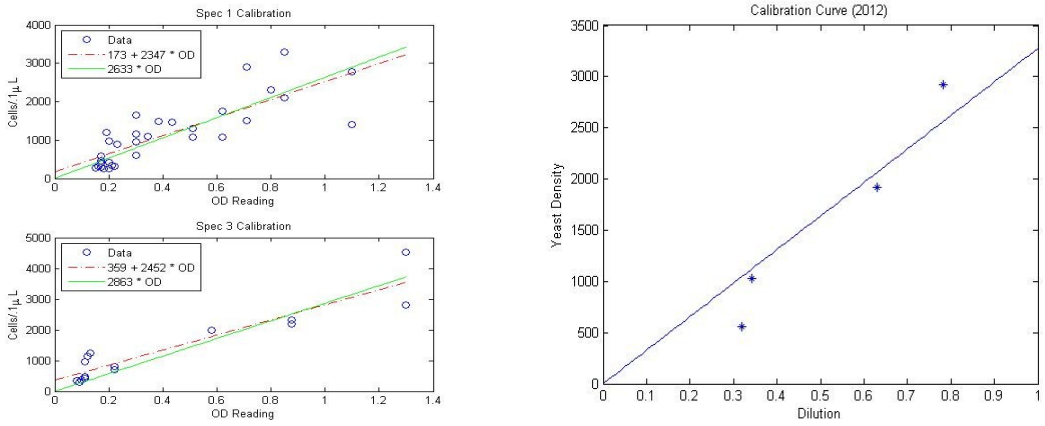


Figure 1

Left: Calibration curves used for the 2010 data. Right: Calibration curve used for the 2012 data. Notice the fitted lines for each year are forced through (0,0).

Again, because yeast grows speedily, measurements of yeast density occurred every two or three hours. To capture enough data, measurements were taken over a time frame of at least 50 hours. Students aided in the measuring, where dilutions and abortions were recorded at each measurement.

APPENDIX B

Lab done in Differential Equations class.

Part I Set up

Given two sets of data, a set with a high amount of sugar added and a set with a lower amount of sugar added, the students were asked to fit the logistic equation to these data sets. With their knowledge of differential equations they parameterized the rate and carrying capacity by fitting a quadratic to the observed data versus the derivative and a line to observed data versus the derivative divided by the observed data.

Differential equations and calculus classes have a larger variety of majors. Below are some of their prospective career paths and what they learned most from the activity.

- Cellular and molecular biology: I learned new ways to find parameters and the repetition of the modeling theme helped with my understanding of the concepts quite a bit
- Economics major: I learned skills on excel which will be useful for my career and understanding models is critical for my intended profession
- Mechanical engineering: It is nice to know how we can apply the differential equations that we are learning to our careers.
- Mechanical engineering: I couldn't see the exact use of this in my field however learned critical thinking and practicing manipulation equations to make them do what we want them to do.

Overall, the skills that can be learned from this lab extend way beyond the ability to model.

Part II Parameterization

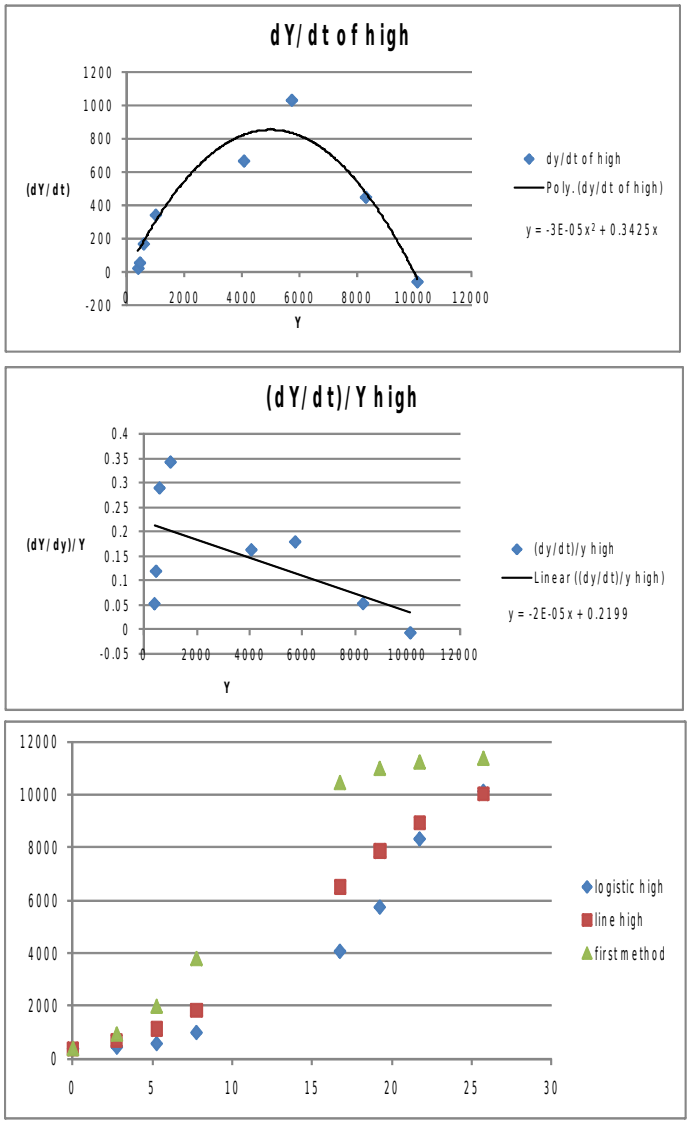
To calculate the parameters the instructor first reviewed line regression and how to manipulate the equations to find parameters using lines. Take for instance the logistic equation, Equation 10

$$\frac{dY}{dt} = rY\left(1 - \frac{Y}{K}\right) \quad . \tag{10}$$

Given already collected data, the concentration of yeast, *Y*, and the time of collection, *t*, is known giving values for the left hand side of Equation 10 and values for *Y*. Divide both sides by *Y*. The only two unknowns are *K* and *r*. Calculation of $\frac{dY}{dt}$ was done by finding the change in Y $(Y_i - Y_{i-1})$ over the change in time $(t_i - t_{i-1})$. Note, this yields one less data point to use for values of *Y*. Dividing by *Y* gives

$$\frac{dY}{dt\,Y} = r - \frac{rY}{K} \tag{11}$$

which is just a line, $y = mx + b$ where $y = \frac{dY}{dt\,Y}$, $x = Y$, $b = r$ and $m = \frac{-r}{K}$. Now we can solve for these parameters. This can be done easily in Microsoft Excel or Matlab. Also, you could fit a quadratic, exactly Equation 10, in the same fashion.



Figure

Work of students in College differential equations class. Fitting logistic equation to data set with high added sugar.

Top: Fitting of quadratic function with observed data against its change over time.

Middle: Fitting a line for the observed data versus the change over time divided by observed.

Bottom: Diamond data follows logistic pattern

Square: Predicted logistic found with parameters found from fitting the line.

Triangle: Predicted logistic with parameters found from fitting the quadratic.

APPENDIX C

Example Matlab Code for Parameterization

```
function err=ErrorSugarWild(data,time, parameters)

parameters=abs(parameters);
a=parameters(1);  %growth wild yeast
b=parameters(2);  %growth yeast
c=parameters(3);  %sugar and wild yeast
d=parameters(4);  %sugar and yeast
q=parameters(5);   %initial number of wild yeast


tobs=time; n=length(tobs);

%%% Find predicted for High sugar
yh1obs=data(1,:)';
yh2obs=data(2,:)';
yh3obs=data(3,:)';

yh1pred=0*yh1obs;
v0=[data(1,1); q; 1.02];
[t, yh1soln]=ode45(@Sugar, tobs, v0);
 yh1pred=yh1soln(:,1);

yh2pred=0*yh2obs;
v0=[data(2,1); q; 1.02];
[t, yh2soln]=ode45(@Sugar, tobs, v0);
 yh2pred=yh2soln(:,1);


yh3pred=0*yh3obs;
v0=[data(3,1); q; 1.02];
[t, yh3soln]=ode45(@Sugar, tobs, v0);
 yh3pred=yh3soln(:,1);

 % Calculate the sum of the error squared
 err=sum((yh1pred-yh1obs).^2 ...%);
  + (yh2pred-yh2obs).^2  + (yh3pred-yh3obs).^2);

  %%% Find predicted for low sugar
 yl1obs=data(4,:)';
 yl2obs=data(5,:)';
 yl3obs=data(6,:)';

 yl1pred=0*yl1obs;
 v0=[data(4,1); q; 1.005];
 [t, yl1soln]=ode45(@Sugar, tobs, v0);
  yl1pred=yl1soln(:,1);

 yl2pred=0*yl2obs;
```

```
 v0=[data(5,1); q; 1.005];
 [t, yl2soln]=ode45(@Sugar, tobs, v0);
  yl2pred=yl2soln(:,1);

 yl3pred=0*yl3obs;
 v0=[data(6,1); q; 1.005];
 [t, yl3soln]=ode45(@Sugar, tobs, v0);
  yl3pred=yl3soln(:,1);

% Calculate the sum of the error squared added to error from high sugar

  err= err + sum((yl1pred-yl1obs).^2  + (yl2pred-yl2obs).^2  + (yl3pred-
yl3obs).^2);

  %%% Find predicted for added sugar
 yc1obs=data(7,:)';
 yc2obs=data(8,:)';
 yc3obs=data(9,:)';

 yc1pred=0*yc1obs;
 v0=[data(7,1); q; 1];
 [t, yc1soln]=ode45(@Sugar, tobs, v0);
  yc1pred=yc1soln(:,1);

  yc2pred=0*yc2obs;
 v0=[data(8,1); q; 1];
 [t, yc2soln]=ode45(@Sugar, tobs, v0);
  yc2pred=yc2soln(:,1);

  yc3pred=0*yc3obs;
 v0=[data(9,1); q; 1];
 [t, yc3soln]=ode45(@Sugar, tobs, v0);
  yc3pred=yc3soln(:,1);

% Calculate the sum of the error squared added to error from high and low sugar

  err= err + sum((yc1pred-yc1obs).^2  + (yc2pred-yc2obs).^2  + (yc3pred-
yc3obs).^2);

    function rhs=Sugar(t, x)

        %x(1)= y
        %x(2)= w
        %x(3)= s
        rhsy=b*x(3)*x(1);
        rhsw= a*x(2)*x(3);
        rhss=-( d*b*x(3)*x(1) + c*a*x(2)*x(3) );
        rhs=[rhsy; rhsw; rhss];
    end
end
```