

# Yeast for Mathematicians – A Ferment of Discovery

Matthew Lewis\* and James Powell†  
Department of Mathematics and Statistics  
Utah State University

April 6, 2016

## 1 Introduction

There is a large gap between mathematics as it is usually taught in postsecondary settings and the professional needs of practitioners in academia and industry. College math classes focus on classical knowledge developed (mostly) 150 years ago, and homework is largely rote, focussing on algorithmic skills, factual knowledge and vocabulary. In direct contrast, a variety of professional, educational and industrial groups have pointed out the need for alternative skills among STEM professionals [5], [7], [18], [24]. These skills include abilities to:

- Develop models and adapt mathematical techniques to novel situations,
- Draw meaning from data,
- Think critically and creatively about quantitative problems,
- Use computational tools to provide timely answers,
- Work effectively in teams,
- Communicate results clearly and concisely.

Over two thirds of employers [24] value these skills among STEM graduates, as compared with only 18% valuing traditional math skills. And while the traditional skills are undoubtedly prerequisite for post-secondary STEM educators, similar integrative, creative and ‘soft’ skills are necessary for research and professional success.

---

\*email: [matthew.lewis@aggiemail.usu.edu](mailto:matthew.lewis@aggiemail.usu.edu)

†email: [jim.powell@usu.edu](mailto:jim.powell@usu.edu)

This is particularly evident in mathematical biology where there are few accepted models and mathematics is continually being developed to describe and explain data generated by biological innovations. Perhaps no other mathematical discipline so depends on non-traditional mathematical skills to make progress. It has been argued [20] that the optimal mental state for a student to learn such skills is precisely the mental state in which a professional applied mathematician would approach a novel biological problem. To make this possible in a classroom context we have created a number of Laboratory Experiences in Mathematical Biology (LEMBs) that allow students to generate their own novel solutions and learn by emulating professional mathematical biologist behavior rather than passive absorption through lecture. Existing LEMBs deal with thermoclines in lakes and oceans ([3]), movement of brine shrimp ([14]), disease dynamics ([17]), and compartmental flows illustrated by leaky buckets ([19]).

While the current set of LEMB resources highlight and encourage development of key skills for math-biologists a lab focused on modeling populations is missing. While there are many data sets available for populations, we believe that collection and organization of data by students is critical for LEMB success ([14], [20], [19]). Nevertheless, data collection in a typical mathematics classroom is routinely problematic. This is especially true concerning modeling populations, a fundamental component of mathematical biology. Generally organisms are too big or too small or require special conditions, and the time scale over which dynamics occur is often too drawn out to be tenable for a math class.

Using yeast as a model population organism addresses many of these difficulties. Yeast grow rapidly in easy media and require no special care. On the one hand, yeast is an ideal candidate for students to model mathematically. As Gause states in *The Struggle for Existence* (1934) "...yeast cells are sometimes subject to perfectly definite quantitative laws. But it has also been found... their trends often do not harmonize with the predictions of the relatively simple mathematical theory." These characteristics allow students to initially engage in modeling yeast dynamics with relative confidence but still challenges students to critically and creatively adjust their models in light of their data. However, yeast are unicellular. Getting population counts directly can be quite

a chore, requiring either microscopes and tedious counting or dilution and photospectrometry, both of which are problematic in math classrooms.

Rather than using specialized equipment, we have developed a lab in which yeast is grown in a small, capped flask, generating carbon dioxide which is trapped in an inverted jar for measurement purposes. The volume of carbon dioxide produced can either be measured directly or by using a time-lapse photo application on a tablet computer. The simple setup produces data that easily integrates into math courses and allows students to behave like a research faced with new biological phenomena.

One big hurdle is the students' natural assumption (and in a typical classroom, correct assumption) that they only need to convince their teacher and not necessarily their peers. That is not usually the case for mathematical researchers, who are primarily concerned with understanding phenomena and persuading their peers. The educational success of a LEMB depends strongly on the instructor being able to adopt a collaborator/mentor role to facilitate student exploration and creativity. Our most successful lab experiences occur when professors "scaffold" student learning, helping them to develop models and techniques to apply these models but *not* directing or judging their progress. This technique is particularly helpful when students have the perception that they are exploring new territory.

To help motivate students and encourage a collaborator/mentor role for instructors we propose model competition as a scientific backdrop. Model competition is a scientific paradigm gaining prevalence in ecology, due in part to the difficulty of performing reductionistic experiments to falsify particular hypotheses in ecological systems. In model competition multiple hypotheses are advanced simultaneously (in the form of models) and allowed to compete in an arena comprised by common data [11]. Quality of fit is adjudicated by information-theoretic metrics like the Akaike Information Criterion (AIC, [2]) or Bayesian (or Schwarz) Information Criterion (BIC, [22]), each of which balances elegance (simplicity, measured as number of parameters) and goodness of fit (measured as deviance between model and data).

For our purposes, model competition, where the goodness of a model is not decided by

the instructor but by a common metric, is a simple means of putting students and instructors in a creative and collaborative frame of mind. Model competition magnifies the instructor's roles as mentor and advisor while minimizing the lecturer and grader components. For our purposes, the Bayesian Information Criterion is generally an appropriate measure for determining which model is best. BIC encourages models that fit the data well with few parameters, has a simple formulation, and can easily be conveyed to upper division students. Given that is commonly used by professional modelers, using BIC to moderate model competition is an authentic way to facilitate lab-based pedagogy in a math class.

In this paper the materials used to launch the Yeast Lab are discussed as is the simple logistic model adapted to  $\text{CO}_2$  observations. BIC is developed in a context suitable for undergraduates. Some of the approaches produced by students in our mathematical biology lab are discussed to give instructors an idea of what students come up with to better scaffold student learning in their own classes. Pedagogical guidance and support is provided so that the Yeast Lab can be incorporated as a real-life example of modeling and model competition into mathematics, statistics and biology courses.

## 2 Launching the Lab

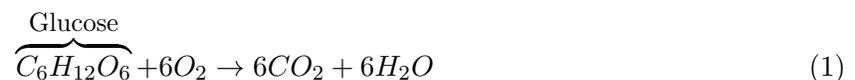
The Yeast Lab is normally run at the same time or just after a computational lab covering techniques of numerical solutions to differential equations. Before the lab is executed, students spend time discussing simple growth models (e.g., linear, exponential) along with parameters (e.g., growth rates, initial conditions) and independent/dependent state variables. How per-capita growth and resource utilization lead to some more advanced population growth models, particularly the logistic equation, is also considered. The behavior and solution of the logistic model is examined as well as methods of parameter estimation (e.g., plotting per-capita growth and using linear regression, using the knee and tail of the curve separately to get intrinsic growth and carrying capacity, logistic regression, or nonlinear fitting techniques) and the ecological meaning of parameters (intrinsic growth rate and carrying capacity). Below we provide a brief introduction to yeast and a derivation

and parameterization of the logistic model in the context of the Yeast Lab.

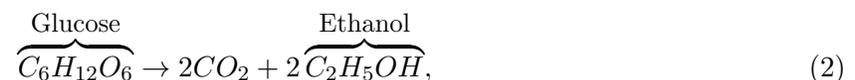
## 2.1 Why Yeast?

Considering that yeast is one of the oldest domesticated organisms, it makes sense that yeast has a long history in mathematical biology. Yeast is popular due to its practical importance in the food and beverage industry, and as a eukaryotic organism it shares many cellular characteristics and mechanisms with more complicated multicellular eukaryotes, like humans. Pragmatically, yeast works as a model experimental organism because it grows readily on simple media, reproduces by budding and gives investigators absolute control over its environmental parameters. While yeast will always be primarily known for its service in producing delicious breads, beers and wines it has a long experimental history. Notably, Gause relied on yeast in *The Struggle for Existence* ??, which is a classic in quantitative population ecology. Yeast was one of the first organisms to which molecular approaches were applied to in the 1950's, and in 1996 yeast (*Saccharomyces cerevisiae*) was the first eukaryote to have its genome sequenced. Ultimately yeast represents “an ideal system to investigate cell architecture and fundamental cellular mechanisms” and thus, yeast has been a common experimental choice [6].

Of the many similarities yeast share with other eukaryotes, cellular respiration is particularly important. Producing energy for most organisms is all about creating adenosine triphosphate (ATP) from sugar molecules. ATP is the basic “molecular unit of currency” of intracellular energy transfer [13], providing energy for most cellular functions, including synthesis of proteins and assembly/disassembly of cellular structures. In the presence of oxygen, the oxidation (burning) of glucose gives



and during this process up to 38 ATP are created (although practically speaking the yield is more like 30 ATP). In the absence of oxygen, however, fermentation occurs and



which produces only 2 ATP – about 15 times less efficient than the aerobic pathway. Presumably the

energy generated keeps the yeast cell alive and allows it to bud (reproduce – yeast cells make little buds, which fall off and become adult yeast cells). As alcohol (ethanol) is produced fermentation declines and as the percentage increases *S. cerevisiae* dies. To generate higher concentrations of alcohol (above 14%) either alternative species of yeast or distillation is required.

After discussing yeast biology we play a video of a past experiment and/or show the data so that students get a sense of what to expect. The class experiment, or experiments in individual groups, can then be set up. While the yeast are fermenting, student groups will meet and create an alternative model (i.e. substantially non-logistic) for population growth. Students need to formulate strategies for explaining their model, generating solutions, finding parameters, and comparing their model with a parameterized version of the logistic model (and data).

## 2.2 Student Expectations and Lab Agenda

In Utah State University's Math-Biology Lab students confront real-world experiments as mixed-educational teams. Students are typically comprised of upper-class undergraduates and graduate students from math, statistics, biology, natural resources and biological engineering. For the Yeast Lab students are divided into teams so that each team has a member with deeper biological background and someone with exposure to nonlinear fitting techniques and numerical methods for solving differential equations.

Teams are expected to:

- Create a model that is significantly different than the logistic model to predict the height of the CO<sub>2</sub> column generated from the growing yeast population.
- Calibrate the model (estimate the parameters) using the collected data.
- Calculate BIC for both the logistic model and the students' alternative model.

We ask students (or student groups) to produce a short written report or present their findings via PowerPoint/Beamer. The reports should include:

- Their alternative model, with a mechanistic explanation for terms.
- Description of solutions and solution procedure, as well as how solution curves do/do not reflect observations.

- Description of parameters required (and their units) as well as procedure used to estimate them.
- A graphical comparison of the logistic and alternate models, along with the data.
- An answer to the questions:
  - Which model better reflects the data, and why? How does this confirm/invalidate any assumptions that you made for your alternate model?
  - For what else could you use this modeling approach?
  - What did you learn from this experience?

Loosely, the in-class portion of the Yeast Lab proceeds as follows:

1. Lecture: Yeast Lab introduction and data collection setup [15 min]
2. Lecture: Derivation of logistic model [20 min]
3. Group Time: Discussion and development of alternate models [20 min]
4. Lecture: Derivation of BIC and computation of logistic model BIC [15 min]
5. Class Discussion: Groups present alternate models, calibration strategy and BIC score [45 min]

This agenda is typically accomplished over the course of a few class periods with the expectation that students are meeting and discussing their alternative models and calibration strategies. The details of the schedule can be compressed or expanded as needed (e.g., setup and recording of reaction can occur before class). If the lab is used as an example of the utility of mathematical technique (e.g., separation of variables or numerical solutions to ODEs) one of the models discussed below can be provided for the class to work with along with data collected in/before class. Assessment of the students' work is done either via written report or oral presentation.

## 2.3 Lab Materials and Methods

### 2.3.1 Materials

Ideally, each group will be able to create and record their own data. This allows groups to get a hands on connection with the data (which is useful when it comes to modeling), potentially reserve a data set as the competition/validation data and protect against a potentially bad batch of yeast. Depending on the class, students can either organize themselves and take turns recording

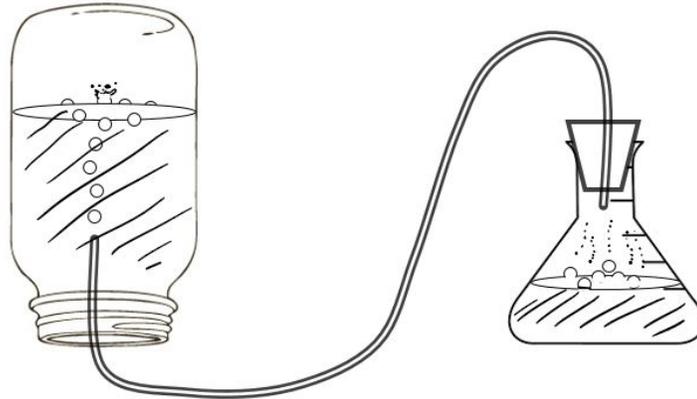


Figure 1: Diagram of the experimental apparatus. Yeast in the flask produces  $\text{CO}_2$ , which is captured in a Mason jar for measurement purposes. Fixing a ruler to the side of the jar and dyeing the fluid in the jar is very helpful for collecting data.

the data every 15 minutes over the 9 or 10 hours required or iPads, or similar, can be used with a time-lapse application to record the data.

The following materials are needed (for each group):

- Approximately 500 ml flask, with cork and flexible tube.
- 1 quart Mason jar (or similar).
- 1 plastic container, maybe about 2 inches tall, large enough to comfortably contain the Mason jar with two or more inches of clearance around the jar.
- iPad (or similar) with charger and application for time-lapse pictures.
- Food coloring.
- Ruler approximately as tall as Mason jar, two rubber bands. Make sure the ruler and its markings make a nice contrast to the food coloring.
- 1 package dry yeast (1/2 tsp or 1.4 gm).
- 1.5 tsp granulated sugar (6 gm).
- 300 ml distilled water (room temperature).

### 2.3.2 Methods

1. Put 300 ml of distilled water in flask and inoculate with 1/2 tsp of dry yeast. Swirl to mix and allow the yeast to rehydrate for 5-10 minutes.
2. While this is happening, organize the visualization apparatus:

- (a) Fill the Mason jar to brim with tap water and a few drops of food coloring for visualization purposes. Cover with plastic container and with both hands flip over so that the Mason jar is inverted and full of fluid. Add a little more colored liquid (around 1cm deep) to help maintain a seal.
  - (b) Affix a ruler to the side of the Mason jar with the rubber bands.
  - (c) Set up the iPad application to take a picture every 15 minutes, and then place the iPad (plugged in so it won't run out of juice) to get a good view of the ruler on the side of the jar. Make sure the area will be lighted during the next 24 hours (either leave the room light on or place a desk lamp nearby to illuminate).
3. Add the 1.5 tsp sugar, swirl again, then cap the flask using a stopper with surgical tubing already attached.
  4. After swirling has settled down and yeast has begun to bubble ( $\sim 5$  min) snake the surgical tubing underneath the lip of the Mason jar as seen in figure 1, being careful not to lose the seal on the liquid. Rest the jar back down on the tubing (it will no longer sit perfectly straight – you can fix this if you want by slipping a couple of short, cut pieces of tubing underneath the edges of the jar). You may want to release enough liquid from the jar so that the fluid height is at a uniform cross-section of the jar.
  5. Start the iPad application and take some data! Example data appears in Figure 2.

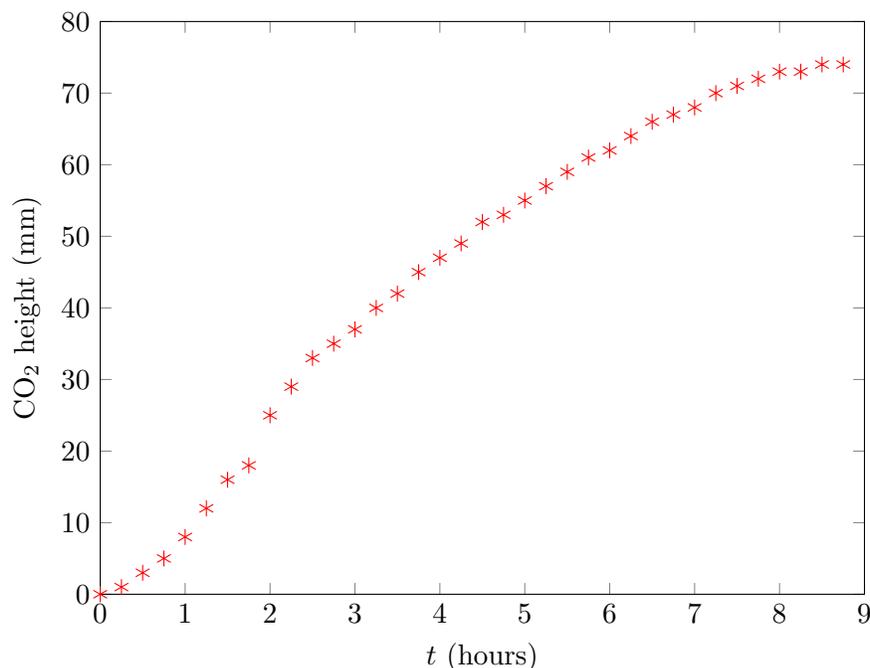


Figure 2: Measured heights of CO<sub>2</sub> produced by yeast growing on sugar in a flask. Note how the data looks somewhat logistic but is not perfectly symmetric. Data was collected by time-lapse video on an iPad, with frames taken every 15 minutes.

## 2.4 Logistic Model

The logistic model is a traditional starting point for describing the yeast populations. The model was originally developed by Pierre-François Verhulst in 1838 [27] to describe when the rate of reproduction is proportional to both the existing population and the amount of available resources. The logistic model was used extensively by Gause, one of the first to study yeast quantitatively. In *The Struggle for Existence* [9] Gause published what is now known as the competitive exclusion principle, or Gause's Law, which states that when two species compete for same survival requirements, the more efficient species will reproduce at a higher rate and drive the less efficient species towards local extinction. In the process of developing the competitive exclusion principle, Gause ran a series of experiments that validated the logistic growth equation for yeast in an environment with a limited nutrient.

The logistic equation can be derived by keeping track of two dependent variables:  $Y$ , the population density of yeast cells in the solution, and  $S$ , the concentration of sugar in the solution. The model will assume that the solution is well-mixed, with no spatial structure to either variable, and that temperature, oxygenation, etc. are held constant.

### 2.4.1 Conservation and Per-Capita Growth

A model of the interaction between yeast and sugar must account for two facts: sugar must be used for the yeast population to grow, and the rate of population growth is proportional to the size of the population. The first of these two facts basically means that for every increase of the yeast population there is a corresponding decrease in the sugar concentration, which may be written mathematically as

$$\dot{Y} = -a\dot{S}. \tag{3}$$

(Recall that  $\dot{Y}$  is the same as  $\frac{dY}{dt}$ ). Here  $a$  is the amount of sugar required to produce a fuzzy baby yeast. The second effect can be written

$$\dot{Y} = g(S)Y,$$

where  $g(S)$  is the per-capita growth rate of the yeast population, which depends on the sugar concentration. Assuming  $g \propto S$  gives a second equation for rate of population growth,

$$\dot{Y} = bSY. \quad (4)$$

The parameter  $b$  can be interpreted as the rate of growth per sugar concentration.

### 2.4.2 Derivation of the Logistic Equation

Equations (3) and (4) can be reduced to a single nonlinear equation. Let  $Y(t = 0) = Y_0$  and  $S(t = 0) = S_0$ , and integrate both sides of (3):

$$-a \int_0^t \dot{S} dt = \int_0^t \dot{Y} dt \quad \Rightarrow \quad S(t) = S_0 - \frac{1}{a}(Y(t) - Y_0) = S_0 + \frac{Y_0}{a} - \frac{Y(t)}{a}.$$

Substituting this result into (4) and factoring gives

$$\dot{Y} = \frac{b(aS_0 + Y_0)}{a} Y \left[ 1 - \frac{Y}{(aS_0 + Y_0)} \right].$$

Now define

$$K = (aS_0 + Y_0) \quad \text{and} \quad r = \frac{b(aS_0 + Y_0)}{a}$$

to get the standard ecological form of the logistic equation:

$$\dot{Y} = rY \left( 1 - \frac{Y}{K} \right), \quad (5)$$

where  $r$  is the intrinsic growth rate of this population for this density of sugar and  $K$  is the carrying capacity of this sugar solution.

### 2.4.3 Logistic Curve and Relation to CO<sub>2</sub> Data

Using standard separation of variables and the initial condition  $Y(0) = Y_0$ , the logistic equation (5) can be solved to get

$$Y(t) = \frac{KY_0}{Y_0 + (K - Y_0)e^{-rt}}. \quad (6)$$

The trouble here is that students are not measuring yeast concentrations, but rather the height of CO<sub>2</sub> in a jar resulting from respiration of the yeast. A simple model (but by no means the only

modell) for the volume ( $V$ ) of  $\text{CO}_2$  produced would be that the volume of  $\text{CO}_2$  grows in direct proportion to the growth of yeast,

$$\frac{dV}{dt} = A \frac{dh}{dt} = c \frac{dY}{dt},$$

where  $h(t)$  is the predicted height and  $A$  the cross-sectional area of the jar containing the  $\text{CO}_2$ . Integrating both sides of the equation and noting that  $h(0) = 0$ ,

$$A(h(t) - h(0)) = c(Y(t) - Y(0)) \Rightarrow h(t) = \frac{c}{A} (Y(t) - Y_0). \quad (7)$$

Plugging in (6) and factoring gives a prediction for the height of  $\text{CO}_2$  produced by the growing yeast,

$$h(t) = \frac{cY_0}{A} \left[ \frac{K}{Y_0 + (K - Y_0)e^{-rt}} - 1 \right] = \frac{cY_0}{A} \frac{K - Y_0}{Y_0 + (K - Y_0)e^{-rt}} (1 - e^{-rt}). \quad (8)$$

This may look like it has five parameters, but if one divides top and bottom by  $K$  and defines

$$\alpha \stackrel{\text{def}}{=} \frac{cY_0}{A} \quad \text{and} \quad \beta \stackrel{\text{def}}{=} \frac{Y_0}{K}$$

then (8) becomes

$$h(t) = \alpha \frac{1 - \beta}{\beta + (1 - \beta)e^{-rt}} (1 - e^{-rt}). \quad (9)$$

Now there are only three parameters, with units commensurate with the data;  $\alpha$  has units of height (mm),  $\beta$  is a unit-free shape parameter representing initial yeast as a fraction of carrying capacity, and  $r$  is a growth rate parameter with units of  $\text{hrs}^{-1}$ .

#### 2.4.4 Qualitative Behavior and Ballpark Parameters

Even if there are only three parameters, parametrization still takes some effort. To do a good job one should do some maximum likelihood or least squares fitting (see below). But 'ballpark' parameter values can go a long way to help students understand the model. First off, the long time limit of (9) gives

$$h(t) \rightarrow \frac{\alpha(1 - \beta)}{\beta} \stackrel{\text{set}}{=} h_{\max}, \quad (10)$$

where  $h_{\max}$  is the end amount of  $\text{CO}_2$  produced (in the case of the sample data  $h_{\max} \approx 74$ ). This means we can replace  $\alpha$  with something a more observationally relevant,

$$\alpha = \frac{\beta h_{\max}}{(1 - \beta)} \Rightarrow h(t) = \frac{h_{\max} \beta}{\beta + (1 - \beta)e^{-rt}} (1 - e^{-rt}). \quad (11)$$

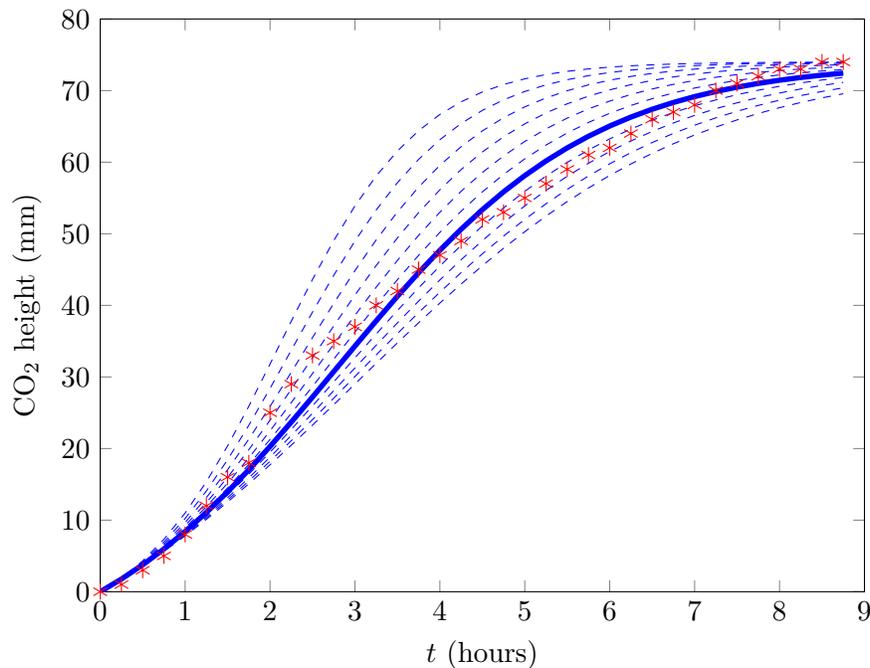


Figure 3: Predicted heights compared with measured CO<sub>2</sub> heights (\*) produced by yeast growing on sugar in a flask. The shape parameter,  $\beta$ , was varied between  $\beta = .075$  and  $\beta = .175$  by steps of 0.01. The highlighted curve, with  $\beta = .135$ , gives a nice visual fit.

In the solution to the logistic equation (6) the initial condition is explicit, but in the case of (10) the initial condition is always zero. However, the slope at zero (let's call it  $m_0$ ) is significant, and reasonably easy to estimate from the first couple of data points. Taking a derivative,

$$h'(t) = \frac{r\beta h_{\max} e^{-rt}}{(\beta + (1 - \beta)e^{-rt})^2} \Rightarrow m_0 \stackrel{\text{def}}{=} h'(0) = r\beta h_{\max}.$$

Plugging  $r = \frac{m_0}{\beta h_{\max}}$  in to (11) gives

$$h(t) = \frac{h_{\max}\beta}{\beta + (1 - \beta)e^{-\frac{m_0}{\beta h_{\max}}t}} \left(1 - e^{-\frac{m_0}{\beta h_{\max}}t}\right), \tag{12}$$

which now has only a single parameter,  $\beta = \frac{Y_0}{K}$ . For the sample data we can estimate  $h_{\max} \approx 74$  and  $m_0 \approx \frac{h(3)-h(1)}{.5} = 6$ ; the effects of varying  $.075 \leq \beta \leq .175$  are illustrated in Figure 3; Something around  $\beta = .135$  seems like a pretty good fit, suggesting that  $Y_0$  is approximately 13.5% of  $K$  in this situation. However, the fit could clearly use some improvement – as Gause suggested, yeast data is full of surprises for modelers! The fact that a standard and accepted model fails to capture the behavior of the data is a springboard for students to develop better models of their own.

### 3 Model Competition

In the process of setting up the Yeast Lab, collecting data, and presenting the logistic-based CO<sub>2</sub> model we challenge students to produce a better model. The question then becomes “What does better mean?” Students are quick to point out that “better” means the winning model hits more of the data points, but many competing models fit well in some places and poorly in others, making model selection difficult. Moreover, it is possible to gain a better fit by adding parameters, but doing so typically results in overfitting. A variety of information theoretic measures of model performance have been developed to assess quality of fit while penalizing for complexity. While Akaike Information Criterion (AIC) [1] is simplest, the Bayesian Information Criterion (BIC) can be derived reasonably in an upper division mathematics class. Presented here is a simple derivation suitable for math juniors and seniors. Alternatively, the BIC or AIC metrics can be presented and discussed qualitatively.

#### 3.1 BIC Derivation

BIC was introduced in 1978 by Gideon Schwarz [22] and is sometimes called the Schwarz Information Criterion. The idea is that in a competition among models, mediated by common data, we should choose the model that is most probable given the data. For our specific *data*, let  $h_j$  represent the observed height of the air column at time  $t_j$  and  $h(t_j; \bar{\theta})$  represent the predicted height at time  $t_j$  generated by the candidate model  $M$  using parameters  $\bar{\theta} = \{\theta_1, \theta_2, \theta_3 \dots \theta_k\}$ .

Using Bayes' Theorem we can write (formally)

$$p(\bar{\theta}|data)p(data) = p(data|\bar{\theta})p(\bar{\theta}),$$

where  $p(\bar{\theta})$  are *prior* probabilities of parameters representing the state of knowledge about model parameterization before the experiment and errors which keep the model from hitting data are captured by the likelihood,  $p(data|\bar{\theta})$ . The *posterior* distribution,  $p(\bar{\theta}|data)$ , for the parameters given the data can be written

$$p(\bar{\theta}|data) = \frac{1}{p(data)}p(data|\bar{\theta})p(\bar{\theta}).$$

The data probability,  $p(data)$ , is unknown but is the same across all models (and therefore irrelevant to model competition). Model probability,  $P(M|data)$ , is the integral of the posterior over all parameter possibilities

$$\begin{aligned} P(M|data) &= \frac{1}{p(data)} \int_{\theta \in \Theta} p(data|\bar{\theta})p(\bar{\theta})d\bar{\theta} \\ &= \frac{1}{p(data)} \int_{\theta \in \Theta} \mathcal{L}(\bar{\theta}; (t_j, h_j)_{j=1}^n)p(\bar{\theta})d\bar{\theta}. \end{aligned} \quad (13)$$

Here  $\Theta$  is the parameter space (or support of the distribution), and we introduce the standard statistical notation for  $p(data|\bar{\theta}) = \mathcal{L}(\bar{\theta}; data)$  as the likelihood function associated with the model and data.

Normally (13) can not be calculated in closed form. However, it can be approximated using maximum likelihood estimates (MLE) for parameters and an asymptotic expansion of the integral. In order to make  $P(M|data)$  as big as possible, we seek the parameters that maximize  $\mathcal{L}$  (i.e., we will follow Fisher's MLE approach [8]). Assuming normal error,

$$\mathcal{L}(\bar{\theta}; data) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(h_j - h(t_j, \bar{\theta}))^2}{2\sigma^2}} = \exp \left[ \underbrace{-\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^n (h_j - h(t_j, \bar{\theta}))^2}_{-NLL(\bar{\theta}|data, \sigma)} \right] \quad (14)$$

where  $NLL$  is the *Negative Log-Likelihood* and  $\sigma$  is the standard deviation of the model error.

The parameters that minimize  $NLL$  also maximize  $\mathcal{L}$ . Let

$$\hat{\theta} = \min_{\bar{\theta} \in \Theta} NLL(\bar{\theta}|data).$$

Then for each parameter  $\theta_i$

$$\frac{\partial}{\partial \theta_i} NLL(\hat{\theta}) = 0, \quad \text{and} \quad \frac{\partial^2}{\partial \theta_i^2} NLL(\hat{\theta}) > 0.$$

Now, consider the integral in (13). Assuming that  $NLL$  is twice continuously differentiable in a neighborhood of  $\hat{\theta}$  we can expand  $NLL$  in a Taylor series,

$$\begin{aligned} NLL(\theta) &= NLL(\hat{\theta}) + \sum_{j=1}^n \left( \frac{1}{2} \frac{\partial^2 NLL(\hat{\theta})}{\partial \theta_1^2} (\theta_1 - \hat{\theta}_1)^2 + \right. \\ &\quad \left. \frac{1}{2} \frac{\partial^2 NLL(\hat{\theta})}{\partial \theta_2^2} (\theta_2 - \hat{\theta}_2)^2 + \dots + \frac{1}{2} \frac{\partial^2 NLL(\hat{\theta})}{\partial \theta_k^2} (\theta_k - \hat{\theta}_k)^2 \right) + \dots \end{aligned} \quad (15)$$

where  $k$  represents the number of parameters. Here we have assumed that the parameters are all independent so that mixed second partials vanish (if not, it is possible to make a change of variables that force the mixed partials to be zero).

Since

$$\frac{\partial^2}{\partial \theta_i^2} NLL(\hat{\theta}) > 0$$

we can scale the parameters such that

$$\frac{\partial^2}{\partial \theta_i^2} NLL(\hat{\theta}) = 1,$$

giving

$$NLL(\theta) = NLL(\hat{\theta}) - \frac{n}{2} \left[ (\theta_1 - \hat{\theta}_1)^2 + (\theta_2 - \hat{\theta}_2)^2 + \dots + (\theta_k - \hat{\theta}_k)^2 \right] + \dots \quad (16)$$

Working this final result into (13) for  $\mathcal{L}$ , we have

$$P(M|data) \propto \int_{-\infty}^{\infty} e^{-NLL(\bar{\theta}|data,\sigma)} p(\bar{\theta}) d\bar{\theta} \propto \int_{-\infty}^{\infty} e^{-NLL(\hat{\theta})} \left( e^{-\frac{-(\theta_1 - \hat{\theta}_1)^2}{n/2}} \dots e^{-\frac{-(\theta_k - \hat{\theta}_k)^2}{n/2}} \dots \right) p(\bar{\theta}) d\bar{\theta}. \quad (17)$$

Schwarz' idea was to look at the dominant exponential contribution to (17), so that the effects of the rescaling outside the exponent can largely be ignored. Additionally, assuming that  $p(\bar{\theta})$  is *noninformative* (i.e., the priors are not informing the parameters) or flat (i.e.,  $p(\theta)$  is not of exponential type), and using

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}},$$

we can write

$$P(M|data) \sim e^{-NLL(\hat{\theta})} \left( \sqrt{\frac{2\pi}{n}} \right)^k = \exp \left[ -NLL(\hat{\theta}) - \frac{1}{2} (k \ln(n) - k \ln(2\pi)) \right] = \exp \left[ -\frac{\text{BIC}}{2} \right]. \quad (18)$$

In traditional BIC it is assumed that  $n \gg 2\pi$  and the constant  $\ln(2\pi)$  may be neglected. The BIC is defined by the exponent:

$$\frac{1}{2} \text{BIC} = -NLL(\hat{\theta}) - \frac{1}{2} (k \ln(n)) \quad \text{or} \quad \text{BIC} = 2NLL(\hat{\theta}) - (k \ln(n)). \quad (19)$$

The  $\frac{1}{2}$  is so that, from an information criterion perspective, a BIC change of one equates to the information effect of reducing variance one standard deviation.

Since BIC captures the negative exponential sensitivity of a model's posterior probability, the model with the *lowest* BIC wins. From a qualitative perspective BIC can be made smaller by a better fit (reducing  $NLL$ ) or by having a more elegant description (fewer parameters, simpler model). Furthermore, when comparing two models, BIC gives a measure of how much better one model is than another. Given Model<sub>1,2</sub> with  $BIC_1 < BIC_2$  then Model<sub>1</sub> is

$$e^{\frac{1}{2}\Delta BIC} = e^{\frac{1}{2}(BIC_2 - BIC_1)} \quad (20)$$

times more probable (this is called the *odds ratio* in favor of Model<sub>1</sub>). Since  $e^{2.3} \approx 10$ , the rule of thumb is that a  $\Delta BIC$  of 5 is convincing evidence that Model<sub>1</sub> is a significant improvement over Model<sub>2</sub>. For greater technical detail regarding BIC, see [23], [12], [16], [10], and [4].

## 3.2 Student Models

We now turn to some of the models created by students from the Fall 2015 Math Biology Lab. These models are presented to illustrate common pitfalls, the range of student creativity, and to help prepare teachers and to scaffold student thinking. Below are a few alternate models generated by former students along with the students' model descriptions. The models were developed in competition with the logistic model as well as one another.

### 3.2.1 Waste Model

The first group focused on the effects of "waste" produced during metabolism. Instead of producing only additional yeast cells, as yeast consume and utilize the sugar also contributes to ethanol production during anaerobic respiration (2), as well as decreasing the pH level as CO<sub>2</sub> bubbles through the solution, making the environment less suitable for yeast reproduction. Rather than model any specific waste product or effect, the students decided that the amount of 'waste' would always be proportional to the sugar used, and the effect would be to reduce growth rates.

The group started with equations for yeast and sugar concentrations as well as the amount of CO<sub>2</sub> produced,

$$\dot{Y} = bYS, \quad \dot{S} = -abYS, \quad \text{and} \quad \dot{C} = cbYS. \quad (21)$$

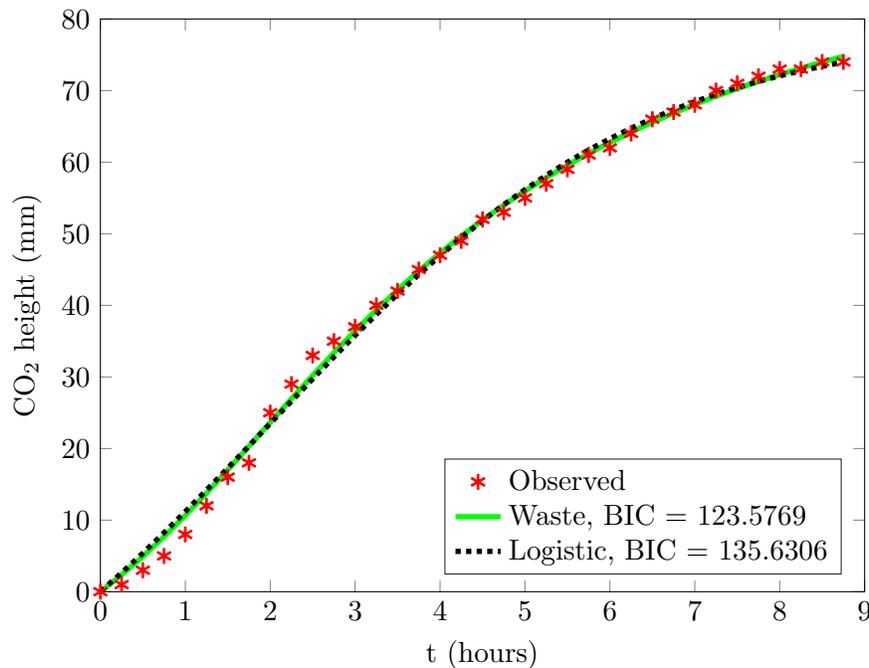


Figure 4: Predicted heights from the Waste Model compared with measured CO<sub>2</sub> heights (\*) produced by yeast growing on sugar in a flask and the initial logistic model derived in class. Parameters  $r_0 = 0.35$ ,  $W_0 = 2.47$ ,  $c = 25.53$  were found using nonlinear least squares and  $a$  was set equal to 1.

To capture how the waste would affect the yeast's growth rate  $b$  in (4) the group proposed (in their words) a “backwards Holling III” equation of the form

$$b(W) = \frac{r_0 W_0}{W_0^2 + W^2} = \frac{r_0 W_0}{W_0^2 + (S_0 - S(t))^2}. \tag{22}$$

In their design the students viewed  $r_0$  as the intrinsic growth rate  $r$  in (5) and  $W_0$  is the threshold where the amount of produced waste decreased yeast population growth by 50%. They made a choice to measure the amount of waste precisely in terms of the amount of sugar metabolized ( $S_0 - S(t)$ ), and their rationale for the absence of a second  $W_0$  in the numerator was that when the waste model (22) was inserted into the system of equations (21) there would be an extra multiplication using units of sugar. Thus their  $r_0$  should be directly comparable to  $r$  in the logistic model. The parameter  $a$  can be interpreted as the amount of glucose necessary to produce a new yeast and  $c$  would be the rate of CO<sub>2</sub> produced per yeast.

The students used  $a$  to make the model dimensionally consistent, but later decided that since

the initial amounts of both sugar and yeast could be measured in grams, they would lose nothing by setting  $a = 1$ . The other parameters  $r_0 = 0.35$ ,  $W_0 = 2.47$ ,  $c = 25.53$  were found using nonlinear least squares (specifically, all the groups ended up using some version of the `fminsearch` function in MATLAB). As seen in figure (4), the model performs better than the initial logistic driven model with a BIC of  $\approx 123.58$  compared to the logistic's BIC of  $\approx 135.63$ . While the students felt they did well, they specifically mentioned how they were frustrated with the model's performance over the first seven or so data points. However, they could not discover a mechanistic means to "bend the curve lower" at the beginning that did not destroy the good fit over the remaining points.

### 3.2.2 Linear (un)Death Model

The second group conjectured that the measured height of  $\text{CO}_2$  released is directly proportional to the yeast cell concentration; they would only need two equations. They felt that the finite life span of yeast would ultimately hinder  $\text{CO}_2$  production and cause the asymmetrical shape of the data. So instead of working with three differential equations like the logistic model presented in class and the Waste Model, this group proposed

$$\dot{Y} = bYS - dY, \quad \dot{S} = -abYS \quad \text{and} \quad \dot{C} = \gamma Y,$$

which reduces to

$$\dot{C} = \beta CS - \delta C \quad \text{and} \quad \dot{S} = -\alpha \beta CS. \quad (23)$$

Here  $\beta$  can be thought of as the growth rate of the  $\text{CO}_2$  column per sugar concentration,  $\delta$  represents the rate at which yeast dying inhibits  $\text{CO}_2$  production and  $\alpha$  is the amount of sugar needed in the mix to produce a 1mm change in height of the  $\text{CO}_2$  column. The students used nonlinear least squares to find their parameters, arriving at  $\beta = 0.1316$  and  $\delta = -3 \times 10^{-4}$ . However, with  $\delta < 0$  the students' term representing the death rate leads to 'un'death and exponential growth, not the linear death they had targeted (see Figure 5).

Another issue the students noted was with the initial conditions. The initial height of the  $\text{CO}_2$  column should be zero, but if they started their model with  $C(0) = 0$  nothing would happen. This was actually pointed out to them during discussions, but they liked their model so much that

they used the second data point as a starting value, thus missing the starting value for CO<sub>2</sub> height. These flaws combined to cause the model to perform poorly when compared to the logistic model (BIC  $\approx 169.05$  compared to the logistic's BIC  $\approx 135.63$ ). In the end students were good-natured regarding their model's failure stating that possibly the yeast had access to alternative resources for growth and therefore a negative death rate is perfectly suitable.

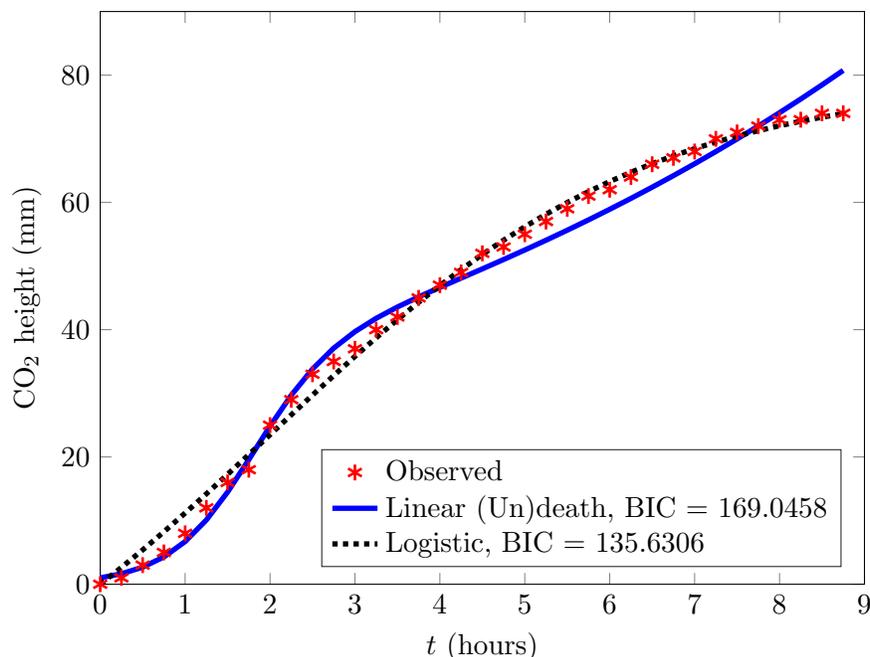


Figure 5: Predicted heights from the (Un)death Model compared with measured CO<sub>2</sub> heights (\*) produced by yeast growing on sugar in a flask and the initial logistic model derived in class. From least-squares approximation students determined the parameters  $\beta = 0.1316$  and  $\delta = -0.0003$  fit the data best. However, with  $\delta < 0$  the students' term representing the death rate leads to exponential growth, not the linear death they had targeted!

### 3.2.3 Pyruvate Model

The next group wanted to account for the differences between aerobic and anaerobic respiration, believing that the growth chamber started out with oxygen in solution which would be used by yeast for more efficient growth until it was depleted. The first step in both processes is glycolysis where sugar molecules are broke down into pyruvate molecules. Once the cell has pyruvate the yeast must continue respiration in either an aerobic or anaerobic direction; this choice is based the presence of O<sub>2</sub> (Figure 7). A cell that can perform aerobic respiration and which finds itself in

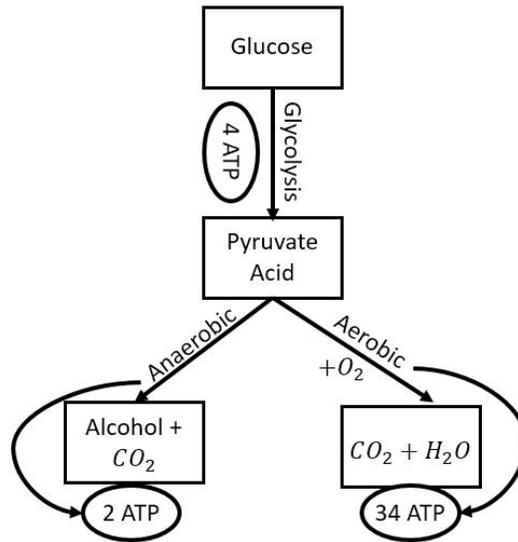


Figure 6: Generating pyruvate through glycolysis is the first step in both the aerobic and anaerobic respiration cycles. Which pathway is chosen depends on the presence of  $O_2$ .

the presence of oxygen will continue on to the aerobic cycle in the mitochondria. If oxygen is not available the yeast will move into anaerobic respiration (or yeast fermentation), which is much less efficient and produces alcohol.

Pulling these ideas together, the students wrote

$$\begin{aligned}
 \dot{S} &= -kYS \\
 \dot{Y} &= -kYS + \alpha_0 \frac{O_2}{O_2 + C} P + \alpha_a \frac{C}{O_2 + C} P \\
 \dot{P} &= kYS - r_0 \frac{O_2}{O_2 + C} P - r_a \frac{C}{O_2 + C} P.
 \end{aligned} \tag{24}$$

Here, the students recognize that the  $k$  used in the  $\dot{S}$  equation cannot be the same  $k$  in the  $\dot{Y}$  and  $\dot{P}$  since the units  $k$  must have to ensure the sugar equation is dimensionally consistent are different than the units needed for  $k$  in the other two equations. However, the students viewed this initial formulation as scratch work since they anticipated nondimensionalizing the equations. The other parameters  $\alpha_0$  and  $\alpha_a$  represent the aerobic and anaerobic growth rate of yeast using pyruvate, respectively.

The students also built equations to deal with the changing environmental conditions within

the chamber to reflect the consumption of  $O_2$  and production of  $CO_2$  that occurs as sugar is converted to pyruvate and ultimately more yeast.

$$\begin{aligned}\dot{O}_2 &= -\beta_o \frac{O_2}{O_2 + C} P \\ \dot{C} &= 6\beta_o \frac{O_2}{O_2 + C} P + 2\beta_a \frac{C}{O_2 + C} P\end{aligned}\quad (25)$$

where  $\beta_o$  represents the rate at which  $O_2$  is consumed in the aerobic respiration pathway and  $\beta_a$  is the rate  $CO_2$  is produced in the anaerobic respiration pathway. The students decided to keep the 2 and 6 as a part of the  $\dot{C}$  equation, reflecting the amount of  $CO_2$  produced in the anaerobic and aerobic pathways respectively as in (1) and (2), although some group members wanted to incorporate those values into  $\beta_o$  and  $\beta_a$ .

The students rescaled the variables to nondimensionalize, using

$$\bar{t} = tr_o, \quad \bar{S} = \frac{Sk}{r_o}, \quad \bar{Y} = \frac{Yk}{r_o}, \quad \bar{P} = \frac{Pk}{r_o}, \quad \bar{O}_2 = \frac{O_2k}{\beta_o}, \quad \bar{C} = \frac{Ck}{\beta_o}.\quad (26)$$

Equations (24) and (25) became

$$\begin{aligned}\bar{S}' &= -\bar{Y}\bar{S} \\ \bar{Y}' &= -\bar{Y}\bar{S} + \gamma_1 \frac{\bar{O}_2}{\bar{O}_2 + \bar{C}} \bar{P} + \gamma_2 \frac{\bar{C}}{\bar{O}_2 + \bar{C}} \bar{P} \\ \bar{P}' &= \bar{Y}\bar{S} - \frac{\bar{O}_2}{\bar{O}_2 + \bar{C}} \bar{P} - \gamma_3 \frac{\bar{C}}{\bar{O}_2 + \bar{C}} \bar{P} \\ \bar{O}_2' &= -\frac{\bar{O}_2}{\bar{O}_2 + \bar{C}} \bar{P} \\ \bar{C}' &= 6 \frac{\bar{O}_2}{\bar{O}_2 + \bar{C}} \bar{P} + 2\gamma_4 \frac{\bar{C}}{\bar{O}_2 + \bar{C}} \bar{P}.\end{aligned}\quad (27)$$

While the students never explicitly resolved the original lack of dimensional consistency in (24) and (25), the dimensionless form of their model was the same as if they had started with consistent dimensions. The students fit the dimensionless equations to the dimensional data for  $C$ , and starting with mass measurements for sugar and yeast. Their reasoning was that a change in units for the sugar and yeast would propagate linearly through the the sequence of equations multiplicatively, resolving itself at the last stage of matching up with the observe  $CO_2$ . With this background, parameters were found using nonlinear least squares,  $\gamma_1 = 0.711$ ,  $\gamma_2 = 0.463$ ,  $\gamma_3 =$

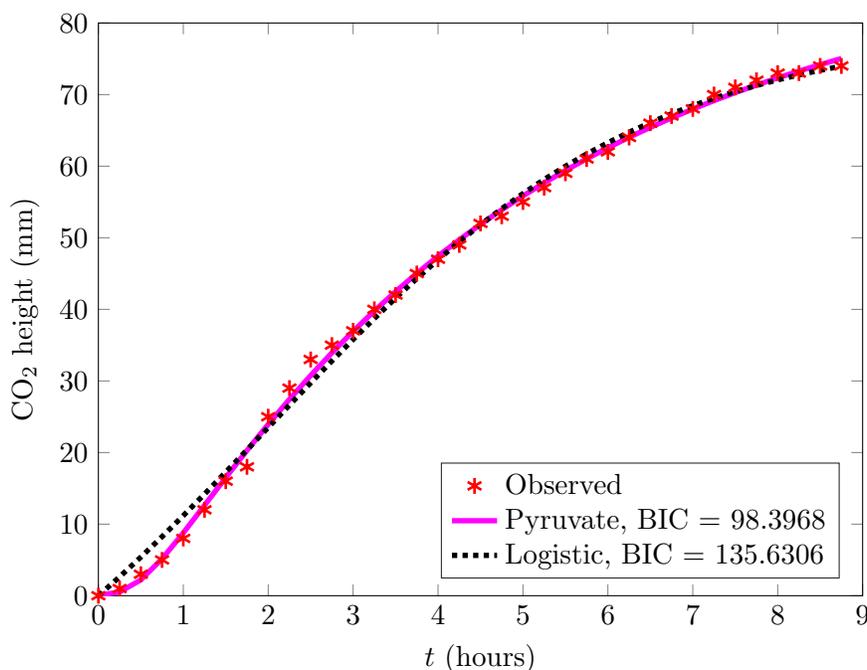


Figure 7: Predicted heights from the Pyruvate Model compared with measured  $\text{CO}_2$  heights (\*) produced by yeast growing on sugar in a flask and the initial logistic model derived in class. Parameters  $\gamma_1 = 0.711, \gamma_2 = 0.463, \gamma_3 = 0.3766, \gamma_4 = 32.4, \bar{O}_0 = 44.3, \bar{C}_0 = -2.062$  were found using `fminsearch` in MATLAB . The model performs well compared to the initial logistic model.

$0.3766, \gamma_4 = 32.4, \bar{O}_0 = 44.3, \bar{C}_0 = -2.062$ . Initially, the students were concerned that the beginning  $\text{CO}_2$  amount is negative ( $\bar{C}_0 = -2.062$ ). However, before the  $\text{CO}_2$  released by the yeast makes it over to the Mason jar and begins displacing the water column it must first fill the reaction chamber. Hence, the students argued that there is a “negative space” that must be filled first. As seen in Figure (7), the model performed well compared to the logistic model (BIC  $\approx 98.40$  for the pyruvate model compared to BIC  $\approx 135.63$  for the logistic model).

### 3.2.4 Oxy-Logistic

The last group’s model is philosophically similar to the Pyruvate model above. Going back to (1) and (2), this group focused on the difference in ATP production (and presumably growth rates) in the aerobic and anaerobic respiration processes. Assuming sugar,  $S$  is in the same units as yeast,  $Y$ , this group started with

$$\dot{Y} = r(O)YS \quad \text{and} \quad \dot{S} = -r(O)YS \tag{28}$$

where the growth rate,  $r(O)$ , a function of oxygen ( $O$ ), is relatively large when oxygen is plentiful and small when oxygen is scarce. The simplest description for the switch is a linear response

$$r(O) = r_a + \frac{r_o - r_a}{O_0} O. \quad (29)$$

Here,  $r_a$  and  $r_o$  are the respective anaerobic and aerobic growth rates,  $O$  is the amount of  $O_2$  amount of oxygen in the solution and  $O_0 = O(0)$ .

For the oxygen component to this model, the group proposed

$$\dot{O} = -fO - \varepsilon OY \approx -fO$$

where  $f$  is the refresh rate, or the rate gases ( $CO_2$  and  $O_2$ ) are both forced from the reaction chamber and replaced with the  $CO_2$  released from the yeast. The students assume that the rate oxygen is used by the yeast,  $\varepsilon = \delta \frac{r_o - r_a}{O_0}$  to be much smaller than the rate  $O_2$  is forced from the chamber. They also assume that the rate of gas production is constant, which is not consistent with the obviously changing rate of  $CO_2$  production. Putting that aside and assuming  $\varepsilon$  negligible,

$$\dot{O} = -fO \quad \Rightarrow \quad O = O_0 e^{-ft}. \quad (30)$$

They proposed a direct relationship between  $CO_2$  production and yeast growth,

$$\dot{C} = b\dot{Y},$$

where  $b$  is the rate  $CO_2$  is produced from the yeast and sugar interaction. Integrating leads to

$$C = b(Y - Y_0). \quad (31)$$

This volume,  $C$ , must match the volume of  $CO_2$  observed;

$$C = Ah(t). \quad (32)$$

Here,  $A$  is the cross-sectional area of the jar and  $h(t)$  is the height of the air column in the jar.

Thus,

$$h = \frac{b}{A}(Y - Y_0). \quad (33)$$

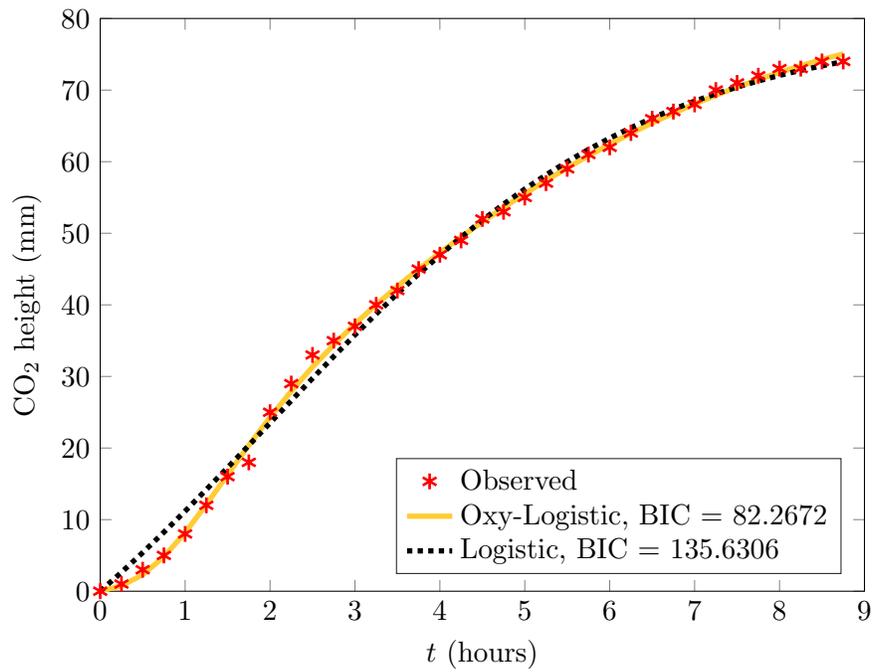


Figure 8: Predicted heights from the Oxy-logistic Model compared with measured CO<sub>2</sub> heights (\*) produced by yeast growing on sugar in a flask and the initial logistic model derived in class. Parameters  $\lambda_a = 0.3860$ ,  $\lambda_o = 5.3592$  and  $f = 1.1662$ , as well as the combinations  $\frac{bY_0}{A} = 0.3406$  and  $\frac{Y_0}{K} = 84.3250$  were found using minimum sum-squared error. The model performs well compared to the initial logistic model and according to the odds ratio (20) is about 3200 times 'better.'

To solve for  $Y$  the group used  $\dot{Y} + \dot{S} = 0$  to remove  $S = K - Y$  from the equations, and substituting (29) and (30) led to

$$\begin{aligned}\dot{Y} &= r(O)Y(K - Y) = K \left( r_a + (r_o - r_a)e^{-ft} \right) Y \left( 1 - \frac{Y}{K} \right), \\ &= \left( \lambda_a + (\lambda_o - \lambda_a)e^{-ft} \right) Y \left( 1 - \frac{Y}{K} \right),\end{aligned}\tag{34}$$

where  $\lambda_a = Kr_a$  and  $\lambda_o = Kr_o$ . Separating variables and solving for  $Y$  finally gave

$$Y = \frac{KY_0}{Y_0 + (K - Y_0) \exp \left[ -\lambda_a t - \frac{\lambda_o - \lambda_a}{f} (1 - e^{-ft}) \right]}.\tag{35}$$

Using (33) they arrived at their final model,

$$h = \frac{bY_0}{A} \left[ \frac{1}{\frac{Y_0}{K} + \left( 1 - \frac{Y_0}{K} \right) \exp \left[ -\lambda_a t + \frac{\lambda_a - \lambda_o}{f} (1 - e^{-ft}) \right]} - 1 \right]\tag{36}$$

This equation has five identifiable parameters,  $\lambda_a = 0.3860$ ,  $\lambda_o = 5.3592$  and  $f = 1.1662$ , as well as the combinations  $\frac{bY_0}{A} = 0.3406$  and  $\frac{Y_0}{K} = 84.3250$ , which the group fit using minimum sum-squared error. As seen in Figure (8), the model outperformed the logistic model (BIC  $\approx 82.27$  for the Oxy-logistic model compared to BIC  $\approx 135.63$  for the logistic model) and was the best of the student models based on BIC.

### 3.3 Winner, Winner

For the Waste, Linear (Un)death, Pyruvate and Oxy-logistic models the BIC's were 123.58, 169.05 and 98.40 and 82.27, respectively. For many of the students the difference between the Pyruvate and Oxy-logistic models seemed minimal. While they both performed much better than the logistic model, the Oxy-logistic model performed

$$e^{\frac{1}{2}(\text{BIC}_{\text{Pyruvate}} - \text{BIC}_{\text{Oxy-log}})} \approx 3200$$

times 'better' based on the odds ratio (20).

The models that performed the best focused on cellular respiration and how the yeast behave depending on the availability of oxygen. As data and each model indicates, yeast rely on sugar availability to grow, reproduce and generate CO<sub>2</sub>. If the experiment were run over a

longer duration perhaps the waste products from yeast growth would impact  $\text{CO}_2$  production more dramatically. However, the current group of models indicates the waste products have less influence over the initial eight or nine hours of  $\text{CO}_2$  production. Rather, as described in both the Pyruvate and Oxy-logistic models, it appears to be the difference between anaerobic and aerobic respiration coupled with sugar availability that initially drives yeast population growth dynamics and in turn, their generation of  $\text{CO}_2$ .

## 4 Discussion and Conclusion

Using readily-available ingredients, the Yeast Lab is a population dynamics LEMB that challenges students and promotes traditional mathematical biology skills (e.g., numerical and analytic solution of differential equations, population modeling, parameter estimation, dimensional analysis, use of computational tools) as well as difficult-to-teach professional soft skills (e.g., leadership, teamwork, communication, the ability to think creatively and critically about models and data). By framing the modeling exercise as a competition mediated by BIC, students were encouraged to take healthy mathematical risks and were more inclined to view their teacher as a mentor and collaborator helping them to learn necessary skills. Throughout the Yeast Lab our students were gripped by a spirit of exploration and competition, and spent many more hours developing their approaches and presentations than would ever have been spent on homework. In exit evaluations students frequently requested *more* experiences of this nature because of the creativity and challenge involved. Students also commented on the lab gave them an exciting insight into the research life of mathematical biologists.

From the models produced by the class, students were able to generate authentic results about a classic mathematical biology species. In so doing, students were able to gain a better understanding of the nature of biomath research. Students often feel that mathematicians are solitary geniuses who prefers to be left alone to work with paper and pencil in the wee hours of the night. Moreover, students feel that, while genius mathematicians may work on interesting, real-life mathematics, they themselves are only capable of solving contrived, algorithmically-solved problems.

Laboratory experiences like the Yeast Lab point out that mathematicians (especially interdisciplinary mathematicians) work on multi-faceted problems where learning from the approaches of others and collaborating is a joy and necessity. In the Yeast Lab students practiced applying the mathematics that they will need as professionals in the context often seen in the profession; they worked as a team.

By fostering students' problem-solving attitudes, the Yeast Lab contributed to a more balanced picture of mathematics and resulted in a handful of creative models. Initially, students expressed doubt that they would be able to produce a model that outperforms the logistic model presented in class. They felt logistic model accounted for all the major phenomena in a logical manner. However, working as a team they altered the assumptions made in the logistic model and added some of their own hypotheses based on their background knowledge and their observations of the data. In the end they were successful in creating multiple mechanistic models for the yeast/ $\text{CO}_2$  relationship that outperformed the logistic model.

For students creating good mathematical models is an important conclusion to the lab, but for the teacher the students' interactions with data, mathematics and their classmates is probably more important than the actual models. Students specifically mentioned that dealing with the data first hand gave them a base from which they could begin constructing a model. They blended their complementary strengths and built on the talents of their teammates. Students with a greater biology background had a tendency to lead the discussion at the beginning of model formulation while those from mathematics lead the group through the fine tuning of the model. From that base students were able to engage in creative and daring mathematics. Rather than focusing on whether each model they proposed in their groups was the correct model, students would voice the connections they saw in the setup and in the data. They then translated those ideas into mathematical relationships, compared their predictions with the data, revised and repeated the cycle. Students did struggle with core applied issues like dimensional analysis, the identifiability of parameters, inconsistencies between models and data, and the uncertainty of nonlinear parameter estimation. But these are struggles (on the one hand) that are critical to do useful, professional

applied mathematics, and (on the other hand) seldom faced by students in classroom mathematics context. During the Yeast Lab our students used mathematics and engaged in the modeling process just as professionals would, and in doing so gained an authentic picture of applied math practice.

Wrapping the Yeast Lab around the paradigm of model competition was a big contributor to the lab's success. With BIC (as opposed to instructor) serving as the judge, students no longer worried about the *correct* model, nor did they worry about the instructor's judgement. That is incredibly liberating for mathematics students used to situations in which solutions and techniques are either strictly right or wrong. The creativity of the models our student produced and the amount of time they spent learning and working attests to the educational impact of LEMBs. The freedom to collaborate/mentor students without judgement is similarly liberating for the instructor. After launching the Yeast Lab we mainly walked around, listened in on group work, answered questions, pressed students for justifications and explanations, encouraged them to write initial models and then think about the consequences of their modeling choices. There were many moments of mathematical facilitation, leading students through techniques necessary to make their ideas practical, and in these moments both students and instructor were fully engaged in the learning process. There were no questions about whether the math being discussed was useful – it was self-evidently useful because students needed it to make use of their creative models and face the challenge of the data. There was no boredom with the explanations, because the students and instructor were both engaged in authentic problem solving.

While modeling and model competition can be done with a canned data set, the students ability to draw from the unwritten/difficult to quantify observations can only occur if the student has a part in collecting the data. With the setup described above, the Yeast Lab has been used in high school, undergraduate and graduate courses as a means to promote and develop various mathematical competencies from an authentic perspective. Giving students a full modeling experience that includes data collection, model creation and model competition does not require difficult to acquire/use technical gadgetry. It *does* require a willingness to push aside some of the normally

tight curriculum in math classes, to give up some of the control over time and content that traditional lecture-format teaching offers. The payoff is that students gain an opportunity to learn soft and creative skills that are more valuable in professional life.

## References

- [1] Akaike, H., 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6), pp.716-723.
- [2] Akaike, H., 1981. Likelihood of a model and information criteria. *Journal of econometrics*, 16(1), pp.3-14.
- [3] Bruder, A. and B.R. Kohler, 2016. Coffee Milk Lab. *Math Journal*.
- [4] Cavanaugh, J.E. and Neath, A.A., 1999. Generalizing the derivation of the Schwarz information criterion. *Communications in Statistics-Theory and Methods*, 28(1), pp.49-66.
- [5] Clayton, M., 1998. Industrial applied mathematics is changing as technology advances: What skills does mathematics education need to provide. *Rethinking the mathematics curriculum*, pp.22-28.
- [6] Feldmann, H. (ed), 2012. Introduction, in yeast: Molecular and cell biology, Second Edition, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi: 10.1002/9783527659180.ch1
- [7] Feser, J., Vasaly, H. and Herrera, J., 2013. On the edge of mathematics and biology integration: Improving quantitative skills in undergraduate biology education. *CBE-Life Sciences Education*, 12(2), pp.124-128.
- [8] Fisher, R.A., 1925. Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge University Press 22:05.
- [9] Gause, G. F., 1934. *The struggle for existence*. Williams and Wilkins, Baltimore.
- [10] Haughton, D.M., 1988. On the choice of a model to fit data from an exponential family. *The Annals of Statistics*, 16(1), pp.342-355.
- [11] Hilborn, R. and Mangel, M., 1997. *The ecological detective: confronting models with data* (Vol. 28). Princeton University Press.

- [12] Kashyap, R.L., 1982. Optimal choice of AR and MA parts in autoregressive moving average models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (2), pp.99-104.
- [13] Knowles, J. R., 1980. Enzyme-catalyzed phosphoryl transfer reactions. *Annual review of biochemistry*, 49:877-919.
- [14] Kohler, B. R., Swank, R. J., Haefner, J. W., and Powell, J. A., 2010. Leading students to investigate diffusion as a model of brine shrimp movement. *Bulletin of Mathematical Biology*, 72(1): 230-257.
- [15] Labaree, D.F., 1997. Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal*, 34 (1):39-81.
- [16] Leonard, T., 1982. Comments on A simple predictive density function, by LeJeune , M. and G. D. Faulkenberry. *J Am Stat Assoc*, 77:657658.
- [17] Lewis, M.J. and J.A. Powell, 2016. Modeling zombie outbreaks: A problem-based approach to improving mathematics one brain at a time. *Primus*, accepted
- [18] Parnas, D.L., 1996. Education for computing professionals. Chapter, 3, pp.31-42.
- [19] Powell, J. A., B.R. Kohler, J.W. Haefner and J. Bodily, 2012. Carrying biomath education in a leaky bucket. *Bulletin of Mathematical Biology* 74: 2232-2264.
- [20] Powell, J. A., Cangelosi, J. S., and A. M. Harris, 1998. Games to teach mathematical modelling. *SIAM review*, 40(1), 87-95.
- [21] Resnick, L. B., 1987. The 1987 presidential address: Learning in school and out. *Educational Researcher* 13-54.
- [22] Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics* 6(2): 461-464.
- [23] Stone, M., 1979. Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.276-278.

- [24] Tatto, M.T., Peck, R., Schwille, J., Bankov, K., Senk, S.L., Rodriguez, M., Ingvarson, L., Reckase, M. and Rowley, G., 2012. Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics in 17 Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-MM). International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.
- [25] Transforming Undergraduate Education for Future Research Biologists, 2003. Board on Life Sciences, Division on Earth and Life Studies, National Research Council of the National Academies. National Academies Press, Washington D.C.
- [26] Usiskin, Z., 1991. Building mathematics curricula with applications and modeling. In: Niss/Blum/Huntley (Eds.) [1991]: *Mathematical Modeling and Applications*; Chichester: Horwood, pp. 30 - 45.
- [27] Verhulst, Pierre-François, 1838. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance Mathématique et Physique* 115-116.