Introduction to Modeling

Bioengineering includes aspects of both science and engineering. For more scientific questions, quantitative models can be used to test a hypothesis that requires a quantitative answer. For engineering projects, quantitative models can also be used to design or optimize something that is being built. Sometimes there is a fine line between the two, such as when we design an intervention into a biological system, such as drugs that affect insulin or thrombotic dynamics, or an orthopedic implant that must mechanically interact with the body.

Mechanistic Models. In some cases, we need to understand the mechanism of a process to answer our question, because we are testing a hypothesis about the mechanism or are trying to change system behavior by altering one or more components. In this case, it is vital that the model correctly reflect the *actual* underlying mechanism, although it is never necessary or even advised to model the *entire* underlying mechanism. That is, the mechanism must be correct in the details which affect the question being asked. We call such a mechanistic model a gray box model, because the mechanism is only partially blacked out, thus still partially apparent in the model. We also call this a parametric model because the parameters of the model have physical meaning that could be independently measured. For example, one might measure the size of a component, the reaction kinetics of two chemicals, a diffusion constant, or the stiffness of an elastic element.

Emperical Models. In other cases, it is not necessary to understand the mechanism of a system because we will not need to change the system. Instead, we may want to change how the system contributes to a larger system. Again, this could involve scientific questions about the role of a protein or cell in a larger physiological system. It can also involve engineering questions such as how to integrate an op-amp into a circuit, or a protein into a genetic network, to get desired behavior of that system. An alternative use of empirical models is to quantitatively compare the behavior of different related systems. An empirical model can allow a few parameters to concisely describe the difference between two data sets. The word empirical refers to knowledge that is based in observation rather than theory, so an empirical model quantitatively describes observed behavior, which is sufficient for these kinds of problems. These are also called black box models because the underlying mechanism is not visible from the model equations. The advantage of empirical models is that there is no requirement that the mechanism be known, and it is often easier to chose a model with fewer parameters.

In reality, most models have elements of both parametric and empirical models. The behavior of an enzyme may be modeled with the Michaelis-Menton Equation, which is a black box model about the enzyme structure, and could never be used to predict how point mutations would affect reaction kinetics. On the other hand, the same enyzyme may be part of a mechanistic model of a genetic network. A model of HIV dynamics may model a T-cell as a black box, but use a gray box model of the interactions between cell and virus; for example, the rate of infection of a naïve cell could be measured independently and be a parameter in the model.

Example 1: Hair cells. We know the viscoelastic response of a hair cell cilia in the inner ear in response to a movement of the tympanic membrane. We want to predict how it will respond to different types of inputs, including different frequency of vibrations. In this case, a black box model of the viscoelastic response is sufficient since we are asking its response to different inputs.

Example 2: Drug delivery. We make a drug delivery particle that will slowly release a drug over time, and we characterize the release profile in vitro experimentally. Now we want to know how this release profile will affect drug levels in a patient, since we want to keep the drug in the therapeutic window, which is above the therapeutic threshold and below the toxic. To do this, we can fit an empirical model to the release profile, and use this within a model of the pharmacokinetics of the drug in the body, which addresses the diffusion of the drug in the bodily compartments and the clearance of the drug. We find that the release profile is not good enough for this purpose, so we want to optimize the release kinetics. For this, the empirical model is not sufficient. Instead, we need a parametric model that describes the mechanism of controlled release correctly.

Verification and Validation. For a model to be useful, we need evidence that *the model is correct*. This is true with all methods; we need to believe that the results are not artifacts of the method used, or even caused by a mistake in methodology, but instead reflect the real system of interest. In experiments, we normally incorporate controls to show that an assay is working properly, and we often do additional experiments to determine how results in a simple system relate to a more complex one. We need to take a similar approach to computational modeling. Verification is the process of ensuring mathematical correctness, which you can think of like a control. Like experiments, you have higher certainty if you get the same answer in two independent calculations (e.g. have two people program the same assumptions into a model), or if you can run control simulations where you know what to expect because the system is greatly simplified. Validation is the process of testing the model's ability to capture the real system, which effectively means testing whether the assumptions in the model match experimental or clinical data. The most common approach is showing that the model reproduces some known behavior. Another approach (for mechanistic models) is to determine the parameters and the model structure in independent experiments. A third method is to predict from the model some system behavior, and test this after the fact. In general, if the model can predict something that was not used, intentionally or unintentionally, in the process of building the model, this is considered a relatively strong type of validation, while reproduction of existing data (including measurement of parameters) that was used to build the model is considered weak.

Innovation/Significance/Impact. However, for a model to be useful, we have to be able to *learn something new from the model*. The model might allow us to confirm or test an unknown hypothesis. Or, it might be used to help design something. However, if it simply reproduces observed behavior, without a plan for how it might then be interpreted or used in future studies, it is of limited use. The usefulness often comes from predicting how the system will respond to different parameters, or identifying the limitations on parameters necessary to provide observed or required behavior. For this reason, we usually perform a systems analysis that addresses quantitative questions like these.

In summary, if a model is too close to what is already known, it has no innovation, but if it is too far from what is known, it has no certainty. The challenge in modeling is to design a model that can be validated but can still provide critical new information. When the project is well chosen, modeling can be a powerful assistance to engineering research. We will provide examples throughout this class, but the project will be your chance to plan a good use of modeling.

General Approach to Building Models

Before we start building mathematical models, recall our terms about variables and parameters:

Independent variables are the variables that have predictable values. Typically, ODE systems have time as the independent variable, while PDE systems have time and one (e.g. x) or more (e.g. x,y,z) position variables. However, this is not required, and some systems have other independent variables. The values of the independent variables do not determine the system state, but the goal is to define the system state for all relevant values of the independent variables.

Dependent variables are the variables that change dynamically in a system as a function of the independent variables. The value of the all the dependent variables collectively describes the state of the system.

Parameters are values that are independently determined, so they do not depend on the dynamics of the system. They are often constant, but they do not need to be, since they may change as a function of the independent variable. Unlike dependent variables, parameters cannot depend on other dependent variables; that is, they may have a predetermined value (fluid velocity may change with position, or an external concentration may be switched at some time), but they do no change dynamically as a result of the system state.

Ordinary differential equations (ODEs) have one independent variable (usually time) and one or more dependent variables.

Partial differential equations (PDEs) have multiple independent variables (usually time and position), and one or more dependent variables. We will focus on modeling chemical transport with convection, diffusion, and reaction terms, so we will consider C1(x,y,z,t).

Stochastic differential equations (SDEs) are ordinary differential equations in which the dependent variables are stochastic variables, because the equations have a stochastic term.*General Model Building Techniques.* For this course, we will learn to build mathematical models and solve them numerically for these three types of differential equation models. Computational modeling involves the following "DIESE" steps for model building.

1) **D**iagram the model.
2) **I**dentify the parameters, dependent variables, and independent variables. Translate the known values and question to be answered in terms of these parameters and variables.
3) **E**quations. Write the equations that translate the diagram and any other assumptions into mathematical form.
4) **S**implify. (combine equations to remove unnecessary variables)
5) **E**rror Check. (Verification)

Once the model is built, you should have the differential equations and initial/boundary conditions. We will learn to…

1) solve these numerically and verify that there are no numerical artifacts in the solution.
2) when possible, solve the model behavior analytically (at least in part) and verify that the numerical and analytic solutions agree.
3) Perform additional analysis to answer your questions and interpret the key results of the in words that are understandable to a non-modeler interested in the original problem.

Verification at the Model Building Stage

As you build your model into mathematical equations, you should verify the equations, which means check them for mathematical correctness, and check that they match the assumptions. Actually, the main tool for verification is to perform a series of tests for mathematical incorrectness. If you find any errors, you can correct them before going further, and otherwise, you can proceed with somewhat more confidence.

A conceptually simple but time consuming verification test is performing the same calculation two independent times, and making sure you get the same thing. If you don't, you know at least one of your calculations was incorrect. However, your two calculations may not be independent, since you will often repeat the same mistake twice. In a class setting, you can compare your answer to that of one or more of your peers, but that won't be possible when you have a problem to solve outside of a class setting. I thus will teach you verification tools that will help you catch your own mistakes as you go. ***We expect you to perform these verifications whether or not we explicitly tell you to, so we will detract more points for any mistakes that should have been caught with our verification tools, than for other algebraic mistakes***. While you can shortcut this in class by comparing your calculations with your peers, we do not encourage this because this won't help you when you solve a problem outside a class setting. Instead, after building a model, we expect you to perform three verification tests:

Model Completeness: The minimal verification you should do is to make sure that you have one differential equation and one initial condition for each dependent variable. For partial differential equations, you will also need a boundary condition equation for each dependent variable on each boundary.

Model Appropriateness. Your model should address the question and assumptions in the model description (verbal or diagram) by relating the variables of interest to each other and to the parameters described, but not to any parameters or variables that you defined during your derivation. In some exceptions, you will keep a parameter or variable you defined in order to present a simpler or more understandable form of the model, but in this case, you must include their definitions as part of the final model.

Dimensional Analysis: Dimensional analysis is a simple verification that should be done when the equations are written and/or simplified, but before trying to solve them. This can be done at each step, or just upon obtaining the final differential equations, after considering volume issues and any simplifications. Dimensional analysis is simply determining the dimensions of each term in each equation, and making sure that they are the same within an equation. Recall that a term is anything that is added or subtracted, on either side of the equation.

Example: A Simple Chemical Reaction Model. ODE, SDE, or PDE?

We will consider here one simple system and consider the different assumptions we might make that require us to use different modeling types and/or different equations. The system is a Surface Plasmon Resonance (SPR) experiment, which is a method that measures the change in mass of material that is very close to a gold surface, which can be converted to the amount of bound material. We can use the SPR to measure chemical reactions in unlabeled sample if we immobilize a ligand (L) on the gold surface, and flow through a receptor (R) in solution, because the receptor-ligand complex (C) has more mass than the ligand L. Thus, response is proportional to the amount of complex formed.

Problem statement: We immobilize a concentration L_0 (in M) of ligand in the chip. We flow a concentration R_0 (in M) of receptor through the chip. How much complex, C, forms as a function of time, if the reaction follows simple one-state kinetics, with association rate k_{on}, in units $M^{-1}s^{-1}$ and dissociation rate k_{off}, in units s^{-1}?

Model Building Comments:

Note that we have one dependent variable, C, and four parameters, R_0, L_0, k_{on}, and k_{off}. Building the model will be easiest if we introduce two more dependent variables, R, and L, which were not provided in the problem statement, to represent the concentration of free receptor and ligand over time. These are not necessarily the same as R_0 and L_0. However, we want the answer to be an equation in C in terms of the four parameters and of the independent variables (time and possibly space), so we may want to remove these additional dependent variables later.

Now we clarify what the association and dissociation parameters mean:

reaction A: $R + L \rightarrow C$, with rate constant k_{on}, in units $M^{-1}s^{-1}$.

reaction B: $C \rightarrow R + L$, with rate constant k_{off}, in units s^{-1}.

<u>*Converting chemical reactions into equations*</u>. The substrates are on the left of the reaction, and the products on the right. So reaction A has substrates R and L, and products C, etc. Recall that that reactions occur at a rate that equals the rate constant times the concentrations of all substrates. That is, reaction A occurs at a rate $k_{on}RL$, where R and L are the concentrations of receptor and ligand, respectively, while reaction B occurs at rate $k_{off}C$, where C is the concentration of complex. The effect of each reaction on the reactants is determined by the difference between the number of that reactant in the product minus the substrates. Thus, each reaction A removes one each of R and L, and adds one C. Thus reaction A contributes - $k_{on}RL$ to $\frac{dR}{dt}$ and $\frac{dL}{dt}$ and + $k_{on}RL$ to $\frac{dC}{dt}$.

That's a good reminder. Now, how do we use this information to build a model that will answer the question? This depends on some additional information or assumptions we need to make.

<u>*PDE models.*</u> If the receptor binds to immobilized ligand faster than it is replenished by convective flow in the device and by diffusion, then the concentration of receptor near the surface will be depleted, which will affect the reaction rate. In this case, we need to model the transport (convection and diffusion) as well as the reaction. This requires a model that describes the concentration of receptor and ligand as a function of position as well as of time, which means a partial differential equation (PDE). We will consider such models in weeks 4 and 5.

SDE models. If you are using a new technology that detects single particle interactions, you may need to model each molecule individually to compare the results to data. If you need to model the movement of single molecules through space, then you will use Brownian Dynamics simulations, which model the stochastic position (x(t), y(t), z(t)) and state of an individual particle. Brownian dynamics provide a discrete stochastic version of transport/reaction PDE equations, and will converge on the PDE solution with large enough numbers, at least for linear systems. We will learn this approach in week 6.

If your new technology does not necessarily detect single molecules, but has nanoliter volumes that involve only 100 or so molecules, then the intrinsic noise due to the stochastic rates of reaction may have a greater effect on your experimental data than will the noise introduced by your measurement methods. In this case, you will want to use stochastic reaction equations, which we will learn in week 7. Both types of stochastic simulations are forms of stochastic differential equations (SDEs).

ODE models. On the other hand, if we have large enough numbers of molecules to neglect intrinsic stochastic noise, and transport is sufficiently fast relative to reaction rates to neglect spatial variations in concentration, then we can use an ODE model. These are much faster to build and to solve than PDE or SDE models, so should be used if these assumptions can be justified. If you have no idea, then solve the ODE model, and do a sanity check with the solution against these assumptions, or compare it to data to see if it fits without the complications of the PDE or SDE model.

Conservation of mass. To write the ODE model equations from the chemical reaction equations, we use conservation of mass, which means that the amount of chemical stays the same except for what is removed or added by any reaction. Thus, we write an equation that expresses that the rate of change for each chemical is equal to the sum of all changes in this chemical due to the various reactions. Since mass in conserved rather than volume, we should convert concentrations to a mass, so that all terms in the differential equation have units mass (or number) per second. To do this, we multiply by the volume appropriate to the reactants in each term in the equation. However, in our SPR model, the immobilized ligand complex, and the free receptor are all present in the same volume, which we will call v, so there will be a 'v' in every term of the ODE, and these will all fall out, so we can disregard it, and simply write the equations in terms of concentrations. In other words, *if (and only if) the volume is the same for all variables,* conservation of mass becomes conservation of concentrations.

$$\frac{dR}{dt} = k_{off}C - k_{on}RL \qquad \text{Equation 1}$$

$$\frac{dL}{dt} = k_{off}C - k_{on}RL \qquad \text{Equation 2}$$

$$\frac{dC}{dt} = -k_{off}C + k_{on}RL \qquad \text{Equation 3}$$

We don't need any boundary conditions, but we need initial conditions, one for each variable. We use the assumptions to match this.

$$R(0) = R_0$$

$$L(0) = L_0$$

$$C(0) = 0$$

Thus, you start with one reaction equation for each reaction, and divide that up differently into one differential equation for each chemical. In general, the number of reaction equations and differential equations are not the same.

Next we look to verify the model:

1. The model is complete, since we have one ODE and one IC for each dependent variable.
2. Woops! the model is not appropriate to the problem; we indeed used only the four parameters, but we still have the two variables that we defined, and that were not defined in the problem statement. We could leave these in the problem, and clarify their definition, or can try to simplify them away.
3. The model passes a dimensional analysis. All equations are similar: $\frac{dX}{dt}$ is M/s. koff is s^{-1} and C is M, so $k_{off}C$ is also M/s. kon is $M^{-1}s^{-1}$, so $k_{on}RL$ is also M/s. Thus, I'll continue with this model to remove the two unnecessary variables.

We notice that the equations are the same, except for the switch in signs between the complex vs reactants. This is not generally the case, and occurs here only because we started with two completely symmetrical reactions. So, we are confident that we can remove some variables. To do this, you must realize that the total amount of ligand in free and complex form must stay the same, because it is immobilized. Thus, $L + C = L_0$, or $L = C - L_0$, which will remove L. In addition, you must remember the assumption we made when we rejected the PDE model: we assumed that the receptor is being replenished by transport faster than it is reacting, so it remains at the inflow concentration, R_0. Thus, $R = R_0$. Indeed, our equation 1 above was incorrect for this problem, since we did not include the transport when we considered things that affect conservation of mass of R. This was not an issue with the conservation of mass equations for L and C, since they are both immobilized and cannot enter or leave the system except by reaction. Thus, we use these two equations to remove R and L from the ODE for C above. This leaves us with the ODE:

$$\frac{dC}{dt} = -k_{off}C + k_{on}R_0(C - L_0)$$

with IC:

$$C(0) = 0$$

Thus we still have one ODE and IC per variable, and the dimensional analysis is the same, but now we have a model appropriate to the question. In fact, this catch was huge, since our previous version of the model forgot to use the assumption about transport and inflow, and instead was assuming a closed system.

How else might we have caught this mistake (or not made it in the first place?) We need to remember to ask whether material can leave or enter the system when writing our conservation of mass equations. Indeed, this would cause us to ask immediately about how much material is flowing in and flowing out, and thus about the flow rate, and about the spatial variation in concentration, which would lead us directly to asking whether we need a PDE model or can add an assumption to allow the ODE. Once we made the R = R0 assumption for the ODE in this context, we would have realized that we did not need to do conservation of mass or write an ODE for R at all.