Modeling for Non-Modelers

A Gentle Introduction to Mathematical Models

What is a Model?

A model, at its most basic, takes one or more inputs, and processes those inputs into one or more outputs.

Another (simple) model

What is the impact of elevated CO2 levels on air temperature?

Also a model

Figure 5.6: Generalized diagram of components of a wetland mass balance.
Causal Loop Diagram of a Model Examining the Growth or Decline of a Life Insurance Company

Source: Robert A. Taylor, U.S. Department of Energy

C = negative (counteracting) feedback loop, || = delays

Things to do with models

- Generate new, testable hypotheses about a system.
- Explore emergent properties of a system.
- Understand our ignorance.
- Predict the future.
- Perform virtual experiments.
- Optimize decisionmaking about the system.

And some downsides…

- They can be difficult to understand. This complexity can pose significant barriers to understanding by end-users.
- Models contain many “hidden” assumptions. Know what these assumptions are before using the model!!!
- The results of model runs may be believed to easily. Be a skeptic!
- Models, particularly complex models, can be expensive and time consuming to develop, implement and validate.

A Hierarchy of Approaches

The Process of Modeling
Is your model any good?

Verify → Calibrate → Validate

- You need data – as much as possible
- Partition your data into two independent sets – one for calibration, one for validation

Tools

- Excel – great for simpler models, regression – a bunch of built-in tools for simple data analysis, statistical modeling
- Visual Tools – Stella, Simile, PowerSim, GoldSim…
- MatLab – OSU site license, simple programming, rich math modeling support
- R, Python – simple, interpreted open source (free) programming languages, rich modeling support
- Traditional Programming Languages – C++, FORTRAN – powerful, computationally efficient, but more hassle to work with – often used for large, complex models.

A Tale of Four Models

Next, we will (briefly) explore four models and modeling approaches:

1. A simple regression model (in Excel, Python)
2. A classification model (Logistic Regression) in Python
3. A population model in Python
4. A systems dynamics model using Python

Model 1 – Empirical Model

Goal: Identify relationship between two variables based on data
This is an example of an ‘empirical’ (or "statistical") model:
  1) data driven (not physically-based)
  2) assume a model form (e.g. \(Y=aX+b\))
  3) use regression to find best fit parameters

To access the Jupyter Notebooks

- [https://bee-jupyter.bee.oregonstate.edu](https://bee-jupyter.bee.oregonstate.edu)
- User: jupyter2
- Password: robot

Linear Regression - Excel

Dataset available at: [http://explorer.oregonstate.edu/Topic/Model/Data/StreamTemp.xlsx](http://explorer.oregonstate.edu/Topic/Model/Data/StreamTemp.xlsx)

Approach:

1. Load Data into Excel
2. Plot the data of interest
3. From the plot, add regression lines
4. Generate summary statistics
Model 2 - Classification

- Classification models categorize sets of inputs in two or more discrete bins.

A common example: predicting species presence/absence based on environmental factors.

We want to understand the probability that a given combination of factors will result in a “positive” result.

Logistic Regression - Python

- Jupyter Notebook available at: https://bee-jupyter.bee.oregonstate.edu/user/jupyter2/notebooks/Classification%20and%20Logistic%20Regression.ipynb

- Approach:
  1. Load Data into Python
  2. Take advantage of pre-existing package (sklearn) for doing all the work

Model 3 – Population Dynamics

Problem: Predator-Prey Interactions - The Lotka-Volterra Equations

The Lotka-Volterra equations were developed in the early part of the 20th century as a simple description of the dynamics of two populations, one predating on the other.

An example might be arctic foxes and arctic hares.

The Math

Problem: Predator-Prey Interactions - The Lotka-Volterra Equations

This is arguably the simplest possible model of predation. Translating to a mathematical description results in the following:

- Rate of Change of Prey \( H \):
  \[
  \frac{dH}{dt} = r_H H - a_H P
  \]

- Rate of Change of Predator \( P \):
  \[
  \frac{dP}{dt} = -r_P P + b_H P
  \]

Questions:

1. Under what circumstances might this model be reasonable?
2. How do we solve the resulting system of equations?
3. Under what conditions (if any) are the populations in equilibrium?
4. Do the equilibrium point imply stability?

Model 4 – Zombie Attacks!

Our final model looks at three interacting populations – Zombies, Humans, and the Dead.

Assumptions (pretty simple!):

- Humans kill Zombies
- Zombies resurrect from the dead and infect Humans

Zombie Math

\[
\frac{dH}{dt} = (k_{Birth} H) - (k_{Trans} H \cdot Z) - (k_{Death} H)
\]

\[
\frac{dZ}{dt} = (k_{Trans} H \cdot Z) + (k_{Res} D) - (k_{Rem} H \cdot Z)
\]

\[
\frac{dD}{dt} = (k_{Death} H) + (k_{Rem} H \cdot Z) - (k_{Res} D)
\]
Things to think about

- All models are wrong, some models are useful. Be sure yours is in the latter group.
- Keep it as simple as possible, but not too simple...
- Model iteratively – start simple, add complexity as needed
- Modeling is a great way to generate insights, explore and understand the world...

Uses of Models

Three Primary Uses of Models:
- Analysis
- Prediction
- Control

Additional Secondary Uses include:
- Conceptual framework for organizing or coordinating empirical research
- Mechanism to summarize or synthesize large quantities of data
- Mechanism to identify areas of ignorance
- Gaming - providing insight to managers by performing “what-if” simulations

A few more benefits...

- Results in a description that is precise and unambiguous – good as a means of communication.
- Models enable predictions to be made in such a way that these predictions can be checked against reality by experiment.
- Capturing knowledge about a system in a durable, distributable form.
- Models can be Fun!

Additional considerations...

- Realism: the degree to which model structure mimics the real world - how completely does the model capture reality
- Precision: the accuracy of the model predictions - how well does the model generate results?
- Generality: the ability of the model to adequately represent diverse systems - is the model broadly useful?
- Often, these properties trade off against each other.

An Example

Objective
- Construct a description of the dynamics of the world's population such that the time when the population size will double can be computed:

Assumptions (Hypotheses)
- We will model the population as whole, not individuals, ages, sexes, etc.
- The population grows at a constant per capita rate – not influenced by internal or external factors

Resulting Model

\[
\frac{dN}{dt} = rN \rightarrow N = N_0 e^{rt}
\]

where \( r \) is the intrinsic per capita growth rate

To calibrate the model, we solve for \( r \)
A slight tweak

Mathematical (Re) Formulation
Assume that the per-capita growth rate is limited by a carry capacity

\[
\frac{dN}{dt} = rN \left(1 - \frac{N}{K}\right)
\]

where \( r \) is the intrinsic per capita growth rate and \( K \) is the carrying capacity.

To calibrate the model, we solve for \( r \) and \( K \).

Which model is better?

Deterministic Models
- known inputs and unique outputs.
- no components of the model are random.
- multiple model runs with the same input will produce the same output.

Stochastic Models
- either inputs or parameters in the model are uncertain or variable, and the statistical distribution of these variables is defined.
- multiple model runs result in a distribution of outputs.
- Examples: Precipitation; flow in the Willamette River.

Do you need to capture variability?

Modeling Process Steps

Modeling Basics

Systems Basics

A Simple Example

\[
\frac{dy}{dt} = \text{inflow} - \text{outflow}
\]

Example: Consumption

\[
\frac{dy}{dt} = -\text{outflow}
\]
**Systems Basics (continued)**

Stock of water in a reservoir

![Diagram of water reservoir system with components like precipitation, water inflow, evaporation, and discharge]

There are two pools: 4600.512 and 7949.685

Evaporation is 2.81475

Feedback loops form when changes in a stock affect the flows into or out of that same stock.

**Feedbacks**

Systems of information-feedback control are fundamental to all life and human endeavor, from the slow pace of biological evolution to the launching of the latest space satellite... Everything we do as individuals, as an industry, or as a society is done in the context of an information-feedback system.

“- Jay Forrester

example – Algal Dynamics - Where are the feedbacks?

- What do the dynamics look like?
- What are the key interactions?
- What is controlling the system?

Assumptions:
- There are two pools: Population and Resource.
- Resources are consumed at a constant per-capita rate, and are non-renewable.
- The population is influenced by birth and death: birth is a constant per-capita rate; the size of the resource base influences the death rate of the population according to: \[ \text{Death Rate} = k_2 \frac{\text{Resource}}{\text{Resource Base}} \]

Mathematics of an Unsustainable System

\[ \frac{dA}{dt} = \frac{b - cR}{R_0} P \]

Where:
- \( P \) = Population
- \( R \) = Resource
- \( b \) = per-capita birth rate
- \( c \) = per-capita consumption rate
- \( R_0 \) = initial resource base

\[ \frac{dR}{dt} = -cP \]

Example:

- Growth = \( A \) \( A_0 \) = Max Algal population
- \( k_0 \) = effective shade
- \( k_1 \) = per-capita mortality
- \( k_2 \) = per-capita growth rate
- \( k_3 \) = overall dynamics

\[ \text{Growth} = k_4 \text{Death} \]

\[ \text{Growth} = k_4 \left(1 - \frac{A}{A_0}\right) \]

\[ dW \over dt = \text{Precip + Inflow - Evaporation - Discharge} \]
Population Overshoot and Collapse

Overshoot and Collapse

What are the possible equilibrium (steady-state) states?

- We find these by setting the derivative(s) equal to zero.

\[
\frac{dP}{dt} = \left[b \left(1 - \frac{R}{R_0}\right) \right] P = 0 \quad \text{and} \quad c = 0, \\
\frac{dR}{dt} = -c P = 0.
\]

Where:
- \( P \) = Population
- \( R \) = Resource
- \( b \) = per capita birth rate
- \( c \) = per capita consumption rate
- \( R_0 \) = initial resource base