

## Introduction to Modeling and Simulation

A **model** is an abstraction from reality used to help understand the object or system being modeled. People use modeling all the time to make decisions in their everyday lives although they usually don't do so in a formal way. Here are some common things are models:

- Maps are models of a portion of the earth's surface. They are made from materials that do not match that surface (traditionally made of paper rather than earth and rock), they have a different form that abstracts many objects shown on the map and leaves other out altogether. For example, a street map leaves out information on the topography, representing the earth as a flat plane. It shows only the streets and their names, direction relative to the north pole, and perhaps a scale and a title. They are generally drawn to scale so that one can calculate the approximate distance on the real earth that is represented by the proportional distance on the map. Computer maps are a special type of map that is "virtual" and goes away when the computer is turned off.
- Most computer games are models of real or imaginary worlds programmed in a computer. The game programmers decide how objects in that virtual world behave – do they follow the ordinary rules of physics or are some of those suspended – allowing them the characters to jump extraordinary distances or fly. They also choose which objects to animate and which are left as static backgrounds even though the real objects in a real world are subject to changes over periods of minutes, hours, or days.
- Many toys are models of real objects, scaled down or changed in their operation so that they are not dangerous or messy like toy trucks, guns, swords, dolls, dishes, stoves.

People naturally use their experiences to create mental models of things they encounter in ways to help themselves learn and survive. Most of us know that brown water is probably not healthy to drink. We know that a car in the snow can possibly skid and therefore we should drive slower and leave more space and time to stop. In Columbus, Ohio, the prediction of snow seems to create a model in many people's mind that has them slow down even if it doesn't snow.

### Types of Models

Models can be characterized and classified in a number of ways, but a complete review of those definitions is not really important to our work in this course. However, let's define a few broad categories of models to help us understand how they work. These categories are not necessarily mutually exclusive since a lot of models involve elements from several types.

### Physical Models



Physical models are scale representations of the same physical entities they represent. They are used primarily in engineering of large-scale projects to examine a limited set of behavioral characteristics in the system. The illustration and site below are a good example of a physical model -- one of a stream and dam used to simulate the rate of outflow from the dam under a variety

of circumstances. In the electronic version of this document there is link that explains more about this physical model (<http://mhl.nsw.gov.au/www/warragvt.htmlx>).

### [Model of river and dam](#)

#### **Mathematical Models**

Models that are run on a computer require the translation of a mental model into a set of rules and structures that can be represented in mathematical terms using a programming or modeling language. Mathematical models use mathematical equations to represent the key relationships among system components. The equations can be derived in a number of ways. Many of them come from extensive scientific studies that have formulated a mathematical relationship and then tested it against real data, just like our "driving to work" example. Some come from laboratory testing of relationships where that is feasible. Sometimes real data are used to derive relationships using statistical techniques to fit a particular relationship to the data and to measure the level of error associated with that representation.

Mathematical models of large-scale systems often use a combination of approaches -- inserting tested equations where the relationships are well known and inserting statistical relationships where there is less certainty. Such models can also use probabilistic relationships for events that are random or exhibit some type of variable pattern. For example, models of weather analyze the long-term weather records for the area under consideration and calculate the frequency of different weather incidents. These are represented as their probability of occurrence, assuming that the past is a strong indication of future events. In Ohio, for example, there is a high probability of many small rainstorms because they occur frequently. There is a much lower probability of extreme events that produce a large amount of rainfall in a short time -- high winds, high snowfalls, etc. These can be incorporated into mathematical models by applying probability distributions in the model and using a random number generator to choose events based on that distribution. Similar rate based distributions are used in population dynamics models relating to birth and death rates, in the modeling of disease (epidemiology) based on risk factors, for earthquake risks, and for many kinds of accident modeling.

## Simulation Models

Simulation models are a special subset of mathematical or physical models that allow the user to ask "what if" questions about the system. Changes are made in the physical conditions or their mathematical representation and the model is run many times to "simulate" the impacts of the changes in the conditions. The model results are then compared to gain insight into the behavior of the system.

## Modeling Terminology<sup>1</sup>

As we work with various kinds of models, we will encounter a whole new vocabulary of words that may have special meaning when applied to models. Here are a few of those terms.

**Accuracy** – The closeness of a measured or modeled/computed value to its “true” value. The “true” value is the value it would have if we had perfect information. We will talk later about various ways to measure accuracy.

**Algorithm** – A set of rules for solving some problem. On a computer, an algorithm is a set of rules in computer code that solve a problem.

**Calibration** - The process of adjusting model parameters within physically defensible ranges until the resulting predictions give the best possible fit to the observed data.

**Conceptual Model** - A hypothesis regarding the important factors that govern the behavior of an object or process of interest. This can be an interpretation or working description of the characteristics and dynamics of a physical system.

**Deterministic Model** - A model that provides a single solution for the variables being modeled. Because this type of model does not explicitly simulate the effects of data uncertainty or variability, changes in model outputs are solely due to changes in model components.

**Empirical Model** - An empirical model is one where the structure is determined by the observed statistical relationship among experimental data. These models can be used to develop relationships that are useful for forecasting and describing trends in behavior but they are not necessarily mechanistically relevant that is they don't explain the real causes and mechanisms for the relationships.

**Parameters** - Terms in the model that are fixed during a model run or simulation but can be changed in different runs as a method for conducting sensitivity analysis or to achieve calibration goals.

**Sensitivity** - The degree to which the model outputs are affected by changes in a selected input parameters.

**Statistical Models** - Models obtained by fitting observational data to a mathematical function.

---

<sup>1</sup> Some of these definitions were taken from Glossary of Frequently Used Modeling Terms. <http://www.epa.gov/crem/library/glossary.pdf>.

**Stochastic Model** - A model that includes variability in model parameters. This variability is a function of: 1) changing environmental conditions, 2) spatial and temporal aggregation within the model framework, 3) random variability. The solutions obtained by the model or output is therefore a function of model components and random variability.

**Variable** - A measured or estimated quantity which describes an object or can be observed in a system and which is subject to change.

**Validation** – Answers the questions “Is the science valid and does the model use current methods and techniques? Is the numerical model adequate to convey the science principles at the level of the question being asked? Is the model arriving at an acceptably accurate representation of the phenomenon being modeled?”

**Verification** – Does the code for the model run correctly and provide a mathematically correct answer? Do the **algorithms** being used accurately represent the mathematical function on the computer?

### First Modeling Example

We will start with an example that everyone can understand – trying to model the time it takes to go through traffic from your house to a destination like work. Let's say you need to decide the best route to take to work. To make this decision you will need to formulate at least one objective for your trip. Here are some of the possible objectives.

- Minimize the amount of time it takes to get there
- Avoid traffic congestion (You hate traffic!)
- Find a route that excludes freeways (You have an old clunker of a car)
- Plot a path between your house and work to make sure you travel by the same spot every day (You could think of this as any stop along the way - for coffee, breakfast, to follow construction development on a friend's new house, etc.)

Assuming that we focus just on the first objective on the above list, we need to decide what will affect that objective. To do that, we must create a **conceptual model** of the system. Our conceptual model should list all of the **variables** that impact our travel time from home to work and what we believe are the **cause and effect** relationships across all of those variables. For any model we are creating or studying, our ideas on the variables and cause and effect relationships come from published information, analysis of data from a real system, and our own knowledge of the system. The phenomena we are modeling may also be constrained by physical laws or prevailing theories of their operation so our conceptual model should reflect those limitations.

For our traffic example, we know that we need to traverse the street system to get from one place to another and that we need to observe traffic laws governing speed, one-way streets, and traffic control devices. Even so, we have a finite but potentially large number of routes we can take and a large number of conditions that could impact how long the trip takes.



The first part of the class exercise is to define as many of the conditions as possible that will impact the travel time to work along with the cause and effect relationships among the variables. For each cause and effect relationship, we need to estimate the direction of

the relationship, the form of the relationship, and, if possible, a quantitative representation of that relationship. For example, we know that bad weather will slow traffic down. The worse the weather, the slower the traffic. We can hypothesize that the impact on traffic is non-linear, that is as the weather gets worse and worse, the traffic slows more and more. We may not have the data to exactly quantify the relationship but we could start with a simple classification of weather events and an estimate of their impacts on the flow of traffic. Here are some examples of weather conditions - can we fill in an estimate of the impacts on traffic flow?

| Weather Condition                       | Impact on Traffic Flow |
|-----------------------------------------|------------------------|
| Clear skies, no wind                    |                        |
| Light rain                              |                        |
| Heavy rain                              |                        |
| Light snow                              |                        |
| Heavy snow                              |                        |
| Heavy snow and blowing, poor visibility |                        |
| Freezing rain                           |                        |
| Other?                                  |                        |
|                                         |                        |

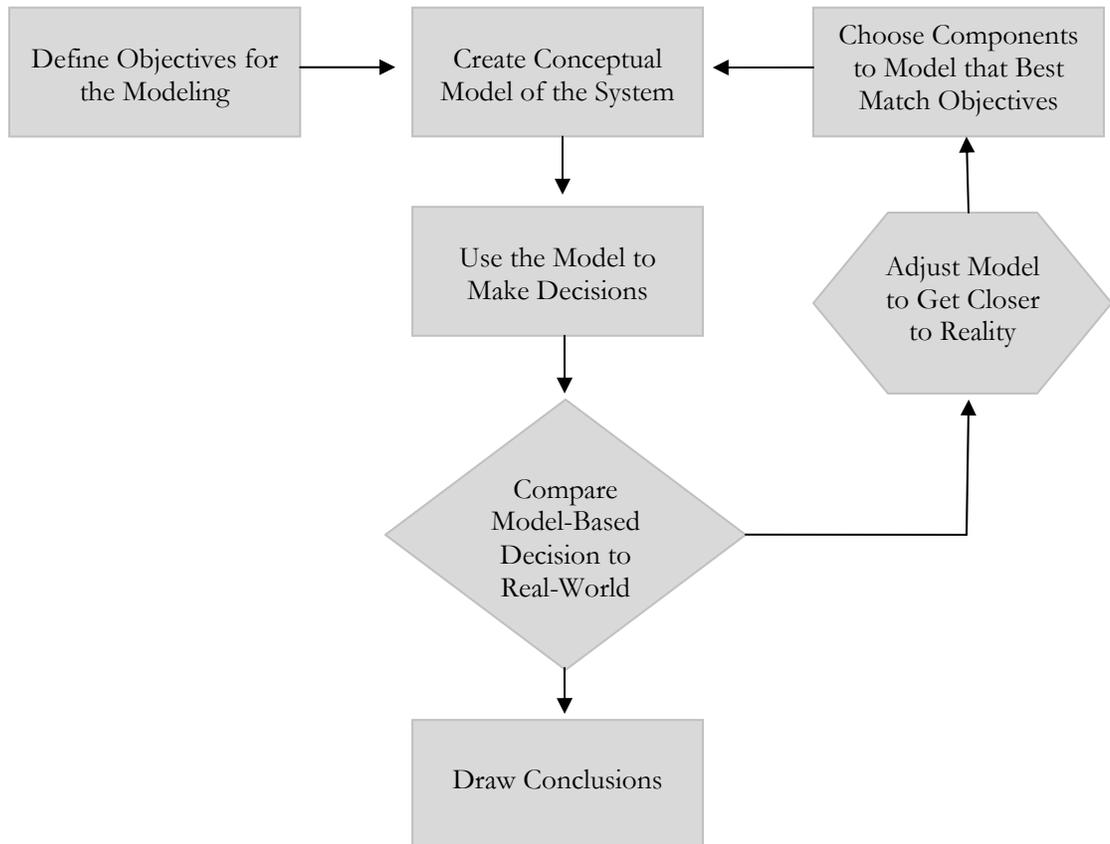
Once we have defined all of the potential conditions, we must decide how to simplify our model to reduce the amount of effort required to arrive at a reasonable estimate of the differences in time without requiring the gathering a huge amount of data. We will often have to trade-off or compromise the **accuracy** of a model with the resources required to operationalize the model. In this case, we will choose to compare the traffic flow across only a few selected routes under good weather conditions.

As we simplify the model, we need to decide which phenomena will be represented as **variables** (items whose values will change based on the relationships represented in the model) and which will be represented as **parameters** (items that are assigned a reasonable constant value to represent a finite set of conditions). For example, every stop sign you stop at may take a different amount of time depending upon how many other cars are approaching the same intersection. However, you may choose to create a parameter that

uses an average amount of stopping time to represent the range of conditions rather than have to gather data or find some other way of estimating the number of cars approaching each intersection during your trip. As we develop the models we use in the class, we will discuss the ways in which we can **verify** whether the computer is giving us the correct result and **validate** whether the answer is an accurate representation of the system.

The process we generally go through on a modeling project is represented by Figure 2.

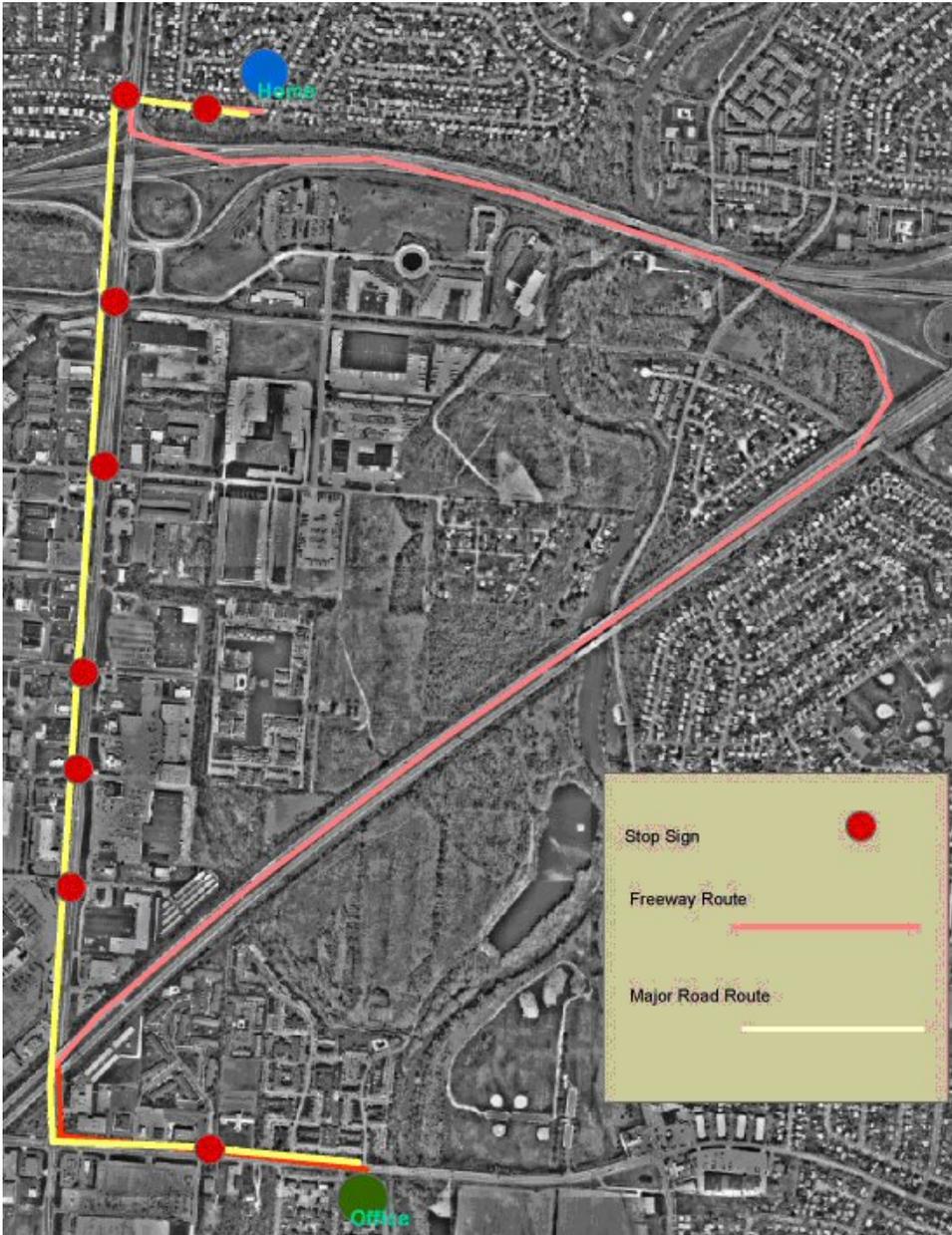
**Figure 2: The Modeling and Inquiry Process**



With our conceptual model in mind, we will use a simple version of the model in a spreadsheet representation of a simple model of a trip to work. Copy the file called traffic1.m from the shared drive in the lab. Open MATLAB on your computer. After the program opens, use the **File Open** command to get to the traffic1 file and open it. At the top of the MATLAB editor window that will open, you will see the definition of the major variables that are used to estimate total travel time. For each leg of the trip on a local street, the travel time is the distance divided by the average travel speed. Time is increased if you get to a stop sign and here is represented as a constant parameter of 30 seconds. The time to traverse a section of highway or a major street is the same as for a local street except that the average speed will be higher. If you get to a traffic light, the time will actually vary depending on the probability that you will hit a red light and the average amount of cycle time for the light to turn green. Finally, our model adds time if we need to make a left turn

at a major intersection as this will delay us further. Assuming there is a green turn arrow, we need to know its cycle time and the number of cars that can make the turn per cycle time.

The middle of the program shows three route options based on given/assumed data. The model considers two routes - a freeway route and a local road route as illustrated in Figure 3. The freeway route is considered for two conditions - with and without congestion. The path associated with each option is described in the program for each of these three cases. For the freeway route one takes the local streets to get to the freeway and then travels 2.5 miles to the freeway exit. From there, one has to negotiate a left turn and travel several blocks to the



destination. For the local street option you use a major road and travel 1.75 miles and face the possibility to be stopped at 5 traffic lights. The difference between congested and normal freeway is the difference in speed (55 mph uncongested freeway, 20 mph congested freeway). If you consider the possible objectives you might have in planning this trip, this

model actually deals with the first three objectives and presents data on the time each option will take.

Take a few minutes to look over the formulas in the MATLAB program and verify how each equation is used to calculate the total time for that segment. You in fact must verify that the mathematics of the equations correctly calculate the travel time given the assumptions that are made about the relationships. If you are confused about any of the calculations, ask the instructors. Compare what is simulated in the model to your conceptual model of travel time. Which of the variables you thought might be important are left out of this model? Under what circumstances would the model probably fail to accurately forecast travel time?

To run the program, click on **Debug** and then **run traffic1** in the editor window. The three answers to the times should appear in the MATLAB command window. Since you don't know how to program in MATLAB, we will have you copy and paste these values into a spreadsheet for further analysis as demonstrated in class.

Once you are sure you understand the model components, take some time to use the model to simulate alternatives. **Simulation** uses alternative sets of conditions to make a series of forecasts, giving the analyst insights into model behavior and helping to make a rational decision about what to do. For this example, use MATLAB to simulate the differences in travel time if the probability of hitting a red light changes. Change the appropriate parameter value starting from .5 to .3 in .05 increments. The program will recalculate the travel times. Copy the resulting times into a spreadsheet along with the parameter value they correspond to, forming a table of these values associated with each run. Testing these types of variations are called **sensitivity testing** for the model. The question is how sensitive is the model output (the time to get to work) to variations in the constants that are used as inputs. One way to measure sensitivity is to quantify the percent change in output per unit change in an input parameter. So in the case above, you should answer the question: for each 5% change in the probability of stopping at a traffic light (between 30 and 50%) how much is the average time to work increased or decreased? Is the model sensitive to this variable? To assist in your analysis, you should use the Excel graphing functions to visualize the relationship between the traffic light probability and the trip time by creating an X-Y scatterplot of the results.

Test the sensitivity of a second parameter – the average traffic speed by assuming that there is an accident on the Freeway which slows the average speed to 15 mph and diverts more traffic to the alternative route, reducing its average speed to 20 mph. Which route should you take?

## Assignment 2 Recap

Here is what you should prepare to hand in for this assignment:

1. A final spreadsheet showing the simulation data and the sensitivity of the model to the parameter changes.
2. A short report describing what you found:
  - a. Can you verify that the model we gave you is correct?
  - b. What things did we leave out of the model that could cause major changes in reality to the time it takes to get to work?
  - c. How sensitive is the model to changes in the parameters you tested? Describe this in quantitative terms, referring to your spreadsheet.

- d. Which route should you take to work under the range of conditions you tested?
- e. If this were a real place in Columbus, describe how you would collect data for three variables or parameters to validate the model.