MODULE 1.2

The Modeling Process

Introduction

The process of making and testing hypotheses about models and then revising designs or theories has its foundation in the experimental sciences. Similarly, computational scientists use **modeling** to analyze complex, real-world problems in order to predict what might happen with some course of action. For example, Dr. Jerrold Marsden, a computational physicist at CalTech, models space mission trajectory design (Marsden). Dr. Julianne Collins, a genetic epidemiologist (statistical genetics) at the Greenwood Genetics Center, runs genetic analysis programs and analyzes epidemiological studies using the Statistical Analysis Software (SAS) (Greenwood Genetics Center). Some of the projects on which she has worked involve analyzing data from a genome scan of Alzheimer's disease, performing linkage analyses of X-linked mental retardation families, determining the recurrence risk in nonsyndromic mental retardation, analyzing folic acid levels from a nutritional survey of Honduran women, and researching new methods to detect genes or risk factors involved in autism. Scientists in areas such as cognitive psychology and social psychology at the Human-Technology Interaction Center of The University of Oklahoma perform research on the interaction of people with modern technologies (Human-Technology Interaction Center). Some of the studies involve "strategic planning in air traffic control" and "designing interfaces for effective information retrieval from collections of multimedia." Buried land mines are a serious danger in many areas of the world (Weldon et al. 2001). Scientists are using a combination of mathematics, signal processing, and scientific visualization to model, image, and discover land mines. Lourdes Esteva, Cristobal Vargas, and Jorge Velasco-Hernandez have modeled the oscillating patterns of the disease dengue fever, for which an estimated 50 to 100 million cases occur globally each year (Esteva and Vargas 1999).

Definition **Modeling** is the application of methods to analyze complex, real-world problems in order to make predictions about what might happen with various actions.

Model Classifications

Several classification categories for models exist. A system we are modeling exhibits **probabilistic** or **stochastic behavior** if an element of chance exists. For example, the path of a hurricane is probabilistic. In contrast, a behavior can be **deterministic**, such as the position of a falling object in a vacuum. Similarly, models can be deterministic or probabilistic. A **probabilistic** or **stochastic model** exhibits random effects, while a **deterministic model** does not. The results of a deterministic model depend on the initial conditions; and in the case of computer implementation with particular input, the output is the same for each program execution. As we see in Module 9.2 on "Simulations" and other modules, we can have a probabilistic model for a deterministic situation, such as a model that uses random numbers to estimate the area under a curve.

Definitions A system exhibits **probabilistic** or **stochastic behavior** if an element of chance exists. Otherwise, it exhibits **deterministic behavior**. A **probabilistic** or **stochastic model** exhibits random effects, while a **deterministic model** does not.

We can also classify models as static or dynamic. In a **static model**, we do not consider time, so that the model is comparable to a snapshot or a map. For example, a model of the weight of a salamander as being proportional to the cube of its length has variables for weight and length, but not for time. By contrast, in a **dynamic model**, time changes, so that such a model is comparable to an animated cartoon or a movie. For example, the number of salamanders in an area undergoing development changes with time; and, hence, a model of such a population is dynamic. Many of the models we consider in this text are dynamic and employ a static component as part of the dynamic model.

Definitions A **static model** does not consider time, while a **dynamic model** changes with time.

When time changes continuously and smoothly, the model is **continuous**. If time changes in incremental steps, the model is **discrete**. A discrete model is analogous to a movie. A sequence of frames moves so quickly that the viewer perceives motion. However, in a live play, the action is continuous. Just as a discrete sequence of movie frames represents the continuous motion of actors, we often develop discrete computer models of continuous situations (Voinov 2003).

Definitions In a **continuous model**, time changes continuously, while in a **discrete model** time changes in incremental steps.

Steps of the Modeling Process

The modeling process is cyclic and closely parallels the scientific method and the software life cycle for the development of a major software project. The process is cyclic because at any step we might return to an earlier stage to make revisions and continue the process from that point.

The steps of the modeling process are as follows:

1. **Analyze the problem**
   We must first study the situation sufficiently to identify the problem precisely and understand its fundamental questions clearly. At this stage, we determine the problem's objective and decide on the problem's classification, such as deterministic or stochastic. Only with a clear, precise problem identification can we translate the problem into mathematical symbols and develop and solve the model.
2. **Formulate a model**
   In this stage, we design the model, forming an abstraction of the system we are modeling. Some of the tasks of this step are as follows:
   a. **Gather data**
   We collect relevant data to gain information about the system's behavior.
   b. **Make simplifying assumptions and document them**
   In formulating a model, we should attempt to be as simple as reasonably possible. Thus, frequently we decide to simplify some of the factors and to ignore other factors that do not seem as important. Most problems are entirely too complex to consider every detail, and doing so would only make the model impossible to solve or to run in a reasonable amount of time on a computer. Moreover, factors often exist that do not appreciably affect outcomes. Besides simplifying factors, we may decide to return to Step 1 to restrict further the problem under investigation.
   c. **Determine variables and units**
   We must determine and name the variables. An **independent variable** is the variable on which others depend. In many applications, time is an independent variable. The model will try to explain the **dependent variables**. For example, in simulating the trajectory of a ball, time is an independent variable; and the height and the horizontal distance from the initial position are dependent variables whose values depend on the time. To simplify the model, we may decide to neglect some variables (such as air resistance), treat certain variables as constants, or aggregate several variables into one. While deciding on the variables, we must also establish their units, such as days as the unit for time.
   d. **Establish relationships among variables and submodels**
   If possible, we should draw a diagram of the model, breaking it into submodels and indicating relationships among variables. To simplify the model, we may assume that some of the relationships are simpler than they really are. For example, we might assume that two variables are related in a linear manner instead of in a more complex way.

e. **Determine equations and functions**
While establishing relationships between variables, we determine equations and functions for these variables. For example, we might decide that two variables are proportional to each other, or we might establish that a known scientific formula or equation applies to the model. Many computational science models involve differential equations, or equations involving a derivative, which we introduce in Module 2.3 on "Rate of Change."

3. **Solve the model**
This stage implements the model. It is important not to jump to this step before thoroughly understanding the problem and designing the model. Otherwise, we might waste much time, which can be most frustrating. Some of the techniques and tools that the solution might employ are algebra, calculus, graphs, computer programs, and computer packages. Our solution might produce an exact answer or might simulate the situation. If the model is too complex to solve, we must return to Step 2 to make additional simplifying assumptions or to Step 1 to reformulate the problem.

4. **Verify and interpret the model's solution**
Once we have a solution, we should carefully examine the results to make sure that they make sense (verification) and that the solution solves the original problem (validation) and is usable. The process of **verification** determines if the solution works correctly, while the process of **validation** establishes if the system satisfies the problem's requirements. Thus, verification concerns "solving the problem right," and validation concerns "solving the right problem." Testing the solution to see if predictions agree with real data is important for verification. We must be careful to apply our model only in the appropriate ranges for the independent data. For example, our model might be accurate for time periods of a few days but grossly inaccurate when applied to time periods of several years. We should analyze the model's solution to determine its implications. If the model solution shows weaknesses, we should return to Step 1 or 2 to determine if it is feasible to refine the model. If so, we cycle back through the process. Hence, the cyclic modeling process is a trade-off between **simplification** and **refinement**. For refinement, we may need to extend the scope of the problem in Step 1. In Step 2, while refining, we often need to reconsider our simplifying assumptions, include more variables, assume more complex relationships among the variables and submodels, and use more sophisticated techniques.

5. **Report on the model**
Reporting on a model is important for its utility. Perhaps the scientific report will be written for colleagues at a laboratory or will be presented at a scientific conference. A report contains the following components, which parallel the steps of the modeling process:

a. **Analysis of the problem**
Usually, assuming that the audience is intelligent but not aware of the situation, we need to describe the circumstances in which the problem arises. Then, we must clearly explain the problem and the objectives of the study.

b. Model design

The amount of detail with which we explain the model depends on the situation. In a comprehensive technical report, we can incorporate much more detail than in a conference talk. For example, in the former case, we often include the source code for our programs. In either case, we should state the simplifying assumptions and the rationale for employing them. Usually, we will present some of the data in tables or graphs. Such figures should contain titles, sources, and labels for columns and axes. Clearly labeled diagrams of the relationships among variables and submodels are usually very helpful in understanding the model.

c. Model solution

In this section, we describe the techniques for solving the problem and the solution. We should give as much detail as necessary for the audience to understand the material without becoming mired in technical minutia. For a written report, appendices may contain more detail, such as source code of programs and additional information about the solutions of equations.

d. Results and conclusions

Our report should include results, interpretations, implications, recommendations, and conclusions of the model's solution. We may also include suggestions for future work.

6. Maintain the model

As the model's solution is used, it may be necessary or desirable to make corrections, improvements, or enhancements. In this case, the modeler again cycles through the modeling process to develop a revised solution.

Definitions The process of **verification** determines if the solution works correctly, while the process of **validation** establishes if the system satisfies the problem's requirements.

Although we described the modeling process as a sequence or series of steps, we may be developing two or more steps simultaneously. For example, it is advisable to be compiling the report from the beginning. Otherwise, we can forget to mention significant points, such as reasons for making certain simplifying assumptions or for needing particular refinements. Moreover, within modeling teams, individuals or groups frequently work on different submodels simultaneously. Having completed a submodule, a team member might be verifying the submodule while others are still working on solving theirs.

The modeling process is a creative, scientific endeavor. As such, a problem we are modeling usually does not have one correct answer. The problems are complex, and many models provide good, although different, solutions. Thus, modeling is a challenging, open-ended, and exciting venture.

Exercises

1. Compare and contrast the modeling process with the scientific method: Make observations; formulate a hypothesis; develop a testing method for the hypothesis; collect data for the test; using the data, test the hypothesis; accept or reject the hypothesis.
2. Compare and contrast the modeling process with the software life cycle: Analysis, design, implementation, testing, documentation, maintenance.

References

Esteva, Lourdes, and Cristobal Vargas. 1999. "A Model for Dengue Disease with Variable Human Population," *J. Math. Biol.*, 38: 220–240. http://147.46.94.112/e_journals/pdf_full/journal_j/j28_380302.pdf

Giordano, Frank R., Maurice D. Weir, and William P. Fox. 2003. *A First Course in Mathematical Modeling*. 3rd ed. Pacific Grove, Calif.: Brooks/Cole-Thompson Learning.

Greenwood Genetics Center. "Welcome to the GGC." http://www.ggc.org/

Human-Technology Interaction Center. "HTIC." The University of Oklahoma. http://www.ou.edu/HTIC/

Marsden, Jerrold. "Control and Dynamical Systems." California Institute of Technology. http://www.cds.caltech.edu/~marsden/

Voinov, Alexey Arkady. "Model Classifications." Course notes for CS 295 Computer Simulation and Modeling, University of Vermont, Gund Institute for Ecological Economics, 2003. http://www.uvm.edu/giee/AV/CS/class2.1.html (accessed March 20, 2005).

Weldon, T. P., Y. A. Gryazin, and M. V. Kibanov. 2001. "Novel Inverse Methods in Land Mine Imaging." *IEEE International Conference on Acoustics Speech and Signal Processing* 2001 (ICASSP 2001). Salt Lake City, UT, May 7–11, 2001. http://wws2.uncc.edu/tpw/papers/icassp01.pdf