0

Modeling Basics: Purpose, Resolution, and Resources

If we were to ask several people for examples of models, we would get a variety of responses which might include mathematical equations, toy trains, prototype cars, or fashion models. What these very different objects have in common is that they are representations of reality. The equations may represent the growth of a population, the toy train is representation of a real train, the prototype is a representation of a future car, and a fashion model is a representation of how clothes will look when worn and maybe other things that one has to live in the world of fashion to appreciate. A model is, then, a representation of reality. This, however, is not a sufficient definition. How can one evaluate different models? A fashion model and a toy train are both representations of reality, but so vastly different that they are incomparable. A street map of a city and a road map of the whole United States are representations of reality which are similar, but neither can substitute for the other. Intrinsic in a model is a sense of *purpose.* We define a *model* to be a purposeful representation of reality. A street map is a representation of reality for the purpose of navigating streets in a particular city. It is useless for driving across country or even for locating traffic jams or construction within a city, but that does not make it a bad model. Other models are needed for those purposes. A successful fashion model serves the purpose of selling clothes, perhaps by creating some fantasy about what people will look like. An artist's model serves a different purpose. For this purpose, perhaps someone who is less glamorous, but with more idiosyncrasies might be a better model. There are different types of toy trains: some are for young children, some are for older children, and others are for adults. Again different purposes result in different models.

This book deals exclusively with *mathematical models* which are models built using the tools and substance of mathematics (including computers and computer software). We do not deal with train sets or fashion models. Further, a major emphasis of this book is on *modeling* which is the process of building mathematical models. Models are presented of course, but more as examples of the results of the modeling process and less as ends in themselves.

Doing a modeling project will help illustrate several modeling concepts. You, the reader, are encouraged to get the following materials and do the tasks that are requested of you. You need a jar, preferably of an interesting shape, and a bag of M&M candies or some similarly sized objects (jelly beans, marbles, or pennies). Here is your project:

Project: Take exactly one minute and determine how many M&M candies will fit in your jar.

Warning: Not doing this project will result in inferior learning, low self esteem, and poorer jobs with lower pay.

Now that you have done this, let us think about what you have done. First the concept of a *resource constraint* should have meaning to you. If you honestly attempted to do this problem in one minute, you may have been frustrated thinking that you could do this problem if you only had more time. There may have been other frustrations involving resources; for example, you could have done this if you had a ruler or some other measuring device. While this deficiency of resources may seem to be an unfair constraint it is something modelers have to cope with all the time. For example, your boss may tell you one morning to determine the effects of some new strategy for a meeting scheduled that afternoon. The boss does not care about, "I could do it if only . . ." The model you construct given a couple hours will be very different from the model you would construct given several days or years. But it is the best you can do within the constraints. Other resource constraints are money, personnel, computing power, and expertise. It may be impractical for you to buy a computer, hire a staff, or go back to school just to solve a problem.

One thing missing from the project above, which is a justifiable complaint, is that the purpose of the project was not stated. Why do you need to know how many M&M's fit in the jar? If the purpose is to determine if one bag of M&M candies would fill the jar, a quick visual inspection of the jar and of the bag would give the answer, and 60 seconds would be more than enough time. If the purpose is to figure out how many bags of M&M's would fill the jar to a point that most anyone would consider to be full, then one needs to do some figuring. Being off by a couple handfuls of candies, however, will not make any difference. Finally, if the purpose is a contest and the person who can fit the most M&M candies into the jar so that the lid can be fully tightened without squashing any candy wins valuable prizes, then it is very important to figure out exactly how many candies fit in the jar. Again the purpose is a vital part of any model that you build.

These ideas bring up the notion of *resolution*. This term may be familiar from photographs or computer images. It is hard to determine details in a photograph with low resolution, but the details are very sharp in a high-resolution picture. Models have resolution also. In the previous paragraph, we discussed three purposes for the problem of finding how many candies fit in a particular jar. In each case, the answer had a different resolution: more or less than one bag; to within a couple handfuls; the exact maximum number.

If the purpose of a model requires only a low-resolution answer, then a low-resolution model is completely acceptable. Even if resources permitted a better answer, the better answer would not be any more useful. On the other hand, if a high-resolution answer is desired, low-resolution models still have their place. Often one starts with a low-resolution model to get a ballpark estimate and successively refines the model until either the desired resolution is achieved or resource limitations prevent one from proceeding further.

Let us go back to our modeling project and revise it a little.

Project 2: A jigsaw puzzle company wants to fill your jar with M&M candies so that a photograph can be taken with the jar looking more or less full (a handful shy will not make any difference). A bag of M&M's holds 50 candies. How many bags should they buy? You have 15 minutes to complete this project. Before you start you may acquire any standard measuring instruments you think you will need. One constraint: do not start with more than one bag of candy (several bags are okay, but you are constrained so that you cannot fill up the jar, dump it, and count the candies).

Assuming you have done this project, let us discuss what you have done. If you thought this would be an easy problem if you just had more time earlier, do you still think this is the case? There is no "right way" to approach this problem. We mention a few. Did you compute the volume of the jar? If you did this, did you approximate the jar as a cylinder, or as a small cylinder surmounting a larger one? Maybe you forgot the formula for a volume of a cylinder and used rectangular boxes instead of cylinders. Maybe you remembered the formula for the volume of a truncated cone and used that. If you had access to a well equipped kitchen you might have found the volume exactly by filling the jar with water and pouring it into a measuring cup. Perhaps your jar had the volume printed on it in some way. Now how about the M&M's? Did you treat them as ellipsoids with volume πabc? Did you treat them as little cylinders, or as little boxes?

Here is one approach to the problem. Approximate the volume of the jar using two cylinders taking care that the volume of the two cylinders is greater than or equal to the true volume. Call this volume V. Next treat an M&M candy as an ellipsoid and suppose its volume is v. Assuming that the volumetric units are the same, V divided by v gives the absolute maximum number of M&M's needed from which we can determine the number of bags needed. Of course this assumes that every millimeter of the jar is covered by candy. If you look at M&M candies poured into a jar, however, you notice that there is a lot of empty space. We can obtain a lower bound by considering the M&M's as cylinders and determining how many lie in one layer at the bottom of the jar and then figuring the number of layers the jar will hold. We may have to use multiple cylinders to represent the jar because of the neck. In any case, we can get a good estimate for how many M&M's would fit in the jar if they were stacked uniformly one on top of the other. Take a moment to convince yourself that, assuming one had the patience to stack them, one bump would knock the M&M's out of their columns and some would fit into the space between M&M's on the previous layer. This is called packing, and when it happens, which is always, more M&M's will be needed to fill the jar. Thus we have two numbers giving us a lower bound and an upper bound, and the true number is somewhere

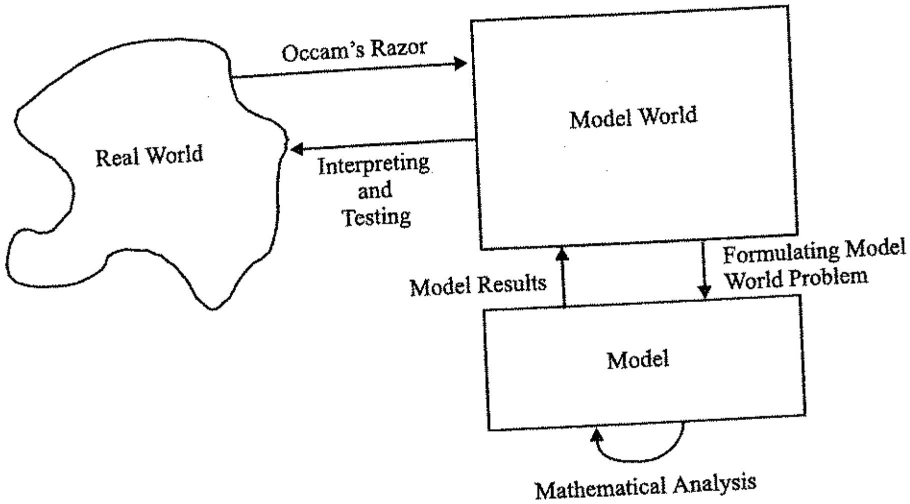

FIGURE 0.1 Model World Diagram.

in between. Despite the simple sounding nature of this problem there is a great deal of complexity. Complexity is associated with any real problem. Consider Figure 0.1.

The problems that modelers wish to solve exist in the real world. The real world is a nasty place with all sort of complications. Bottles are irregularly shaped. Candies are shaped so they can fit together in many ways, but always having some space between them. To illustrate this, the real world is given a nebulous shape. The first step in the modeling process is to simplify the real world to create a model world. As the picture depicts, the model world leaves out much of the complexity of the real world. For example, a bottle becomes two cylinders. The M&M's become cylinders. The original question gets translated into a question involving the model world. In the model world, the question is how many M&M cylinders fit inside the two bottle cylinders. Next we construct a model of the problem in the model world. All the tools and techniques of mathematics can now be applied to the problem in the model world to get an answer in the model world. A very important mistake to avoid is to assume that the solution to the problem in the model world answers the question in the real world. The final step is to interpret the answer found for the model world problem, back in the real world. Here we notice that the M&M's can pack and that the shape of the jar is not two cylinders. Depending on our purpose, resolution, and resources, we may need to work through the cycle one or more times to get a satisfactory answer.

In most mathematics classes, we start with an equation and solve it. Occasionally we start with a problem in a model world which we solve and answer back in the model world. In this book, we do both of these, but also carry the process further to begin with problems in the real world and work our way completely through this diagram to get answers in the real world.

The process of converting the real world into a model world is probably the most important step in the modeling process. This process is referred to as using "Occam's razor." William of Occam, a fourteenth-century philosopher, coined a phrase in Latin which literally translates to: "things should not be multiplied without good reason." In other words, don't make things harder than they need to be. In a modeling setting, exclude the details which are irrelevant given the purpose, or which cannot be handled given the constraints. We are cutting the world down to manageable size, as if with a razor, hence the term Occam's razor. Occam's razor is a sensitive and dangerous tool. Cut out too much, and the model solutions have nothing to do with reality. Cut out too little, and the problem is too difficult to solve with the available resources.

A model that every calculus student has probably seen is that projectiles follow parabolic trajectories. This is derived by starting with a constant acceleration rate g and initial conditions (s_0 and v_0) and integrating the equation $a(t) = g$ twice to get the displacement as a function of time $s(t) = \frac{gt^2}{2} + v_0t + s_0$. Let us examine this model in light of Figure 0.1 and Occam's razor. The problem in the real world was originally predicting the motion of cannon balls and perhaps things like bullets and stones. (A great deal of technological advances have been made exploring more efficient ways of killing people.) This problem was moved into the model world. Occam's razor sliced away and left a world which was flat, with constant gravitational force, in a vacuum, with no other forces of any kind. Within this world, the model for parabolic motion was set up and solved. Then it was tested in labs and on battlefields. For these purposes, the model was very good. One danger is to assume that because the model predicts that all projectiles have parabolic motion, in reality they all will. Try throwing a feather. Try throwing a light plastic ball into the wind. Try dropping a BB into a jar of honey. Missile designers realized that they could not assume constant gravitational force or neglect air resistance. However, for throwing rocks and firing cannon on windless days, the parabolic model is a good model.

Occam's razor creates some assumptions which should be spelled out when reporting on a model that has been built. With the projectile motion, we spelled out that the model world was flat, was a vacuum, and had constant gravity and no other forces. This is a vital part of a model. When someone throws a feather and says, "Ah ha, your model is wrong," we can reply, "Well, the feather would have followed parabolic path if you had thrown it in a vacuum." The accuser would have to accept this. The next complaint might be that your model is unrealistic because we do not live in a vacuum. To which we reply that for objects of certain masses projected certain distances, the effects of not being in a vacuum are negligible. Of course we would have to back this up, and this is what the lab and field tests verify. Assumptions and model testing are two very important lines of defense against skeptics and critics.

Here is another modeling project which we will only think about and not actually solve. It serves to illustrate many of the concepts of this book. Suppose we are starting a limousine service, driving people from a particular town to the nearest airport. For the sake of illustration, suppose that this is 70 miles one way. We want to compute the cost in gasoline of these trips for budgeting purposes. Our car has an average gas mileage for this type of driving of 25 miles per gallon.

A first pass at this problem would be to divide the distance to the airport by the mileage and multiply by the current gasoline price. However, after thinking about things, we realize that the current gasoline price is changing. We look up records for the last five years and compute a function which predicts the price of gas into the future, understanding that while reality may be very different from this prediction, it will probably be good for the short term. We repeat the model with this gas price function. Now we have a *time dependent* model. The questions it answers are not just how much will it cost, but how much will it cost at different times. Both of the models we have constructed are *deterministic* models. They include no random factors and consequently predict one number or one number at each point in time. Do we believe these numbers? We might suspect that different trips to the airport might result in different gas mileages due to random factors such as traffic and weather. If we include these into our model, we can run simulations and end up, not with a single number, but with a distribution of numbers. Models which include randomness are called *stochastic*. We may think of other random factors such as varying distances based on where customers are picked up or where the driver is able to park. At some point the model may become unwieldy and we must decide (use Occam's razor) whether all this detail is important to our purpose and worth the necessary resources.

The first chapter of this book deals with deterministic time dependent models where the change in time occurs in jumps (gas prices change weekly or daily, but not from instant to instant). The second chapter adds stochasticity into models and examines ways of analyzing the output. The third chapter adds additional complications such as a limousine service with multiple cars with different gas mileages. The fourth chapter deals with ways of looking at data such as gas prices over time and fitting functions to them. The fifth chapter deals with time dependent models where the change over time is continuous (for example if the gas mileage steadily decreased between service intervals). The last chapter covers randomness which occurs continuously. All these chapters provide different ways of analyzing essentially the same problem. Occam's razor must be used with the issues of purpose, resolution, and resource constraints to decide which approach is best for a particular situation.

While we do not often think about it, we do modeling all of the time. Here is a another project. How long will it take to finish reading this chapter? This is a problem we may solve frequently when reading before bed and get tired. We could answer this question by computing our reading rate for this type of material and determining the number of words left to go. Maybe we could include a factor for reading rates slowing down as we get tired. But if we are trying to figure out whether to finish the chapter now or later, this degree of resolution is not necessary. For this purpose, a sufficient answer is "not long."

0.1 Further Reading

These first two books are good continuations of the ideas presented in this chapter. Both go considerably deeper:

John Harte, *Consider a Spherical Cow: A Course in Environmental Problem Solving*, University Science Books, Mill Valley, California, 1988.

Anthony M. Starfield, Karl A. Smith, and Andrew L. Bleloch, *How to Model It: Problem Solving for the Computer Age*, McGraw-Hill Publishing Company, New York, 1990.

The next book is an introductory modeling classic. It is recommended additional material for the whole book so we list it at the beginning and will not repeat it every chapter. The material is at a similar level to the material in this book, but it places greater emphasis on engineering-type problems:

Frank R. Giordano and Maurice D. Wier, *A First Course in Mathematical Modeling*, Brooks/Cole Pub. Co., Monterey, CA, 1985.

0.2 Exercises

1. How many 1,000 cubic-inch cylindrical tin cans can be formed from a sheet of metal one yard wide and two yards long?

2. How many restaurants are there in the United States? Clearly it is impossible, for all practical purposes, to find the right answer as restaurants open and close daily. First establish an acceptable error range for your answer and then try to answer this question. You may consult any source for information (such as population size, or number of cities) with the exception of anything which directly deals with this question. Clearly state the assumptions you are making. Next find a source which states the number of restaurants there are. Compare your answer to this answer. Realizing that any "official" number is also an estimate and not truth, which answer are you more confident in and why?

3. Estimate the number of dots in Figure 0.2? Your method should not involve counting a large number of dots; counting a sample is acceptable. Suppose this figure represents

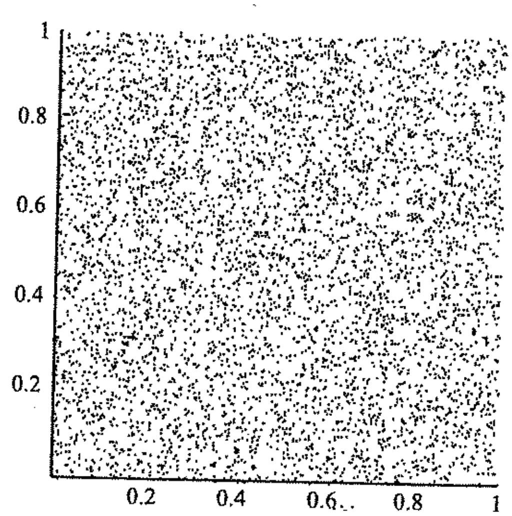

FIGURE 0.2 How Many Dots?

aerial photograph of a tree farm with each dot representing a tree. Your purpose is to tell a lumber company how many trees are in this tract of land so they can make revenue projections.