MODELS AND MODELING: AN INTRODUCTION FOR EARTH AND ENVIRONMENTAL SCIENTISTS

JERRY P. FAIRLEY

with website

WILEY Blackwell

Models and Modeling

Models and Modeling

An Introduction for Earth and Environmental Scientists

Jerry P. Fairley

WILEY Blackwell

Registered Office
John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

Names: Fairley, Jerry P., author.
Title: Models and modeling : an introduction for earth and environmental scientists / Jerry P. Fairley.
Description: Chichester, UK ; Hoboken, NJ : John Wiley & Sons, 2016. |
Includes bibliographical references and index.
Identifiers: LCCN 2016024807 (print) | LCCN 2016033087 (ebook) | ISBN 9781119130369 (pbk.) |
ISBN 9781119130383 (pdf) | ISBN 9781119130376 (epub)
Subjects: LCSH: Earth sciences–Mathematical models. | Environmental sciences–Mathematical models.
Classification: LCC QE33.2.M3 F35 2016 (print) | LCC QE33.2.M3 (ebook) | DDC 550.1/5118–dc23
LC record available at https://lccn.loc.gov/2016024807

A catalogue record for this book is available from the British Library.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Cover image: iStockphoto/Alexander Shirokov

Set in 9.5/13pt Meridien by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

To my father, for teaching me to do math one step at a time.

Contents

About the companion website

This book is accompanied by a companion website:
www.wiley.com/go/Fairley/Models

This website includes:

- Powerpoints of all figures from the book for downloading
- Solutions to many of the problems given in the chapters

Introduction

Before serious discussion on any topic can take place, it is usually necessary to define one's terms. In the present instance, we need to ask the following question: "what, exactly, do we mean when we speak of a 'model'?" A reasonable definition of the term might be *a simplified or idealized representation of reality.* Although this is a pretty broad definition, it is perhaps the only one that really embraces everything we refer to as "a model."

In order to better understand what is, and what is not, a model, we can examine some possible examples. Which of the following would be considered models?

- A map
- A model train
- A model United Nations
- A fashion model
- An atom

Of course, by our definition all of these things would qualify as models. Students will sometimes argue that "an atom" is a physical object and, therefore, not a model. However, everything we know about atoms comes from observing their effects, and our understanding and depiction of atoms has changed and evolved over time. In fact, our current understanding of atoms and subatomic particles is so far outside our everyday experience that our descriptions are entirely probabilistic. As a result, I maintain that any discussion of "atoms" is necessarily a discussion of models—simplifications and idealizations of things that are beyond our direct knowledge.

In this book, we will primarily concentrate on a subset of models known as *mathematical models*. A mathematical model is a kind of model that is formulated in terms of mathematical concepts such as constants, variables, functions, derivatives, and so on. The nominal goal of a mathematical model is to give quantitative indications of the behavior of some aspect of a system.

Note that I said we would *primarily* be concentrating on mathematical models. In fact, the word "primarily" glosses over an important fact about the development of mathematical models: *every mathematical model is the quantification*

Models and Modeling: An Introduction for Earth and Environmental Scientists, First Edition. Jerry P. Fairley.

Companion website: www.wiley.com/go/Fairley/Models

of an underlying conceptual model; that is, *a mathematical model is the quantification of our idea of how a system works*. This is a very important, but often overlooked, aspect of model development. The fact is that model development requires many decisions to be made about which aspects of the system are important, and which aspects may be neglected. If the decisions are poor, a poor model will result. If the decisions are good, a good model *may* result (no guarantee). In any case, before a *mathematical* model can be developed, it is first necessary to define a *conceptual model*, or conceptual understanding of the system. The usefulness of the quantitative model, and the reliability of its representation of the system, will be wholly dependent on this conceptual understanding.

Another aspect of models that is deserving of discussion is the uses to which models may be put. A sample of these might be the following:

- To teach
- To "predict" system behavior or the consequences of actions
- To develop or test theories
- To plan or design field/laboratory tests
- As propaganda to support a partisan viewpoint

Of these, the use of models as propaganda is probably the most common application for models (e.g., think about the usefulness of fashion models and model cars in selling products). It is tempting to think that "scientific" (i.e., mathematical) models are exempt from issues of partisanship, but this is far from the truth. Because every model is a quantification of an individual's conceptual understanding, all models are developed from a partisan point of view; that is, *every model contains bias!* Therefore, the first step in evaluating any model is to understand the motivations and biases of the developers.

Another topic it would be appropriate to touch on is *model fallibility*. It is not uncommon to hear statements similar to "it must be true, because the model says so," and models are commonly depicted in popular culture as unfailing guides to the future. In reality, the knowledgeable modeler understands that a model embodies our conceptual understanding of some system, and is therefore incomplete and prone to error. Furthermore, there are an infinite number of alternative models that could be devised to represent any given system. Since it is impossible to test all the alternatives for a superior model, *a model can never be shown to be the "true" representation of a system.*

It is also easy to make the mistake of thinking that a model is reliable because it has been "validated" or "verified." In my opinion, the idea of "model validation" is misleading at best; it implies that, because a model has produced good results in the past, future predictions may be taken uncritically. In fact, every new prediction or new situation to which a model is applied will take a model into scenarios outside of the conditions for which it was calibrated and tested. Under these circumstances, the modeler must always be alert to the possibility that new processes, changes in parameters, or insufficiencies of the underlying conceptual model will cause the model to fail.

Instead of using the terms "validated" or "verified," I prefer to think in terms of *confidence building*. If a model performs well when tested against observations, we gain confidence in the model's output, and each additional test increases our confidence. We retain our confidence as long as the model continues to perform reliably; however, we never lose sight of the fact that, at some time in the future, the model parameterization or underlying conceptualization may be insufficient for the task at hand, and the model may fail. Keep in mind that, although a model can never be proven to be "right," it only takes one counterexample to prove it wrong.

Usually, models are created for some useful purpose—to understand some phenomenon, plan some test, or make some prediction—and I am a believer in the idea that models should be useful. However, I would be less than honest if I didn't say that one of the reasons for developing models was the pleasure to be had in their development. I believe this is as good a reason to develop a model as any, and perhaps better than many others. As you work through the models presented in the following chapters, I hope that a little of the enjoyment and gratification that comes from working with models will shine through.

CHAPTER 1
Modeling basics

Chapter summary

Regardless of whether the subject to be modeled is a groundwater flow system, the growth of a bacterial colony, the flight of a projectile shot from a cannon, or any other system, the construction of a model is not a simple step-by-step procedure. Because the modeler must make decisions about which processes should be included in the model and which will be neglected, what domain will be used, and so on, the development of a model is as much art as science. Notwithstanding the creative aspects of model development, there are some simple but important rules that should be followed. In this section, we will examine three basic rules for model development, briefly discuss some important aspects of model formulation, and make some suggestions for evaluating a model's performance. These rules and suggestions will form the basis for all of the examples of model development in the following chapters.

1.1 Learning to model

How to develop a model of a physical system, when to believe and when not to believe the model output, and how to determine whether the model predictions have any relevance to real life are common and central questions that must be answered by the would-be modeler every time a new situation is encountered. Most commonly, modelers are shown how to use a software package (or asked to read the documentation for a software package), and then assumed to be sufficiently competent to produce reliable predictions of system behavior. Even a moment's reflection will show that this is a nonsensical way to go about learning the craft of modeling, and this attitude has been largely responsible for the proliferation of bad models and the subsequent lack of confidence in modeling (and modelers).

Ideally, a modeler would gain experience and a deep understanding of the process of model development while working as an "apprentice" under a skilled modeler. This desirable state of affairs is seldom met with in the real world, however. The purpose of this book is to provide some guidance for those aspiring

Models and Modeling: An Introduction for Earth and Environmental Scientists, First Edition. Jerry P. Fairley.

Companion website: www.wiley.com/go/Fairley/Models

modelers who do not have the advantage of serving such an apprenticeship. Although not a substitute for the teaching and advice of an experienced modeler, it is to be hoped that the rules and examples in this and the following sections will at least keep the novice modeler from falling into some of the more obvious pitfalls associated with the mathematical modeling of physical systems.

1.2 Three cardinal rules of modeling

It is probable that, over time, many hundreds of "rules" have been made up regarding the construction, evaluation, and application of mathematical models. Most of these purported rules would be better classified as "suggestions," "considerations," or even, in some cases, as "superstitions." Over many years of making and using models, however, I have become convinced that the following three cardinal rules should be followed at all times in the development of a mathematical model:

1. Always know exactly the objective of model development.
2. The model you develop should be appropriate for the available data.
3. Start with the simplest possible model of the system, even if it is completely unrealistic. Once you thoroughly understand this preliminary model, add complexities to the model *one at a time* until you arrive at a satisfactory representation of the system.

In my experience, all three of these rules are routinely overlooked by modelers, and many poor and inappropriate models have resulted. We will briefly consider each of these rules here, but, more importantly, they are bound into the fiber of every model developed in the following chapters.

1.2.1 Rule 1: Know your model objective

It is common for a modeler to start a modeling investigation with the objective of "making a model of the aquifer" or with a similarly vague idea of what is to be accomplished. I cannot state strongly enough that a modeler must know exactly what s/he is trying to accomplish before ever putting pen to paper (or typing an input parameter). The more precisely the objectives of the model are known, the more likely the investigation will be successful. Make a habit of writing down the objective of your model, and be ready and willing to reduce your objective to a single sentence. The objective "to model the wells in the Grande Ronde Aquifer" is a very poor statement of purpose; a better (although still insufficient) objective is "to determine the influence of pumping well MW-4 on nearby wells." Better yet (and possibly sufficient to begin an investigation) is "to estimate the change in head in wells MW-1 and MW-3 that results from a 24-hour constant rate pump test in well MW-4." The examples in the following chapters always include a statement of that which is to be found; hopefully, after working through the examples, the reader will have a clear idea of how to formulate model objectives.

Although the temptation to "just get modeling" and "show some results" may be strong, you will always be better off if you first make certain you understand exactly what it is you want to achieve, and formulate a plan to reach that goal.

1.2.2 Rule 2: Make your model appropriate for your data

Hydrogeologists and environmental scientists are often working in data-poor environments. There is usually little to be gained from building a three-dimensional (3D), coupled saturated–unsaturated zone model with heterogeneous property sets when the only data to constrain the model come from a single aquifer test. Often in these situations a simple analytical model will give results that are as reliable as (or more reliable than) a complex numerical simulation. Furthermore, complex numerical simulations are often misleading, since it is tempting to think that, because they are complicated, they are realistic. What non-modelers (and many modelers) are unaware of is the fact that any computer model is solving the same equations that the modeler can write down with a pencil and paper. If there are few data to support the added complexity in terms of spatially varying properties, time-dependent recharge or boundary conditions, and so on, then the complicated numerical simulation may in fact be a worse representation of the system than a greatly simplified analytical model.

It should be said that many clients, regulatory agencies, and other downstream users of model output will push for complex numerical simulations in spite of the paucity of data to support such simulations. Although economic, political, or regulatory pressures may force a modeler to undertake the development of 3D simulations when only 1D simulations are justified, or a transient model when a steady-state model would do, the modeler should at all times be aware of the limitations of the models s/he is working with. By following Rule 3 (Section 1.2.3), the savvy modeler will be able to develop the more complex model demanded by the client while still maintaining her or his integrity and a high standard of modeling ethics.

1.2.3 Rule 3: Start simple and build complexity

When faced with a complex and difficult real-life situation, it is tempting to start out by developing a model that includes the most important processes. For example, if the goal is to understand the impact of a pumping well on other nearby wells, a novice modeler might want to build a model that includes variations in the rate of pumping, recharge from rainfall or snowmelt, the influence of changing water levels in a nearby lake, and other similar items that are clearly needed for a realistic representation of the system. The problem is that there is no way, in such a complex conceptualization, for the modeler to determine if the model output makes sense or not. Your first attempt at modeling a system should always be the simplest possible representation. Rather than modeling a 3D transient system, begin by modeling a 1D or 2D steady-state system with no source terms or other complexities. Although this may not be a realistic representation of the system, at least the modeler will know if the results are reasonable. Next, the modeler may add a spatially and temporally constant source term; again, evaluate the results. Do they make sense? Can you convince yourself the output is reasonable? If so, add another complexity, reevaluate, and so on, until the final product is one that you both understand and believe in. Never go on to the next step until you have complete confidence in, and an intimate understanding of, the current step—as well as all the steps that lead to the current step.

1.3 How can I evaluate my model?

Particularly with regard to Rule 3 (Section 1.2.3), one may rightly ask the question: "how do I know my model is giving reasonable results?" There is no easy answer to this question; in part, knowing a model is giving reasonable results is the product of experience with models. There are, however, some simple suggestions that can help to uncover problems with a model; these are described in the following text.

1.3.1 Test model behavior in the limits

The quickest way to begin an examination of model behavior is to check the behavior of a model in the limits. What is the model behavior at time $t = 0$? What about as $t \rightarrow \infty$? If you set a parameter to 0 or check the limit as it goes to ∞, is the result what you would expect? These kinds of tests are most readily carried out with analytical models, but, with some ingenuity, they can usually be applied even to complex numerical simulations.

1.3.2 Look for behavior congruent with the governing equations

Even very complex numerical simulations are based on a few well-known equations such as the Laplace equation, Poisson's equation, and the transient diffusion equation. Each of these equations implies particular behaviors, and you should check to make certain your model is behaving in a fashion that accords with your understanding of the underlying governing equations. (We will examine the characteristic behaviors of the most common equations in the following chapters.) Check for maxima and minima, look at the curvature of the predicted model surface, and examine any discontinuities in the output. Are these features present (if you expect them) or not (if you don't expect them), and are they in the appropriate places? It is a good sign if your expectations, based on your understanding of the characteristics of the governing equations, are realized. If the behavior you observe in your model output is different than your expectations, you need to understand why these differences arise before moving forward.

1.3.3 Nondimensionalization

Whenever possible, you should nondimensionalize your model (nondimensionalization is described in Appendix A, and illustrated throughout this text). Note that, although you will rarely be able to nondimensionalize a commercial simulation package, the model is based on mathematical equations that can always be nondimensionalized. Nondimensionalization carries with it a number of benefits; in particular the following:

- Nondimensionalization reduces the governing equations to their most basic functional form, which helps to clarify the expected behavior.
- Examination of the nondimensional form of the governing equation is the easiest and most certain way to identify which parts of an equation are relevant and which parts may be neglected. In this way, simplifications of the original equations may often be made (and, equally important, justified).
- Nondimensional plots of model output are the most compact way of presenting the model results. Admittedly, your target audience may not be equipped to understand dimensionless results; in this case, it is up to the modeler to either present dimensional results or educate their audience regarding the nondimensional ones.
- The nondimensionalization process results in the identification of the controlling dimensionless parameters. These parameters control the behavior of the equation, allowing the modeler to readily identify parameter ranges over which behavioral changes will take place. Furthermore, the dimensionless parameters show the modeler which dimensional parameters can be uniquely identified and which cannot.

1.4 Conclusions

As was stated in Section 1.1, following the rules and suggestions laid out in this chapter won't guarantee success, nor will it make you a modeler (only time and experience will do that). Hopefully, however, the ideas presented here will help you avoid some of the most common traps that inexperienced modelers tend to fall into. In the following chapters, I will develop a number of models; as you follow these developments, watch for the application of these basic principles. Ask yourself how you could apply these principles to your own problems. As with any creative endeavor, there are rules and guidelines that can be applied, but the ultimate responsibility for the final product belongs with the artist.

CHAPTER 2
A model of exponential decay

Chapter summary

Possibly, the most basic building block for mathematical models of physical systems is the process of exponential growth and decay; models of exponential decay processes are encountered in every branch of science and engineering. Here, we examine the problem of modeling discharge from a high alpine basin, which leads to the representation of basin discharge as an exponential decay process. Although simple, the model is quite general and may apply to many springs and streams, as well as to the electrical charge stored in a capacitor, the temperature of a heated piece of steel quenched in cold water, and the decay of radioactive materials.

2.1 Exponential decay

Exponential decay processes are ubiquitous in nature. The essence of exponential decay is that the rate at which some quantity is lost is determined by the amount of the quantity that remains behind. We are all familiar with this process on an intuitive level; for example, a can of soda placed in the refrigerator contains quite a bit of heat (relative to the temperature of the refrigerator). The temperature gradient between the refrigerator and the soda is large at early times; thus, the soda at first cools rapidly. However, as time goes on, the difference in temperature between the soda and the refrigerator lessens; as a result, the rate of cooling also decreases. Theoretically, the soda will never actually reach the temperature of the refrigerator, but it will come arbitrarily close over a long period of time. We say that the temperature of the soda decays exponentially to the temperature of the refrigerator.

Much of the behavior observed in the natural world can be expressed in terms of differential equations—equations that show relationships between the rates of change in different quantities—and, because they possess the property of being their own derivatives, the solutions to differential equations very commonly involve exponential functions. Thus, it should be no surprise to find that exponential functions are encountered widely in natural phenomena.

Models and Modeling: An Introduction for Earth and Environmental Scientists, First Edition. Jerry P. Fairley.

Companion website: www.wiley.com/go/Fairley/Models

Although exponential decay models are common to many disciplines, each discipline refers to them by a different name. In heat transfer, models of exponential decay are called “lumped capacitance” models; similarly, electrical engineers use the term “RC models” (for “resistance–capacitance”). Hydrogeologists, who use exponential decay models to represent spring flow, basin discharge, and similar processes, often refer to such models as “baseflow recession models.” In the following sections, we will develop a simple exponential decay model for basin discharge, using as an example a hypothetical watershed in central Idaho.

2.2 The Bandurraga Basin, Idaho

Suppose you are approached by the US Forest Service regarding the management of water yield from an alpine basin in central Idaho. The basin receives a variable amount of precipitation in the form of rain and snow each year; in general, there is little or no precipitation after April. As one would expect, discharge in the stream that drains the basin (the “main stem”) peaks about the time precipitation in the basin ceases, and stream flow decreases over time until precipitation resumes in the fall. Water from the basin is captured in a down-valley reservoir and used to operate a micro-hydroelectric generator system that provides power for the Forest Service wilderness visitor center. The water manager’s problem is this: some years there is an overabundance of discharge from the basin, and the excessive runoff can damage the reservoir; in other years, the water manager drains (draws down) the reservoir in early spring, making room in the reservoir to store the basin yield, but there is not always enough water to refill the reservoir. When spring and summer flows are too low to refill the reservoir, the visitor center can experience power shortages in late summer. The water manager’s question is, can you help?

2.3 Getting organized

Most modelers are familiar with this kind of request: can you help? Can you make a model of this? Whenever you receive a request of this type, the first thing you should do is work to understand the exact nature of the problem and what is expected of you (i.e., Section 1.2.1). Requests for models of physical systems are rarely well-thought-out, at least in part because few non-modelers really understand the craft of modeling. Regardless of whether or not the client is clear in her/his description of the problem and desired objectives of the model, you can be certain that you, as the modeler, will bear the brunt of the client’s unhappiness if your deliverable does not meet the client’s (often unstated) expectations. As a result, it is up to you to discuss the problem with the client until you clearly understand what is required; furthermore, you must share your understanding with the client, to be certain that the client knows what to expect when s/he receives your final report.

2.3.1 Observable quantities

This is probably a good point to pause and reflect on what I call "observable quantities." When formulating your model objectives, it is important to keep in mind what can be observed and what cannot. A model that makes predictions that cannot be checked is rarely useful; as a result, the quantity you select as your model output must, at least in principle, be amenable to measurement. In this example, there would be very little point to making predictions of average saturation in the basin, or even of the quantity of water stored in the basin, because there is no way to measure an average basin saturation or independently verify the amount of water in the basin at any given time. Furthermore, it is not clear how the amount of water in the basin (or the saturation) can be related to the quantity of discharge from the basin. In a sort of "worst case," the Forest Service's water manager may equate the "total water stored in the basin" with the amount of water available for power generation. This will almost certainly translate into a resource shortfall, and the modeler will bear the responsibility for such a shortfall because s/he provided misleading model predictions.

2.3.2 Stating the model objective

On the basis of the foregoing discussions (Sections 2.3.1 and 1.2.1), it is clear we must formulate our model objective with consideration of our observable quantity (in this case, the discharge from the basin) and with other expectations we have for the solution (i.e., to what use will we put the model output?). In this case, I propose a good expression for that which is to be found may be the following:

FIND: An expression for discharge from the Bandurraga Basin, Idaho, at any time following the cessation of precipitation in an annual cycle. The expression must allow the calculation of cumulative discharge between any two specified points in time (e.g., from May 1 to August 1).

Note that we have stated exactly what the model output should be (the basin discharge as a function of time) and, additionally, we have stated the expectation that we will be able to integrate the solution between any two points in time (either analytically or numerically). The first of these objectives tells us we don't necessarily need to know the spatially distributed head in the basin, the temperature of the discharge, the dissolved load in the main stem, the amount of power that can be generated from the water in the basin, or similar extraneous information. The second part of the statement of model objectives starts us thinking about what kinds of models might or might not be appropriate; for example, we probably don't want to devise a statistical model that gives the mean discharge of all previous years on day X, because such a model can't give predictions of cumulative discharge specific to the current water year (although it would certainly be possible to devise a probabilistic model for this situation).

2.3.3 What data are available?

Once we have a clear statement of our model objective, the next step is to marshal our resources in terms of how much we know about the situation. What data are available? What facts about the model domain are known to us? In this case, we will assume we have the type of information we could reasonably expect to be available for a wilderness basin under Forest Service management.

KNOWN:

- Discharge from the basin at some initial time (say May 1)
- Basin area, topography, and geometry
- Type of vegetative cover in the basin
- Twenty-year record of precipitation, discharge, and temperatures in the basin
- Basin geology (rock types, geological structures, faults, etc.)
- Information on soil types and thicknesses

It is likely that not all of this information will be useful to us, but it isn't always clear in advance what will be and what will not be useful. Furthermore, the process of thinking through what is known about a system is a crucial part of developing an overall understanding of the problem and of conceptualizing an approach to its solution.

2.3.4 What can we assume?

The next step in conceptualizing an approach to model development is to begin a list of our model assumptions. This list of assumptions will change as we formulate the model; it may grow or shrink as we relax some of our initial assumptions. Regardless of the way the list evolves over time, *you must keep careful track of your modeling assumptions!* You will need to know your model assumptions for your final report (or journal article, thesis, etc.), and it will be helpful to have them all gathered together in one place.

There are many possible assumptions that could be made at this point in the development of our model; for example, we could assume the basin comprises homogeneous and isotropic geological materials (this is a common assumption of groundwater models), or that precipitation in the basin is equal to the average of the precipitation over the past 20 years. In general, however, it is best to make a minimal number of assumptions at the start, and then to impose additional assumptions as needed during model development (while keeping careful track as new assumptions arise). In this case, we can (at least initially) make do with a very short list of assumptions.