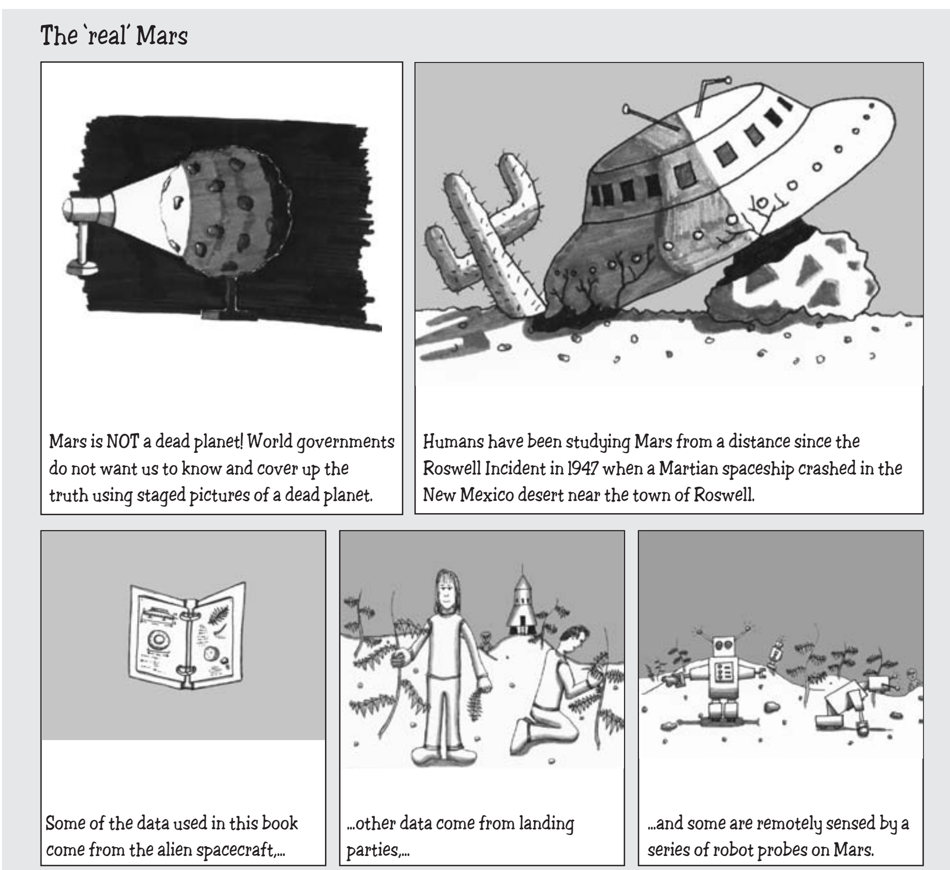
1 Introduction

1.1 What if there was life on Mars?

Mars is a dead planet. All of the information we have from remotely sensed data and, more recently, from landing craft, show that, even if life did ever exist on Mars, it does not now. But was not the Mars of 1960s B-movies more intriguing? Imagine a Mars that is very much like Earth, teeming with life and facing similar environmental and ecological problems to those we face here. The setting for this introduction to environmental and



ecological modelling is such a Mars. As you use the book and solve the problems, the story of the environmental and ecological problems faced on Mars will unfold, given in a storyline that appears through the text as picture boxes. The picture boxes not only provide the examples for which models will be developed, tested and applied, but they also act as markers to help you skip through the book to find specific subjects in the main text.

Through the picture boxes, we will present a number of environmental and ecological problems faced on this imaginary version of Mars, which could be addressed using models. You will be given the results of previous modelling attempts and you will discover why those models failed. You will then be given the opportunity to develop new approaches to do a better job. On the way, you will learn how to model and how to apply good modelling practice. The book uses guided problem solving to develop modelling skills. It will allow you to learn the essential philosophies and techniques of environmental and ecological modelling in a scientifically rigorous way by learning and applying these techniques to solve problems on Mars. All of the skills you learn from using this book will be applicable to real-world problems—indeed these techniques were developed on Earth!

Mars needs YOU!

Previous attempts to model the environment on Mars have failed. Your job would be to lead the team of exobiologists (people who study life on other planets) to develop models to help address the environmental and ecological problems on Mars. Good luck at the interview!

1.2 What is a model?

A model is a simplified representation of reality; this definition applies equally to a physical model, such as might be sculpted in clay, and a mathematical model, constructed using mathematical equations (Fig. 1.1).

When creating a model, a sculptor will observe reality, decide which aspects of reality are important, and use a selection of techniques to represent these important aspects; this representation is the sculpted model. A mathematical modeller follows exactly the same procedure; the only difference between the mathematical modeller and the sculptor is in the methods they use. Like the sculptor, a mathematical modeller observes reality, but this is usually done by some form of experimentation, not just by eye. The mathematical modeller decides which aspects of reality are important to the representation. Section 1.4 introduces a systematic procedure for selecting the aspects that are important, in which the research question is translated into a hypothesis, which then guides the modeller as to the type of model needed, and the important characteristics of the model. As

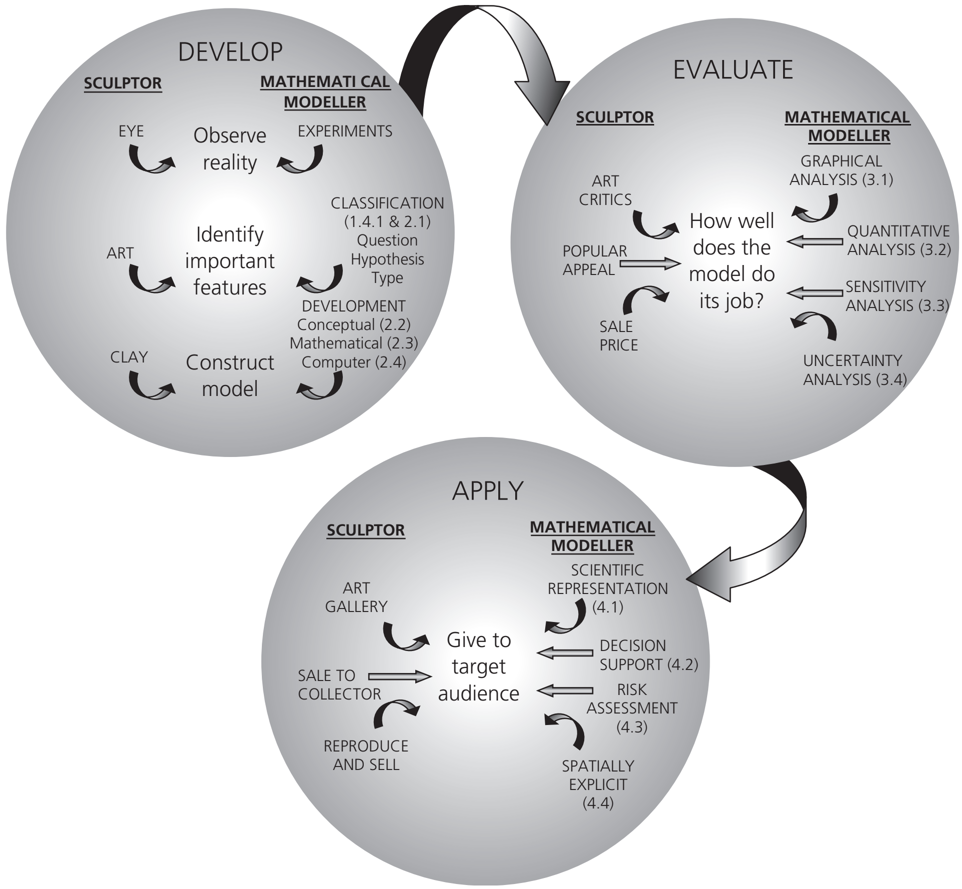
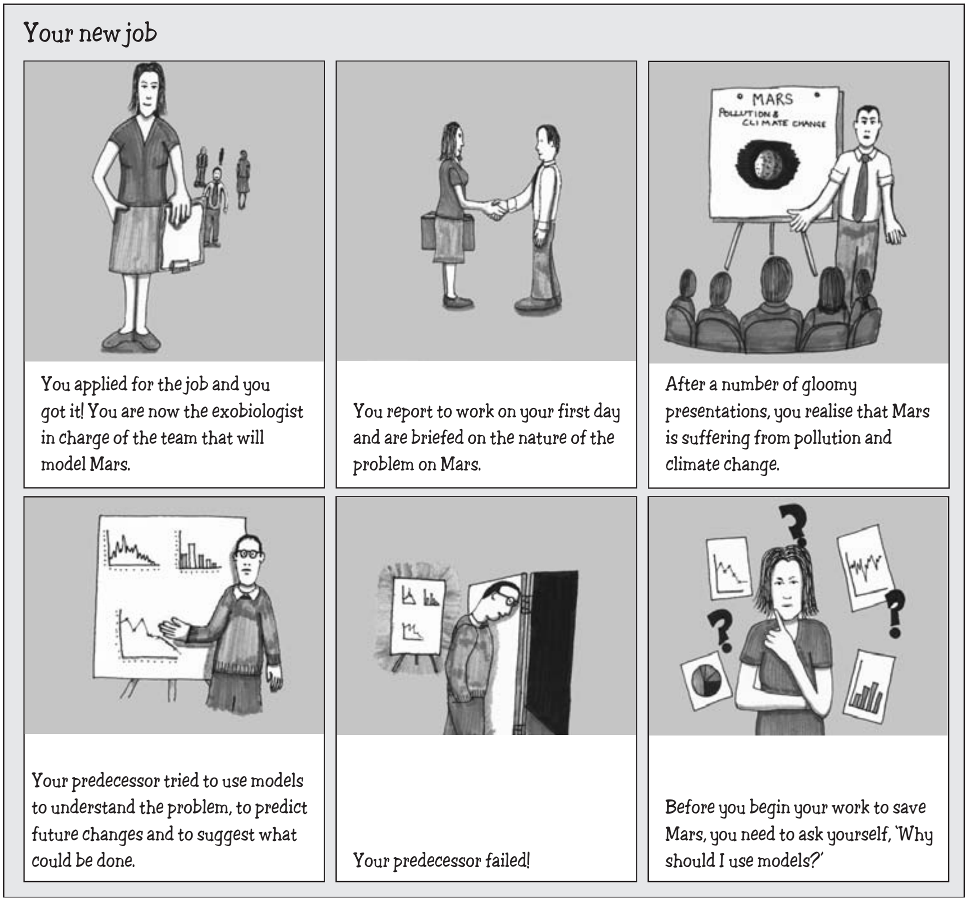


Figure 1.1 How to make a model: a simplified representation of reality.

will be discussed in Chapter 2, the mathematical modeller then uses a combination of mathematical techniques and computer programming to represent these important aspects of reality as equations; this representation is then the mathematical model.

The value of a model lies in its ability to do what it was created to do. Critics may evaluate a classical sculpture according to how much it looks like the subject, whereas an impressionist sculpture is evaluated on an entirely different basis. Similarly, a mathematical model should be evaluated with respect to its ability to achieve its objectives. If it is a quantitative model, the evaluation should also be quantitative. Different aspects of model evaluation are discussed further in Chapter 3.

The value of the model can only be fully realised if it is made accessible to its target audience. A sculptor might do this by exhibiting the model in an art gallery, selling it to a collector, or making reproductions to be more widely sold. Due to the ease of exact reproduction provided by computers, mathematical models are usually packaged in some form that makes them easy to use, and to be reproduced and distributed to potential users. Some of the trials and tribulations of applying mathematical models in this way are described in Chapter 4.



1.3 Why use models?

Many people regard modelling as a combination of difficult mathematics and technically demanding computing. People who spend too much time with models are often classified by others as '**modellers**' and are considered to be rather strange. However, modelling is a ubiquitous tool that is widely used in all branches of science, although it is often known by a different name. One of the most elegant models in history (described as a 'table') was developed by Mendeleyev to describe the electronic structure of elements. The model is, of course, the periodic table, and was used to predict the properties of many elements before they had even been discovered. Without this model, the discovery of many elements would have been delayed for many more years.

In fact, modelling is something that we all do every day. Take the example of choosing a queue at the railway station. We look at the length of the queue, and assess how quickly each queue is moving; in the blink of an eye, we have constructed a model of the time it will take to get to the counter, entered the input data (the relative speed at which the two queues are moving), and obtained the result upon which we choose the best queue to join. In doing this, we have mastered the main stages of model conception and development. These instinctive thought processes are disentangled when we look at model development in Chapter 2.

Many of us even go on to evaluate the model, by noticing which person is at the back of a neighbouring queue, and checking that we do in fact reach the front before them! This evaluation process is rationalised in Chapter 3. Modelling is something that comes naturally to us all, but this is precisely the reason why it is so often misused and misrepresented. Issues surrounding application (and misapplication) of models are discussed throughout this book and are the focus of Chapter 4.

If we are to make proper use of our inherent capability to model, we must have an unambiguous idea of what the model is intended to do (i.e., the **scope** of the model), follow procedures for model development, evaluation and application, and understand that the natural world is unpredictable. After all, who could have predicted that the family of five at the front of the queue would all need new rail passes?

Translating the model from the assessment in your head, to the back of an envelope, to a spreadsheet, to a computer program, allows us to use the model to do science. Science starts with a hypothesis: a hypothesis is a simple model (see Section 1.4.1). An unambiguously stated hypothesis can be tested, allowing a comparison between two alternative theories. A numerically precise hypothesis can be quantified, allowing observations to be explained and future events to be predicted. Collecting hypotheses together allows their combined effects to be understood, and real environmental and ecological systems to be modelled. Rather than asking why we should use models, we should be asking why not!

Having started the modelling process by testing a qualitative hypothesis, it is a small step to make the hypothesis unambiguous and quantitative (see Section 1.4.1), adding value to our understanding. By measuring the time needed to serve customers at the railway station over a period of an hour, unambiguous quantitative hypotheses can be written to describe the movement of the queues. Consider the following example.

'On average, the time needed to serve a customer is:
a seconds for a male customer aged less than 21;
b seconds for a male customer aged from 21 to 65;
c seconds for a male customer aged over 65;
d seconds for a female customer aged less than 21;
e seconds for a female customer aged from 21 to 65;
f seconds for a female customer aged over 65.'

These hypotheses can be tested for this station by measuring the time needed to serve customers for a further period of time. If the new measurements are close to the hypotheses, I might decide to use them to predict which queue I should choose. This will need input information about the number of customers in the different categories in each queue. By doing this, I have written unambiguous and quantitative hypotheses, and I have quantified the expected result; this is modelling (although by the time I have done all this I would have missed my train and would possibly have been arrested!).

All of these measurements were taken at a railway station in London. When I am at a railway station in Paris, I bump into another modeller who has developed a different model for choosing which queue to join. His model is based on the following hypotheses.

'On average, the time needed to serve a customer is:
x seconds before 9.30 am;
y seconds between 9.30 am and 4.30 pm;
z seconds after 4.30 pm.'

We might use modelling to compare the two different theories (i.e., the time to serve the customer depends on the type of customer, or the time to serve the customer depends on the time of day). To settle our argument, further measurements should be taken at different railway stations both in London and Paris. The measurements should include the time of day, the number of customers in the different categories in each queue, and the time taken to reach the front of the queue. Examination of the measurements will put an end to all the arguments, and determine which model, if any, is likely to correctly predict the best queue to join.

Having collected together all the data needed to test the hypotheses, a small additional effort could give us so much more. We can use statistical analysis to determine any relationships that exist between the type of customer, the time of day and the time needed to serve the customer. We can determine the range of possible times for each queue, and so quantify the likelihood that our predictions are correct. Having quantified these relationships, we can try to understand the processes contributing to the observed result. If we understand the processes, we can extrapolate to railway stations in other countries: do the results hold equally well in England and France, and are they likely to be very different in North America? If we have established a good rule of thumb, I can tell my friends how to choose the best queue at any railway station around the world!

This example illustrates many capabilities that are common to most modelling activities.

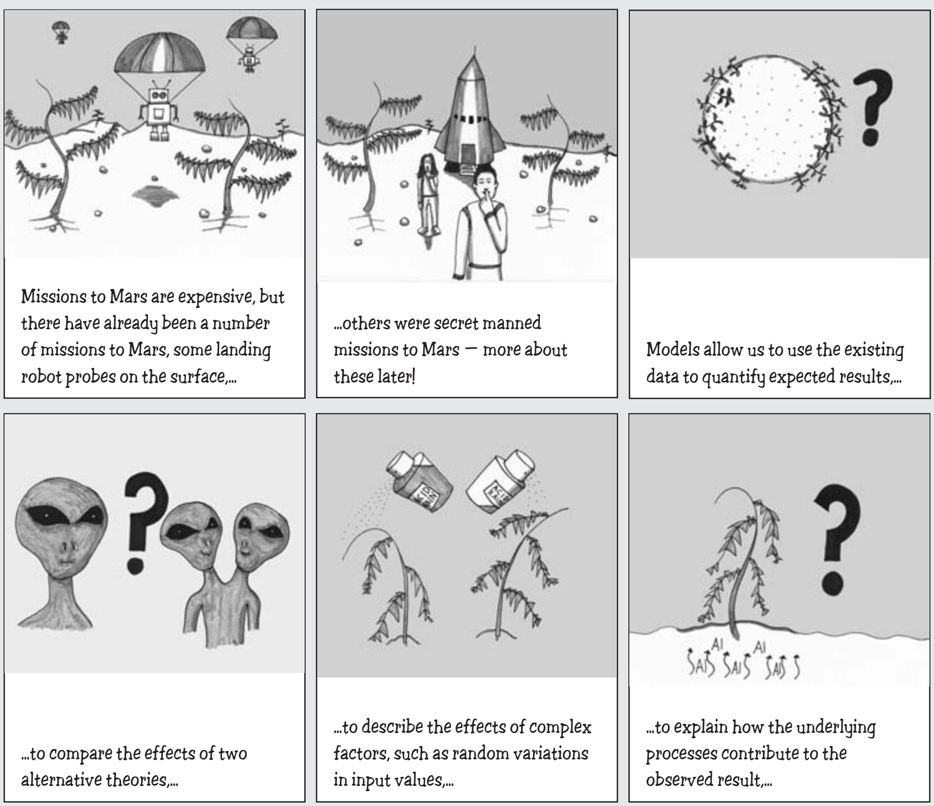
In general, modelling has the potential to:

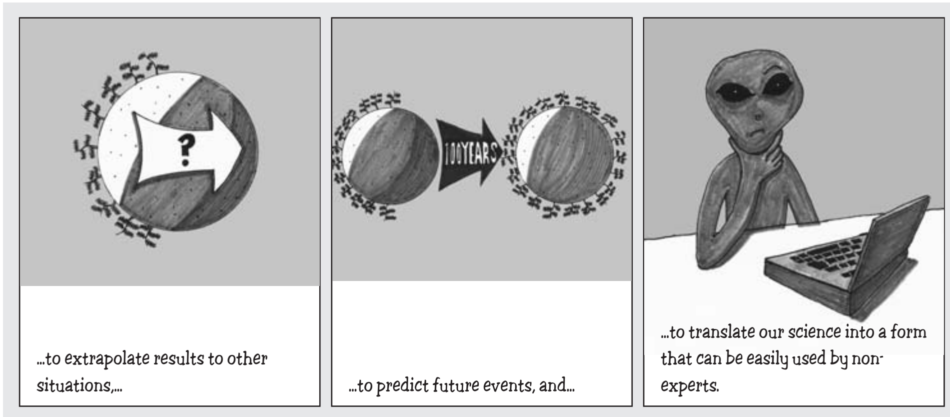
1. quantify expected results;
    - For example, given recent weather trends, do we expect sufficient rain to fall over the summer to keep reservoirs topped up, or should water companies be thinking about hosepipe bans?
2. compare the effects of two alternative theories;
    - For example, is organic food really better for our health, or is there no health benefit over non-organic foods?
3. describe the effects of complex factors, such as random variations in inputs;
    - For example, how does uncertainty in carbon dioxide emissions affect predictions of climate change?

4. explain how the underlying processes contribute to the observed result;
   - For example, describe how changes in the number of ragworms living in an estuary can cause changes in the populations of other organisms through complex food web interactions.
5. extrapolate results to other situations;
   - For example, what can the spread of BSE within the UK cattle herds tell us about the potential spread of BSE in France?
6. predict future events;
   - For example, in 2050 what will the global human population be?
7. translate our science into a form that can be easily used by non-experts.
   - For example, weather forecasts allow us all to make use of complex meteorological science.

All of these accomplishments are just a few easy steps from the starting point of the hypothesis.

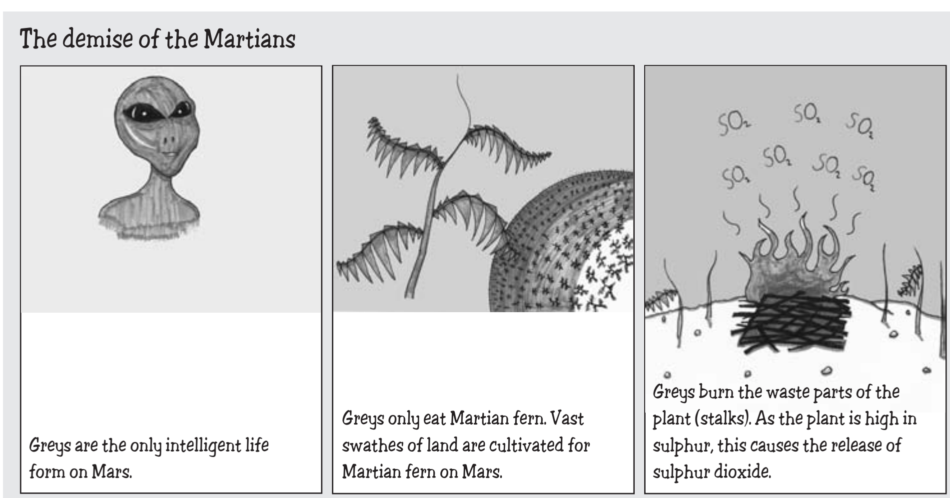
Why use models on Mars?

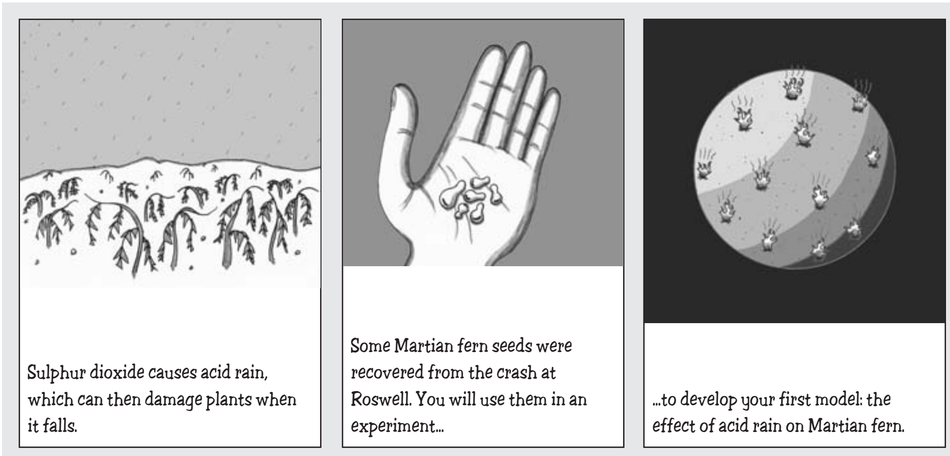




1.4 Which model should I use?

Descriptions of mathematical models in the scientific literature often use a large number of different terms. It may appear that the only purpose of this jargon is to confuse. However, these terms can have one of two important purposes: to determine the type of model or to describe the type of mathematics to be used. Models and mathematics are often confused or equated to each other; but they are different things. The type of model tells us what the model does, that is, the sort of inputs used and outputs produced, the limitations of its application and the way it is used. The type of mathematics tells us how the model does that, that is, how it translates the inputs into the outputs. In the next two sections we attempt to demystify the jargon, using an example of a model of Martian fern to explain the purpose and meaning of some of the more commonly used terms. The importance of these terms will become clear throughout the book as we show how the type of model and mathematics needed can be used to guide the development process.





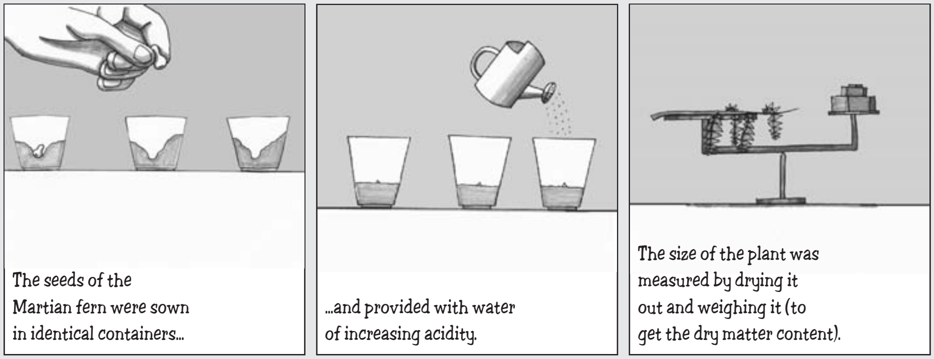
1.4.1 Determining what type of model to use

When we make efforts in modelling that go beyond the simple starting hypothesis, we need to define the type of results expected, and the type of model needed to achieve those results. If I want to tighten up a screw, I must first choose the screwdriver that fits the screw I am trying to tighten; I cannot use a flat-head screwdriver on a cross-head screw. In the same way, if I want a model to do a particular task, I must first choose the model that is appropriate to achieve that task. The classification of screwdrivers, by the shape and the size of the head, is extremely helpful when choosing the right screwdriver for the job. The classification of models, if it focuses on what the model can *do*, can also be extremely helpful when choosing the right type of model for the job.

The classification of models allows us to decide what type of model we need, and helps us to search for any models that have already been developed. There are many different ways of classifying models. In this discussion, we have started at the beginning, with the hypothesis. By examining the characteristics of the underlying hypotheses, all of the major features of models can be characterised, long before we start adding any numbers or equations.

Models can be based on anything from a simple hypothesis to a complex collection of hypotheses, but even the simplest hypothesis includes all the characteristics needed to classify a model. As an example, consider an experiment to study the effect of water acidity on the size of Martian fern.

Setting up a Martian fern pot experiment



First, we carefully and exactly state the question to be addressed by the model. From the question addressed we can develop the **null hypothesis** and its alternatives. These will form the basis of our model.

The **null hypothesis** is a negative statement of the thing you want to test.
The **alternative hypotheses** list all possible alternatives to the null hypothesis.

The null hypothesis, referred to as H_0, for the research question 'Does the acidity of water have any effect on the size of Martian fern?' is

'The acidity of water has no effect on the size of Martian fern.'

The alternative hypotheses (H_1 and H_2) can then be formulated as follows:

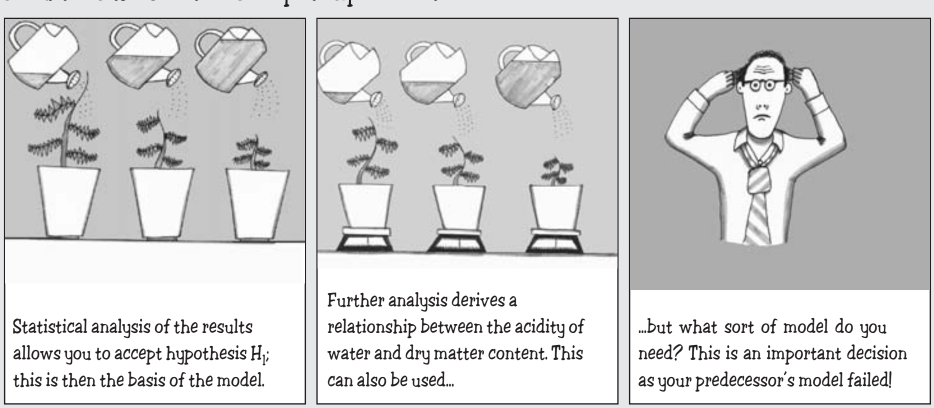
H_1: *'The size of Martian fern decreases with the acidity of water.'*

and

H_2: *'The size of Martian fern increases with the acidity of water.'*

The model is then constructed from the hypothesis that is believed, from experimental observations, to be true.

Results of the Martian fern pot experiment



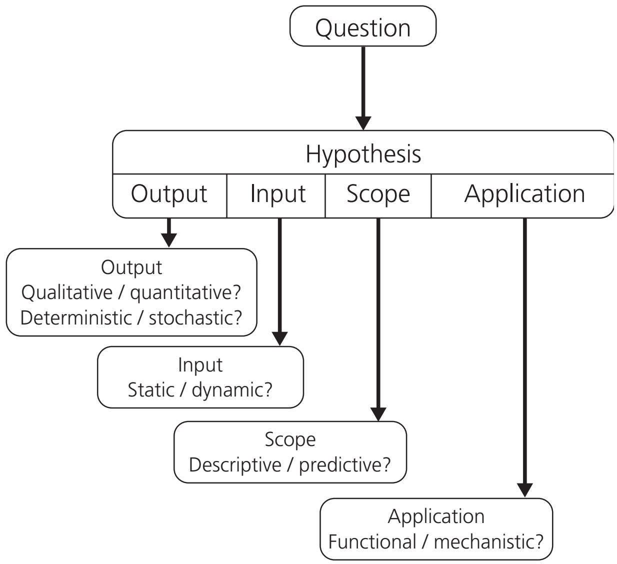


Figure 1.2 Classification of a model according to its underlying hypotheses.

The type of model that we should use can be defined by the **outputs**, **inputs**, **scope** and **application** of the hypotheses that it is based on (Fig. 1.2). The model will then be defined in the same way as the hypotheses.

The output from the above hypothesis describes the size of the Martian fern. The input is the pH and it describes the acidity of the water. The scope defines whether the model is to be used to explain the results of the current experiment, or whether it will be used to extrapolate to new situations. The application tells us what the model will be used for.

Outputs—information produced by the model.
Inputs—information needed by the model to run.
Scope—can the model be used outside the experiment used to develop it?
Application—is the model used to explain processes?

The **outputs** from the hypothesis could be **qualitative** or **quantitative**. A qualitative value describes the nature of the output, whereas a quantitative value will provide a numerical measurement or count. If we were interested in the more extreme effect, where acidic water kills the Martian fern, we might only need to know whether the fern was alive or dead (a qualitative measure):

H_1: *'The acidity of the water determines whether Martian fern is alive or dead'*

(Output: qualitative).

The input measurement of water acidity is quantitative, but the tested status of the fern is qualitative, that is, alive or dead. If a model provides any quantitative output, it is usually described as quantitative. A model based on the above hypothesis provides no quantitative output and so would be qualitative.

By contrast, a model based on the following hypothesis:

H_1: *'The size of Martian fern plants will decrease by 15% with each decrease in pH unit'*

(Output: quantitative)

would be a quantitative model.

A quantitative output variable can be given as a specific (or **deterministic**) value, or it can be given as a range, specifying the probability that the result falls within the given range. The quantitative model above would output the deterministic result of 15% for each decrease in pH unit. Models working with ranges of values are termed **stochastic**. If we were weighing up alternative risks, we might need to know the chances of acid rain killing the Martian fern, and so we would need to use a stochastic model. The above hypothesis can be rewritten for application in a stochastic model as follows:

H_1: *'There is 95% probability that the size of Martian fern plants will decrease by 10–20% with each decrease in pH unit'*

(Output: quantitative stochastic).

The stochastic result of the above model would be 10–20% for each decrease in pH unit.

The **inputs** to a model can be fixed values for each run of the model, or they can change over a series of measurements. The input measurement of the acidity of water does not change, and can be described as a **static** variable. However, we might be interested in how the size of Martian fern is affected by the acidity of water at different times from the start of the experiment. In this case, the inputs will change over a series of measurements taken at different times, and can be described as **dynamic** with respect to time. The hypothesis should be rewritten as follows to reflect our interest in time:

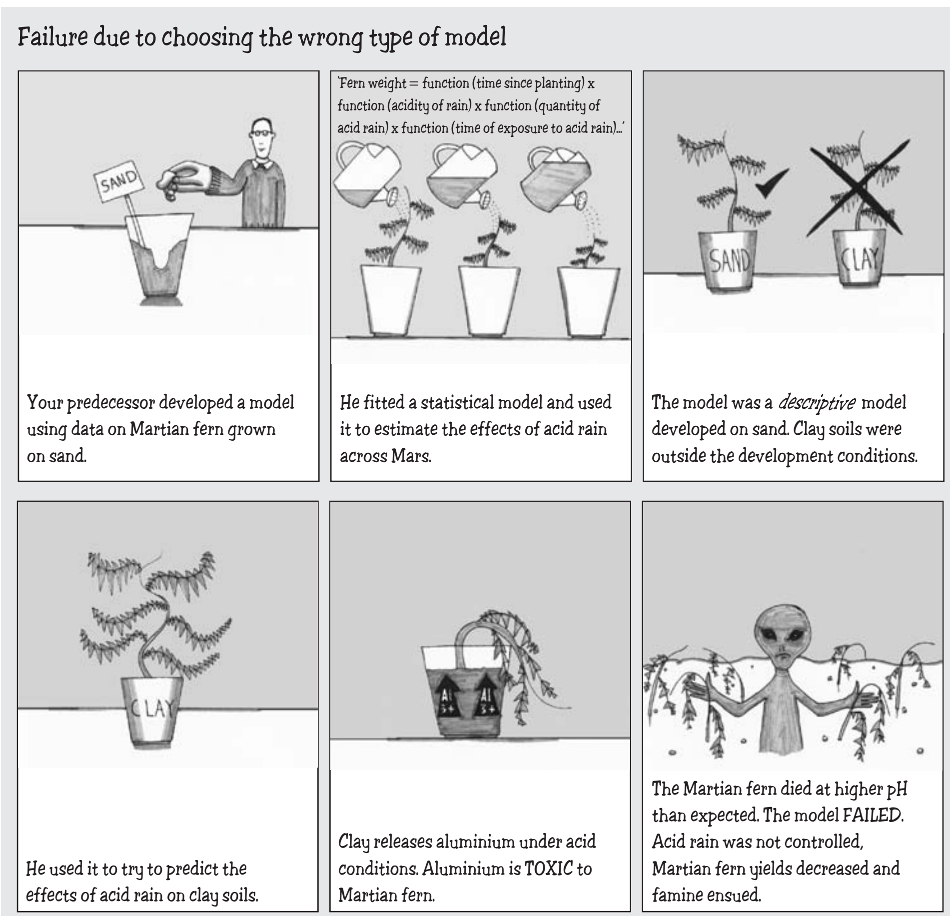
H_1: *'The size of Martian fern plants will decrease by 2% every day with each decrease in pH unit'*

(Input: dynamic with respect to time).

If a model uses any inputs that are dynamic, it is classified as a dynamic model. A model can be dynamic with respect to any of its inputs, such as successive changes in pH, but, in practice, the term 'dynamic' usually refers to changes with respect to time. Only if the model uses static input variables alone, will it be classified as static.

The **scope** of the hypothesis can be termed **descriptive** or **predictive**. A descriptive model is used to describe observations within the conditions of the current experiment. A predictive model is used to extrapolate beyond the scope of the experiment, and provides results that extend beyond the current observations. Models may be predictive with respect to time, space, species or any other input variable. If we wanted to use the above model to predict the effect of a higher acidity than included in the experiment, the model would be predictive with respect to acidity. A predictive model usually requires some degree of understanding of the processes causing change, as an understanding of the processes improves our ability to extrapolate to new situations.

The **application** can be **functional** or **mechanistic**. If the purpose of the hypothesis is to explain the underlying processes responsible for the overall result, the hypothesis can be described as mechanistic. If the hypothesis merely aims to represent or predict the experimental observations, the hypothesis will be described as functional. The inputs needed to drive a functional model are usually less complex than those needed to drive a mechanistic model. The need for mechanistic understanding in predictive models can result in confusion regarding the terms ‘functional’ and ‘mechanistic’. A model that has been constructed around a mechanistic understanding of the underlying processes so that it can be used to predict, but that is used only to represent and predict the experimental observations, should be termed **functional**. This model should only be termed **mechanistic** if we are interested in using the model to find out what is happening and explain the underlying processes.



SELF-CHECK QUESTIONS: WHAT TYPE OF MODEL SHOULD YOU USE?

1. Q: Which term best describes each of the following?
   a. Whether the model is predictive or descriptive?
      i. Inputs
      ii. Outputs
      iii. Scope
      iv. Application

   [A: iii. Scope]

   b. The values used by the model to predict the outcome?
      i. Inputs
      ii. Outputs
      iii. Scope
      iv. Application

   [A: i. Inputs]

   c. The type of relationships used in the model to explain observed results?
      i. Inputs
      ii. Outputs
      iii. Scope
      iv. Application

   [A: iv. Application]

2. You want to know whether a Martian fern is alive or dead at the end of an experiment in which it is exposed to acid water.

   Q: Is the output from the model qualitative or quantitative?

   [A: Qualitative—it is either alive or dead.]

3. You want to know by how much growth is reduced when the Martian fern is exposed to acid water.

   Q: Is the output from the model qualitative or quantitative?

   [A: Quantitative—the output should show the relative reduction in growth of plants grown using acid water compared to plants grown in neutral water.]

4. You want to know by how much each day the size of Martian fern plants decreases when exposed to acid water.

   Q: How would you describe the input variable, the acid water added each day?

   [A: Dynamic with respect to time.]

5. You use a statistical model to describe your results on the impact of acidity on Martian fern.

   Q: Is the scope of the model descriptive or predictive?

   [A: Descriptive—used within the same experiment from which the model was derived.]

6. You develop a model that simulates the effect of acidity on the growth of Martian fern using your results, which you then apply to fern growing in the wild.

Q: Is the scope of the model descriptive or predictive?

[A: Predictive—used outside the original data used to develop the model.]

7. You want to develop a model to predict what will happen to Martian fern under different possible future concentrations of acid rain. You develop the model from the relationships you found in an experiment where plant growth was measured under different acidities.

   Q: Should the model application be described as functional or mechanistic?

   [A: Functional—as it does not simulate how acidity affects plant growth.]

1.4.2 Determining what type of mathematics to use in the model

In Section 1.4.1, we described how the type of model is defined by the nature of the hypotheses that it is constructed from. The purpose of this section is to illustrate how the same question can be answered using many different types of mathematics. Many different mathematical approaches can be used to explore, explain and model the same data, and this adds more to the jargon of modelling and can be very confusing. However, the important distinctions between models do not lie in the type of mathematics used, but in the purpose for which the model was developed, that is, the type of model as described in Section 1.4.1.

The mathematics chosen are no more than the materials used to build the model. It is like choosing to build a house out of bricks, sticks or straw; different materials are more suitable for different purposes, but the important characteristic of the house is not the material it is made from, but rather the purpose it can be put to. The important characteristic of a straw house is that it is quickly-built temporary accommodation, not that it is made of straw. The important characteristic of a brick house is that it is strong and will not blow down, not that it is made of bricks.

However, because we know brick houses tend to be strong and do not blow away, a house is often described by the material it is made from rather than by its purpose. The same applies to models, so you will often hear a model being described by the mathematics it is constructed from. A model might be described as statistical, geostatistical, Bayesian, a neural network, cellular automata, process-based; the list goes on and on! As with the brick house, these terms become important only because we know something about the characteristics of models built using the different mathematical approaches. The characteristics of the approaches and how you choose an approach to develop your model will be discussed further in Chapter 2. Here, we discuss how the jargon describing the mathematical approach relates to the type of model and so helps to further describe its purpose. Do not worry about exactly how you should use each type of approach. Instead, focus on what the approach *does*, as it is this that allows you to assess what type of model can use this approach.

Many different mathematical approaches could be used to address the general research question ‘How does acid rain on Mars affect Martian fern?’ This question could be asked in many different ways; it is the exact form of the research question that determines the nature of the underlying hypotheses, the type of model associated with the hypotheses and the choice of mathematical approach.

If the question is 'How does acid rain on Mars affect the size of Martian fern?', then the *size* of Martian fern must be related to the pH of the rain water.

As already discussed in Section 1.4.1, this can be done through a hypothesis such as:

H_1: *'The size of Martian fern plants will decrease by 15% with each decrease in pH unit of the rain water at a chosen site.'*

The model developed from this hypothesis would be

- quantitative (because the output, *size*, is a quantitative measure),
- deterministic (because we are not specifically interested in risks so the simpler single-value result—*15% with each decrease in pH unit*—is sufficient),
- functional (because the model must describe *what* the effects of decreasing pH are, not how decreasing pH has this effect) and
- descriptive (because we do not intend to extend the results beyond the conditions of the original experiments).

As the model is descriptive and functional, there is no need for process-based equations; a **statistical** model is a good approach. A statistical model derives the mathematical equations used in the model directly from measured data using standard statistical techniques. The data could, for example, be analysed using an analysis of variance (ANOVA), as found in most statistical software packages. This is one type of mathematics that can be used to translate the hypothesis into a working model. When a model is described as statistical, you immediately know that it contains no understanding of the processes involved, so it is unlikely to be accurate if used to predict results well outside the conditions for which it was developed. Furthermore, you know that it can provide no mechanistic understanding of the observed responses because this information is not included in the statistical relationships derived from the data. The term 'statistical' immediately tells us that the model purpose *should* be only functional and descriptive.

If, however, you are interested in how the Martian fern is **distributed** with respect to the pH of the rain across Mars, the question would be reworded as 'How does the distribution of acid rain on Mars affect the distribution of Martian fern?' The hypothesis on which the model is based remains unchanged, but the research question dictates that the model now requires a spatial component to describe the spatial distribution. One set of mathematical approaches that can be used to describe the spatial distribution is termed **geostatistical**. To develop a geostatistical model, measurements of acid rain and Martian fern are collected at a series of points on a map. As with a statistical approach, the geostatistical procedures are used to fit the equations that determine the occurrence of Martian fern with respect to the pH of the rain water. These equations then form the body of the model. If a model is described as geostatistical, you know, as for the statistical approach, that it should only be used for functional and descriptive applications, but that it can also handle information about the spatial distribution of input data and provides a spatial distribution of results.

Alternatively, it might be the **risks** to Martian fern of the acid rain on Mars that are of most interest. In this case, the question becomes ‘What are the risks to Martian fern of acid rain on Mars?’ The hypothesis must be changed to express the risks associated with acid rain, for example:

H_1: *‘The size of Martian fern plants will decrease by 10–20% with each decrease in pH unit of the rain water at a chosen site.’*

The model constructed from a hypothesis such as this will be quantitative, functional and descriptive (as before), but the outputs will now be stochastic, providing an idea of the risks associated with acid rain (a decrease of 10–20% with each decrease in pH unit). There are many different mathematical approaches for constructing such stochastic models.

Bayesian statistics incorporate prior knowledge and accumulated experience into probability calculations, based, for example, on the previous year’s observations. Bayesian statistics could be used to address questions of risk. To do this, you would need to collect, in the previous year, measurements of the size of Martian fern and the pH of the rain water. From this information, the range of the likely decrease in Martian fern with the pH of the rain water can be ascertained. If a model is described as Bayesian, you know that it contains no understanding of the processes involved, and so should be used as a functional model, but that it can be used to predict risk.

A **neural network** has a similar function, and can be trained on the first year’s data. A neural network is a piece of software that is ‘trained’ by presenting it with examples of input and the corresponding desired output. Neural networks mimic the vertebrate central nervous systems, to develop rules about relationships between inputs and outputs such as the decline of Martian fern with the increasing acidity of the rain. To use a neural network to assess the risks to Martian fern of decreases in the pH of rain water, measurements of Martian fern and rain water pH should be collected for a number of years and used to train the neural network. The neural network can then provide a very accurate representation of how the size of Martian fern has changed with changes in pH. A model constructed using a neural network is functional, but if sufficient data have been used to train then the model can be used to predict.

If, on the other hand, you know the source of acid rain, and want to know *how* the movement of acid rain from the point source affects the growth of Martian fern, then the question could be restated as ‘How does the movement of acid rain from the point source affect the size of Martian fern?’ The same hypotheses could be used to describe the effect of acid rain on Martian fern, but other hypotheses would be required to describe the movement of acid rain from the point source, such as the following:

H_1: *‘The movement of acid rain is given by the direction and speed of the wind.’*

A number of different approaches exist for describing spatial changes in variables and could be used to construct a model of the movement of the acidic pollution from these hypotheses. The geostatistical approaches already discussed can be used in this context. Another approach is to use **cellular automata**. Cellular automata separate continuous space into discrete cells (see Fig. 1.3). The cells then react, by a series of rules or relationships, to the local conditions around the cell, for example, the condition of

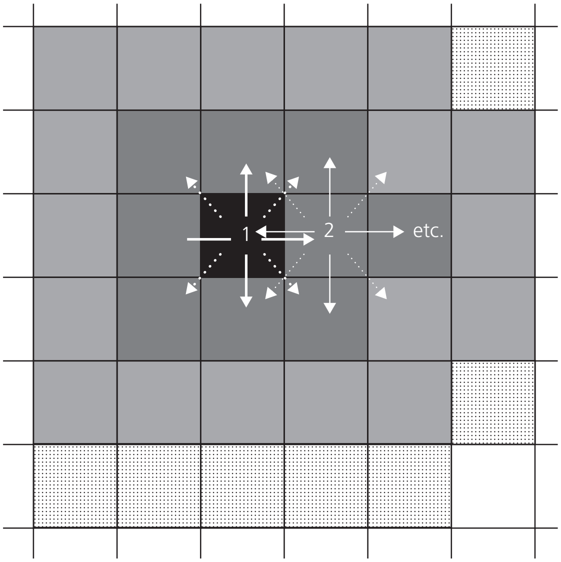


Figure 1.3 Using cellular automata to describe the movement of acid rain on Mars.

neighbouring cells. This creates a simple model describing the movement of the acidic pollution, allowing the amount deposited in rain in all cells around the point source to be calculated. The statistical relationship then provides an estimate of the effect of the acid rain in each cell on the Martian fern. If a model is constructed from cellular automata, you know that it describes movement across a region. As this requires a large number of calculations, the relationships used in the model tend to be simple statistical relationships, so cellular automata are usually functional models, providing process understanding of the movements only.

A very different approach is used if you want to know *why* the acid rain affects the Martian fern in the way it does. The research question must be restated as 'Why does acid rain on Mars affect Martian fern?' The simple hypothesis that is the basis of the statistical models must be replaced by a series of hypotheses describing the processes causing the Martian fern to be affected by acid rain water; hypotheses such as the following:

H_1: '*Aluminium is released from clay minerals according to the equilibrium constant for the acid reaction*',

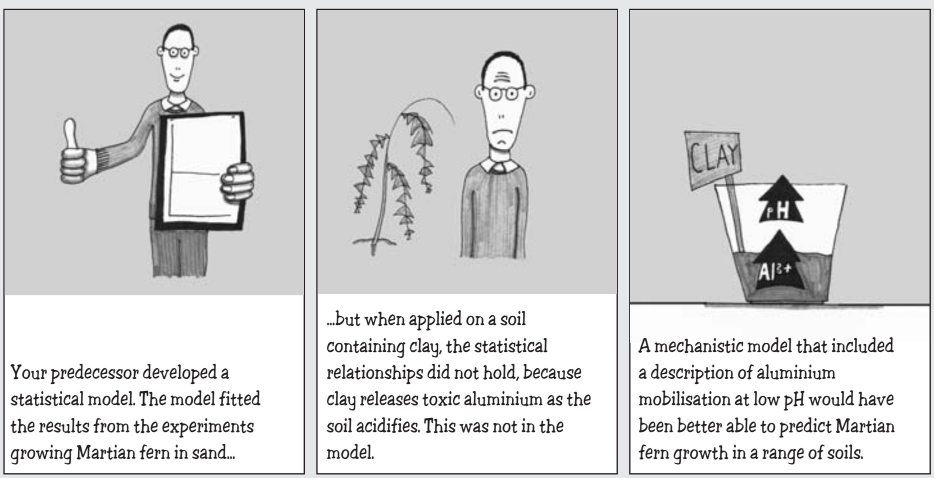
H_1: '*Aluminium is toxic to Martian fern at concentrations of over 20 ppm*',

and so on.

The model constructed from this series of hypotheses is no longer functional; it is mechanistic because it explicitly contains information about the processes in the system. Models of this type are usually described as **process-based** models. If a model is described as process-based, you immediately know that it can be used to understand the mechanisms affecting the results, and, if the processes are adequately described, then it should be accurate when used to predict.

This discussion illustrates how the mathematics used to construct a model influence the capabilities and robustness of the model. Knowing the type of model needed will determine the type of mathematics that are appropriate; this helps in model development. Knowing the capabilities of the mathematics will determine the application and scope of an existing model; this helps us to decide if an existing model is appropriate for a given purpose. Far from being confusing, the jargon associated with models can greatly increase our understanding of models. The list of jargon presented here is not intended to be exhaustive; this is not feasible in a constantly developing field where different people use different names for the same thing. The examples given are intended to demonstrate how you should respond when you come across a new mathematical approach. A builder, faced with a new type of building material, will find out the characteristics of the new material (How heavy is it? How strong is it? What is its shape? How can it be fixed together?). The builder can then assess what type of house this new material can be used to build. Similarly, a mathematical modeller, faced with a new piece of jargon describing a mathematical approach, should find out the characteristics of the approach. If the modeller knows what type of model is needed, then the suitability of the new approach can be assessed more easily.

Failure due to choosing the wrong type of mathematics



SELF-CHECK QUESTIONS: WHAT TYPE OF MATHEMATICS SHOULD YOU USE?

1. Q: To address the question 'Why does acid rain on Mars affect Martian fern?', which of the following mathematical approaches would most likely be inadequate?
   a. Mechanistic
   b. Statistical
   c. Neural network

   [A: b and c. Neither are process-based so cannot describe why acid rain affects Martian fern.]

2. Q: To address the hypothesis

   'The size of Martian fern plants will decrease by 15% with each decrease in pH unit of the rain water at a chosen site',

   which of the following mathematical approaches would most likely be inadequate?
   a. Mechanistic
   b. Statistical
   c. Neural network

   [A: None—all would potentially be capable of addressing the hypothesis.]

3. Q: For modelling spatial data, such as the distribution of Martian fern on Mars, which of the following mathematical approaches *could* potentially be used?
   a. Mechanistic
   b. Statistical
   c. Neural network
   d. Cellular automata

   [A: All could be used; d is ideally suited; a and b could be used if linked to a spatial data set of input data, and c could be used if enough spatial training data were available.]

4. Q: Which of the following are important when choosing the type of mathematics to use in a model?
   a. The question to be addressed
   b. The types of input data available
   c. Whether or not the mathematics is clever
   d. The hypothesis formulated from your research question

   [A: a and d are of primary importance; b also needs to be considered; and c is not important.]

1.5 Choosing an existing model

Before any time is wasted in developing an inappropriate model, the model required should be carefully classified using the above analysis. The model classification (whether static or dynamic, qualitative or quantitative, deterministic or stochastic, descriptive or predictive, or functional or mechanistic—see Section 1.4.1) will assist with searches of databases and the scientific literature for existing models in this area. You can then assess any available models for their suitability for your purposes. Developing your own model takes time and commitment, and like other areas of science should always build on the work of others. Carefully defining what is needed before you start will help you to do just that.

By classifying the type of model needed to answer the specific question, asked in a specific way, and understanding how the choice of mathematics relates to the type of model, you can quickly ascertain if an appropriate model already exists. If it does, and it is available to you, then you can try it out, and see if it does the job. Often, however, the right type of model does not exist (see Fig. 1.4). In this case, classifying the type of model

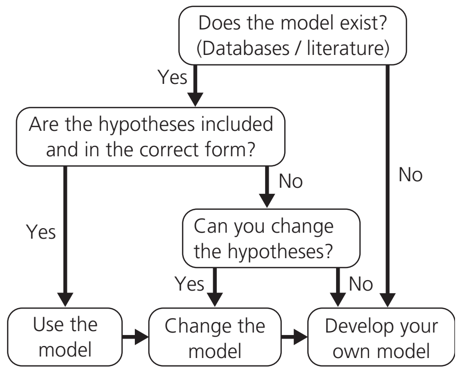


Figure 1.4 Should I develop my own model?

needed has saved you a huge amount of time learning how to use an inappropriate model, and wondering why it does not tell you what you want to know. If no suitable model exists then you have no choice but to make your own.

1.6 How is a model made?

Making a model that does just what you want it to do involves a series of important stages: model development, evaluation and application. These stages can be cumbersome and labour intensive, but if, in the interests of more rapid progress, you miss out any of these steps, then you are likely to end up with a model that does not behave itself and causes you embarrassment!

We start with the research question or practical objective (Fig. 1.5). It is valuable to write this down before you start to ensure that you do not get carried away with what

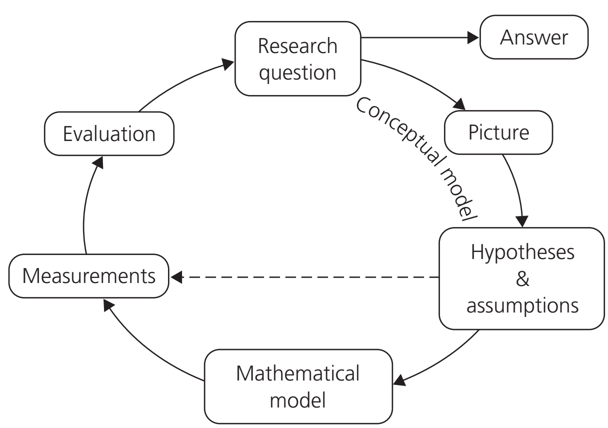


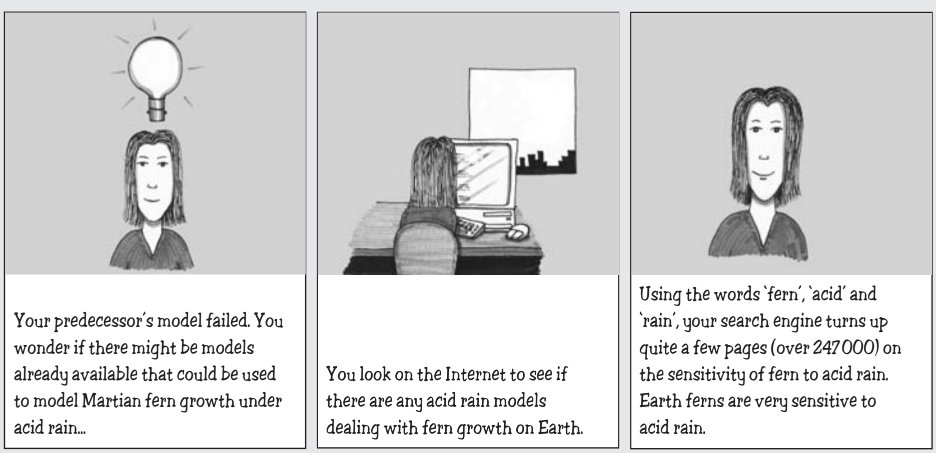
Figure 1.5 Stages in making a model.

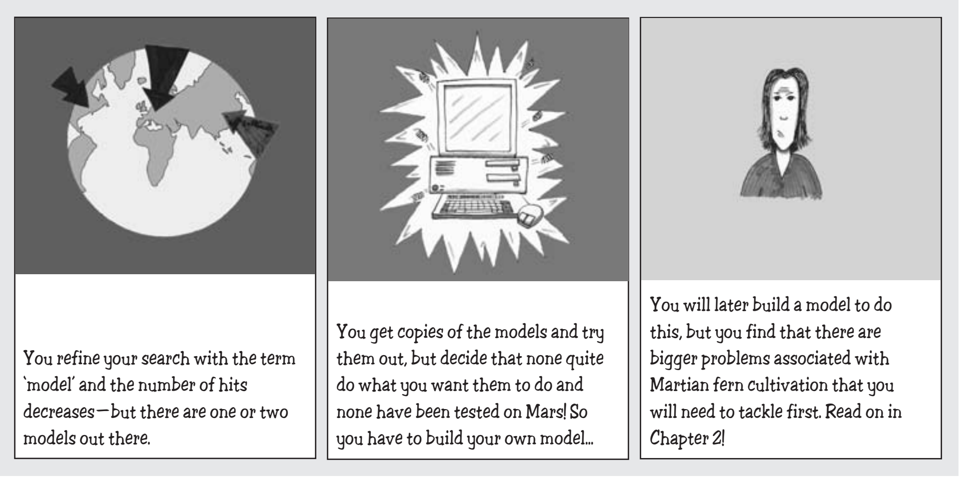
you *could* model and forget what it is that you *need* to model. It is often useful to visualise the problem by drawing a picture to summarise the key issues. The picture helps us to list the main hypotheses and assumptions that constitute the model. You have now made a **conceptual model**. This conceptual model can be examined to determine the characteristics needed in the working model. The hypotheses and assumptions are transformed into mathematical equations and linked together to produce the model. Some of the many different approaches used to produce and link together the mathematical equations needed (i.e., to develop the mathematical model) will be described in more detail in Chapter 2.

Developing the model is not the end of the process. It is crucial that, before it is ever used in anger, the working model be evaluated against independent measurements from the range of situations in which it needs to work. The response of the model to changes in input variables should be determined using a sensitivity analysis and the results should be compared to the measured responses of these factors. The uncertainty associated with the simulations should be quantified, and an uncertainty analysis performed to quantify changes in uncertainty in different situations. This will be described further in Chapter 3.

Following successful model evaluation, you have confidence in your model, and can begin to use it. The responsibility for ensuring appropriate use of the model rests with the user interface. Complete tables of input defaults, comprehensive error and warning messages, full help support and system documentation, and the clear presentation of results, are all inherent in the art of model application, which will be discussed in Chapter 4. To enable us to apply models in the real world, we must understand how models are conceived, developed and tested. If you want to do this, read on!

Choosing an existing model





SUMMARY

1. Models have the potential to
   a. compare the effects of two alternative theories,
   b. quantify expected results,
   c. describe the effects of complex factors, such as random variations in inputs,
   d. explain how the underlying processes contribute to the observed result,
   e. extrapolate results to other situations,
   f. predict future events, and
   g. translate our science into a form that can be easily used by non-experts.
2. Models can be classified by the hypotheses on which they are based.
3. A hypothesis can be classified by its outputs, inputs, scope and application.
4. A hypothesis can be classified by its outputs as
   a. qualitative (variable is a score or category), or
   b. quantitative (variable is a number).

   If a model includes any quantitative variables it is considered to be quantitative.
5. A quantitative hypothesis can be further classified as
   a. deterministic (variable is a single value), or
   b. stochastic (variable is the range of possible values).

   If a model includes any stochastic variables it is considered to be stochastic.
6. A hypothesis can be classified by its inputs as
   a. static (variables are fixed for any model run), or
   b. dynamic (variables change during the model run).

   If a model includes any dynamic variables it is considered to be dynamic.

7. A hypothesis can be classified by its scope as
   a. descriptive (interpolates observations), or
   b. predictive (extrapolates beyond observations).
8. A hypothesis can be classified by its application as
   a. functional (represents observations), or
   b. mechanistic (explains the underlying processes).
9. Models are further classified by the transformation of hypotheses into mathematical equations (e.g., statistical, geostatistical, Bayesian, neural networks, cellular automata, process-based).
10. Working models are made in the following three stages:
    a. development (research question leads to a conceptual model which leads to a mathematical model),
    b. evaluation (graphical, quantitative, sensitivity and uncertainty analysis), and
    c. application (user interface, defaults, error trapping, help support, presentation of results and system documentation).

PROBLEMS (SOLUTIONS ARE IN APPENDIX 1.1)

1.1. How would you classify the following model? A model has been developed to allow farmers to predict the potential increase in the weight of their cows with the amount of concentrate feed given each day. The model is structured around the following hypothesis:

> '*There is a 95% probability that the weight of cows will increase by 0.1–0.2 kg with each additional kilogram of feed given each day*.'

Specify the type of
- **a.** outputs,
- **b.** inputs,
- **c.** scope and
- **d.** application.

1.2. How would you classify the following model? A model has been developed to help policy makers estimate the changes in European soil carbon with changes in land management. This model must be used to quantify how soil carbon stocks might change in response to changes in land management and to determine improved management methods that might be used to increase soil carbon stocks. It is structured around the following hypotheses:

> Hypothesis 1: '*Land management (manure management, tillage practice and crop residue management) determines soil carbon stocks in cropland*',

> Hypothesis 2: '*Changing land management will change soil carbon stocks to a level determined by the new management regime*.'

Specify the type of
- **a.** outputs,
- **b.** inputs,

c. scope and
d. application.

1.3. How would you classify the following model? A model has been developed to help researchers estimate the size of a dolphin population, based on the probability of sighting a previously tagged dolphin on successive visits. The model is based on the following hypothesis:

> '*The probability of sighting a previously tagged dolphin is dependent on the size of the dolphin population*.'

Specify the type of
a. outputs,
b. inputs,
c. scope and
d. application.

1.4. How would you classify the following model? A model has been developed to help bioenergy growers to determine which biofuel crops can be grown at different sites. The model is based on the following hypotheses:

> Hypothesis 1: '*A biofuel crop will not grow outside a specified temperature range during the growing season*',
>
> Hypothesis 2: '*A biofuel crop will not grow above a specified elevation*',
>
> Hypothesis 3: '*A biofuel crop will not grow outside a specified rainfall range*.'

Specify the type of
a. outputs,
b. inputs,
c. scope and
d. application.