2. General introduction

2.1 I think, therefore I am (a modeler)

If thinking is a philosophical proof of existence, according to Descartes, then crafting models can be seen as a logical consequence of thinking. The notion that life is complex is as old as man itself. As are the models developed to understand nature. Assuming that any representation of the world is in itself a model, wildlife drawings of deer, oxen, and other creatures in a cave wall can be considered the primitive models (Figure 2.1).

Whether cave dwellers use those models as simple ornament (the primeval wallpaper), a feeble record of their own history, or to study the behavior of the animals and to delineate better hunt strategies, we can only speculate. But we do know they are probably the oldest representations of natural elements, and of men interaction with them. So the first models used in the study of biological systems can be traced back to the caves where our ancestors lived. Out of curiosity or necessity, sometimes both, men has been creating conceptual models ever since. With increasing levels of complexity and details, they stand among science's most notorious messages, from Copernicus representation of the solar system, to Mendeleev's periodic table, to Watson and Crick's DNA double helix (Figure 2.2).

Simply put, 'a model is a logical machine for deducing the latter from the former'[1], so both drawings and differential equations are just a way to express the ideas on which they are based. But what's so special about them? One can say that there is nothing really special about models, apart from the fact that they provide understanding, data and means to study the world that is impossible to achieve by the simple combination of analytical methods [2, 3, 4, 5]. Models try to capture the complexity of the subject they address and mechanistic models, for example, 'can be of great value in applied science, and offer many benefits over pure empiricism'[6], and provide 'the attractive capability to extrapolate and generate the bigger picture'[7].

Since it is impossible to fully address the complexity found in the living world, given the significant amount of limitations, starting with knowledge itself, models only capture a part of it; the most relevant aspects, modelers like to believe. And for that, they sometimes leave out details that may well be irrelevant, even though they seem necessary, ending up with functional, yet not so elegant representations of the real world, such as legless spherical cows.

Figure 2.1: Cave paintings: the first models of nature produced by humans.

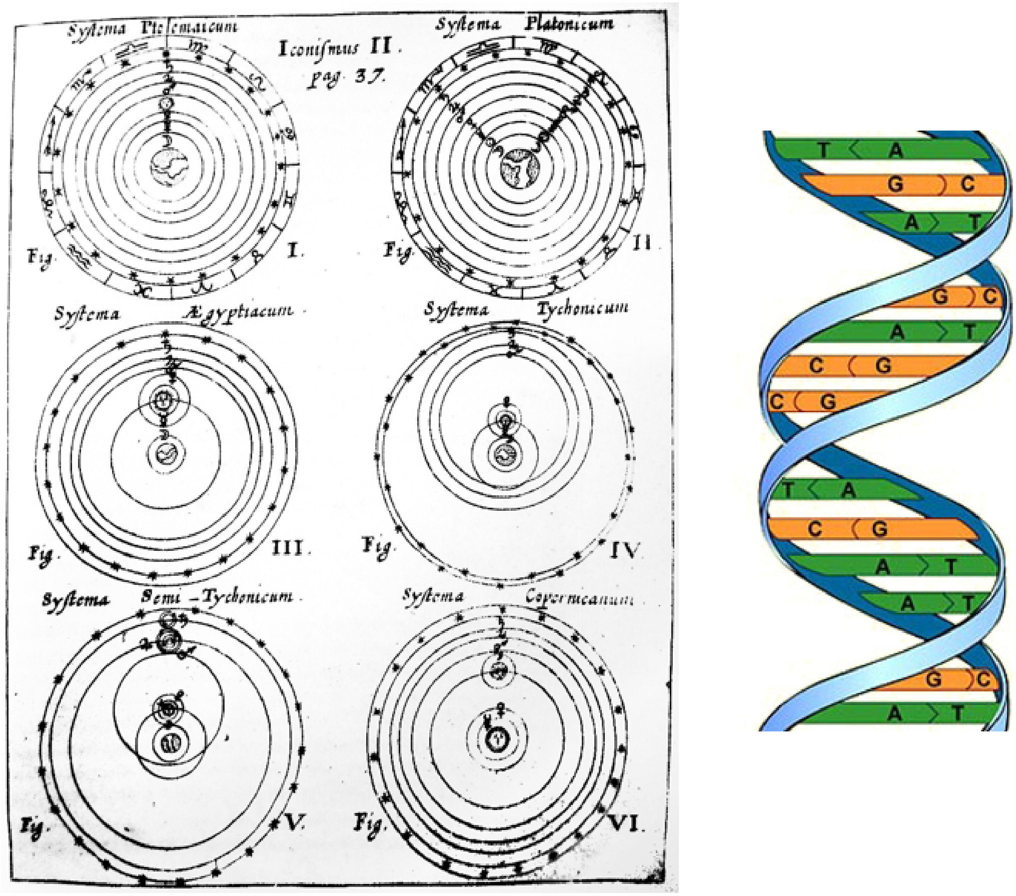


Figure 2.2: Nicolaus Copernicus model for the solar system (1543), originally presented in his book *De revolutionibus orbium coelestium* (left), and the famous DNA double-helix from Watson and Crick, originally published in 1953 (right).

2.2 The spherical cow – the modelers joke

The spherical cow is a modelers' tale that gives the title to a handbook in environmental modeling [8]. The story, used as an introduction to the volume, provides an honest critic to science by satirizing scientists in general and modelers in particular, though physicists are used as a proxy to modelers that are not mentioned by name. For the argument sake, the story is outlined here. A farmer wants to know why his cows aren't giving milk, and goes to Loomis Lab seeking for a physicist. Once there he requests a theorist to come up with a solution to solve the problem of his cows not giving milk. The theorist takes some moments to reflect and then rushes to the blackboard, draws a circle and starts his explanation: "Consider a spherical cow ..."

As with most jokes, the humor in the spherical cow is a vehicle for some elements of truth [9]. All ecological models fit the comparison to the spherical cow in some way, irrespective of their degree of complexity. But besides the physiognomy of the cow (shape or size) and the complexity of its physiology (the internal processes) the attention must also be focused on their product. Following the symbolism of the metaphor, modelers have to find the right way to milk their spherical cows or, as recently shown [10], they have to demonstrate that something useful can be done with their results.

2.3 The meaning of "Let's consider..."

The definition of a *model* is not the same for everyone, and can mean either a precise reproduction or a broad depiction of reality, or something in between. The same applies to the perception of detail that can range from simplistic to overly complex. But frequently it's not just a matter of definition, but a matter of expectation. To most end users a model may as well look as suspicious and unrealistic as a spherical cow to a farmer. This has been a major shortcoming attributed to models by non-scientists who are frequently unaware of the starting point of a model, namely the baseline assumptions that determine its complexity. Prepositions, conditions and assumptions are part of every analytical method, and frequently a key to explain its results. Assumptions stand as a simplification that facilitates the approach and study of processes and systems functioning (Figure 2.3), leading to solutions that otherwise would be unreachable. But they also can stand as a fatal flaw because they can undermine that same results.

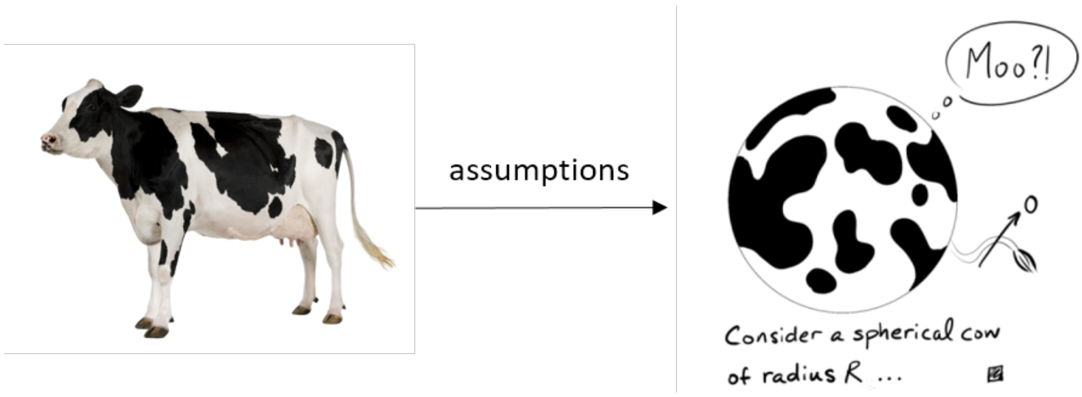


Figure 2.3: The spherical cow joke starts with the assumption that a cow is properly represented by a sphere.

Considering a spherical cow may imply a baseline assumption that shape or legs are irrelevant. Farmers know that cows are not spherical and have functional legs, nonetheless these assumptions may be realistic if the focus of the study is on milk production. Unless, of course, there is some kind of relation between milk and shape, or if the cow needs the legs to move to greener pastures in a way that is not possible by just rolling over. But if three-dimensional movements of the ruminant in a heterogenic space covered with random patches of grass are irrelevant, the legless assumption is viable. And the same applies for shape. 5 If correctly handled, assumptions are an effective way to balance model complexity. Unfortunately the same assumptions that are used to keep models 'as simple as possible' can doom the modeling exercise. By assuming a spherical cow and analyzing the problem, one can never reach the result that in fact the sphere is the proper shape to best represent a cow, i.e., every prediction of the model is based on the fact that the cow is spherical, so it's not something that the model attempts to validate. Unless, of course, if the model proves to be wrong. In that case the adequateness of the shape can be questioned.

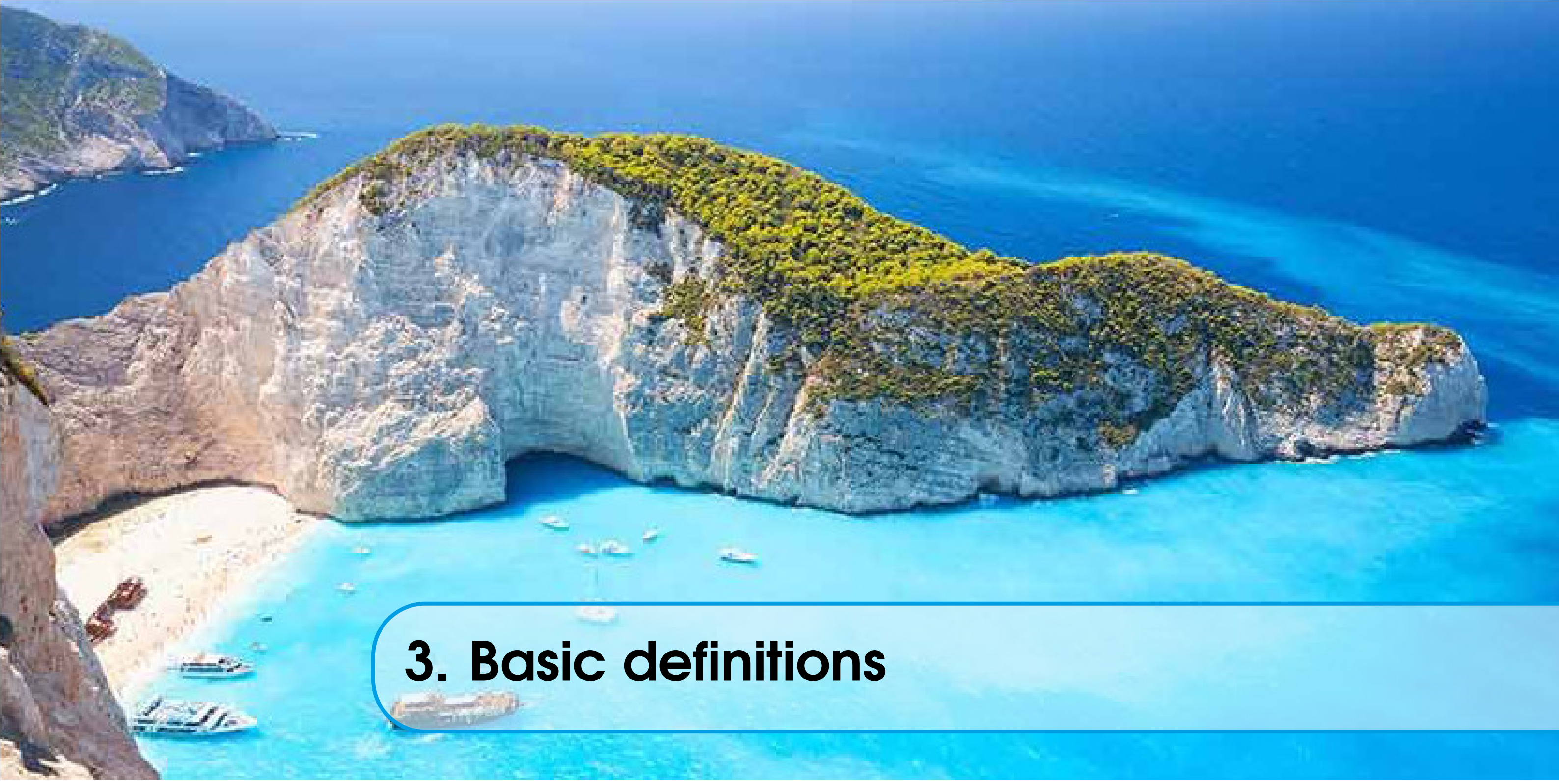
A model is by definition a small object, usually built to scale, that represents another object that is often much bigger and more complex. In that sense, a model is always a simplified representation of reality. Following this logic, a model is not reality at a smaller scale, but rather a small part of reality; the model only addresses some of the original features of the real object. So, there is an obvious question that haunts modelers: how much is enough? The specter of complexity is a constant presence because there is not a straightforward answer for the question of how much reality we must embed in our models, so to achieve any degree of realism.

For ecological modelers, for example, this question can have a slightly different form: how many parameters do I need to describe a particular process in an ecosystem model? The question may have different answers, depending on the processes and ecosystem, but it still is a fundamental step on the evaluation of any model and also a good starting point for any model developed from scratch. Clearly, the number of parameters required in a particular model varies with the purpose of the model, that is, the intended degree of realism and expected end products. Some authors say that the challenge is to describe biological processes and interactions adequately with a minimum number of additional parameters. And that seems to be the rule! Unfortunately this rule does not necessarily comes with a guide for its implementation!

R Some more elements on more a philosophical discussion on the role of models, on the importance of assumptions and on the ongoing debate around model complexity can be found in:

Mateus, M. (2017). Milking spherical cows – Yet another facet of model complexity. *Ecological Modelling* 354: 172-175.

(DOI:`http://doi.org/10.1016/j.ecolmodel.2017.03.001`)



3. Basic definitions

3.1 Defining a model

Models are simplified representations of reality, either some sort of miniatures (e.g., physical models) or equations (e.g., mathematical models). Consequently, models capture some aspects of reality, decided by the modeler, based on what he thinks to be important or relevant. So, mathematical models are based on observations, and by some form of experimentation, and there is a systematic procedure for selecting the aspects that are important and should be included in them. This process of selection may follow a systematic procedure, as illustrated in Figure 3.1 and, once there is a set of equations to start with, the model is then evaluated and applied, eventually. This process is illustrated in Figure 3.2.

Ultimately, the value of a model is determined by its ability to do what it was created to do, meaning that mathematical models should be evaluated with respect to its ability to achieve its objectives. Also, and equally important, they must fully realized if it is made accessible to its target audience.

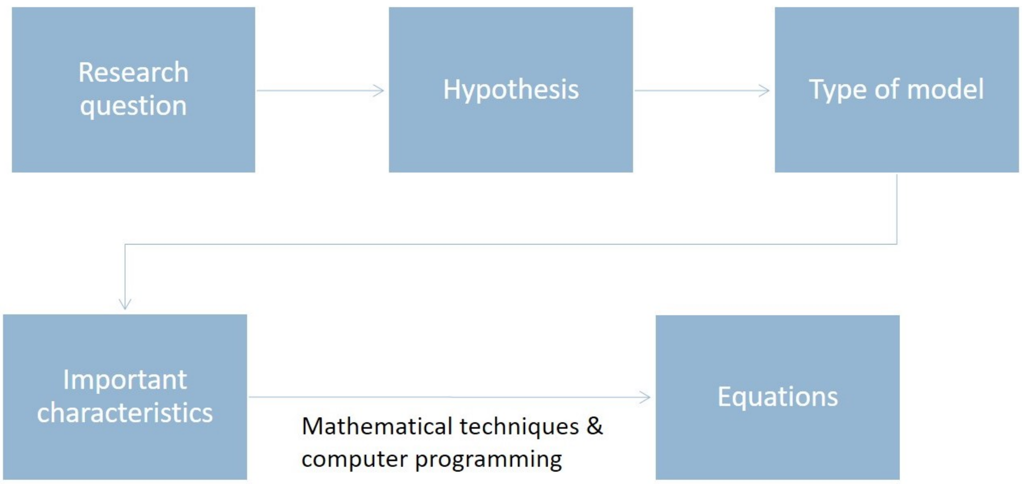


Figure 3.1: Systematic procedure for selecting important components of a model.

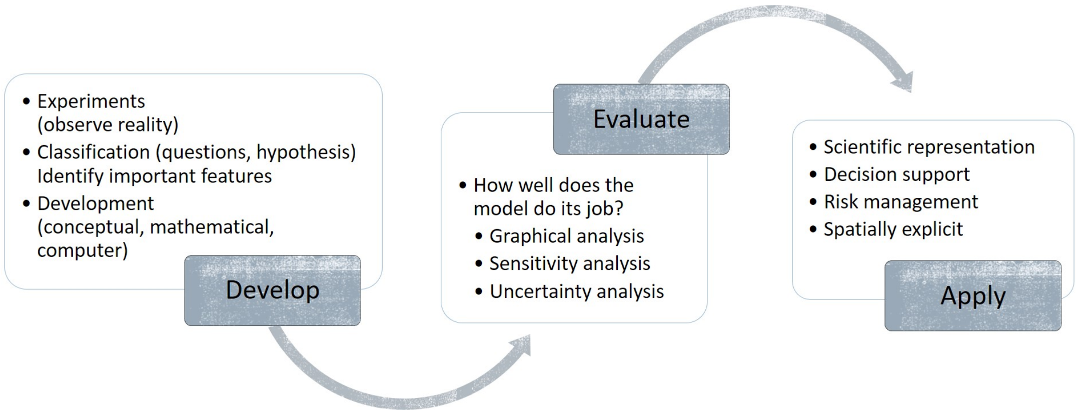


Figure 3.2: Different stages in the creation of a model.

3.2 After all, what is a model?

Models are ubiquitous tools, meaning that they are widely used in all branches of science. Frequently they go by different designations, the example being the periodic table, or the DNA-double helix, as mentioned before. The importance of such models is that they may generate new information and insights into the real world. The periodic table, for example, lead to the discovery of the properties of many elements even before they were discovered.

The interesting thing about modeling is that it comes natural to humans. But to make proper use of our inherent capacity to model the world around us, we must:

i. Have an unambiguous idea of what the model is intended to do (scope of the model);
ii. Follow procedures for model development, evaluation and application, and;
iii. Bear in mind that the natural world is unpredictable.

Models have the potential to:

- Compare the effects of alternate theories;
- Quantify expected results;
- Describe the effects of complex factors, such as random variations in inputs;
- Explain how the underlying processes contribute to the observed results;
- Extrapolate the results to other situations;
- Predict future events;
- Translate science into a form that can be easily used by non-experts.

In a nutshell, numerical precise hypothesis can be quantified with models, allowing observations to be explained and future events predicted.

3.3 Why computational models?

The best answer to this questions is probably another question: why not? Since most of our tasks have been increasingly performed by computers, why do the same for model building and simulation performance. After all, computers have the ability to perform complex calculations in a way that humans cannot.

Over the last decades, computer models have become standard tools in engineering and scientific research. In fact, they now rank among the most relevant scientific and technological advances of the last decades, with countless applications in every sector of engineering and scientific area. Their simulation and forecast skills, for instance, makes them irreplaceable tools in monitoring and management of natural resources.

Data from observations is seen as the most reliable way to know the state of a system at any given time. Although correct, this begs the question: do we have enough data to understand the world around us? The answer is usually no, and data shortcomings may lead to an incorrect interpretation of the behavior of the world around us, as exemplified in Figure 3.3. Models, on the other hand, can provide continuous temporal and spatial outputs, allowing to visualize the evolution of the object of study (example of model output in Figure 3.4). Finally, models equip us with the capacity to predict the evolution of the studied systems into the future, based on past and present knowledge of the system. This predictive, or forecast, capacity is one of the most important – and unique – features of numerical models. This capability of numerical models is schematically illustrated in Figure 3.5, in which the role of models is integrated in the classical method of gathering data (monitoring).

If, for example, our model is an ecosystem model, the leading principles for development and its relevance may be summarized as follows [11]:

1. Interpolations to fill data gaps, for instance to provide information regarding what is happening between two observations in time or to fill in the three-dimensional picture of a system from two-dimensional data.
2. Forecasting or hindcasting approaches, i.e., to make predictions for operational management when a system is varying within historical bounds.
3. Enhancement of systems understanding by quantification of a conceptual model (e.g., to calculate materials budgets) or to quantitatively test the plausibility of that conceptual model.
4. Developing ecological theory and generalizable ecological hypotheses.
5. Extrapolation and projection, i.e., to generate hypotheses regarding the function and likely responses of a particular system when perturbed beyond its previously observed state.
6. Scenario evaluations for operational or strategic management.

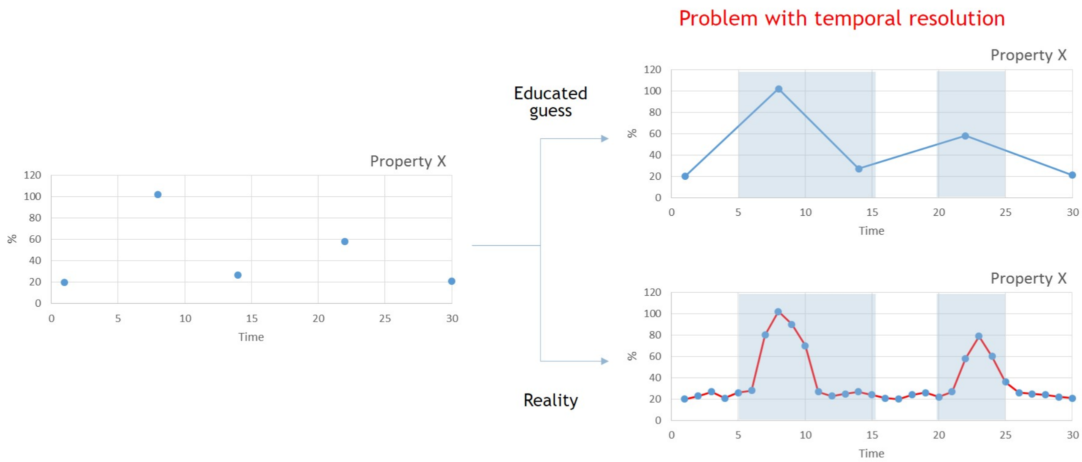


Figure 3.3: Data is frequently insufficient to describe the temporal (and spatial) variation of most properties and processes occurring in nature.

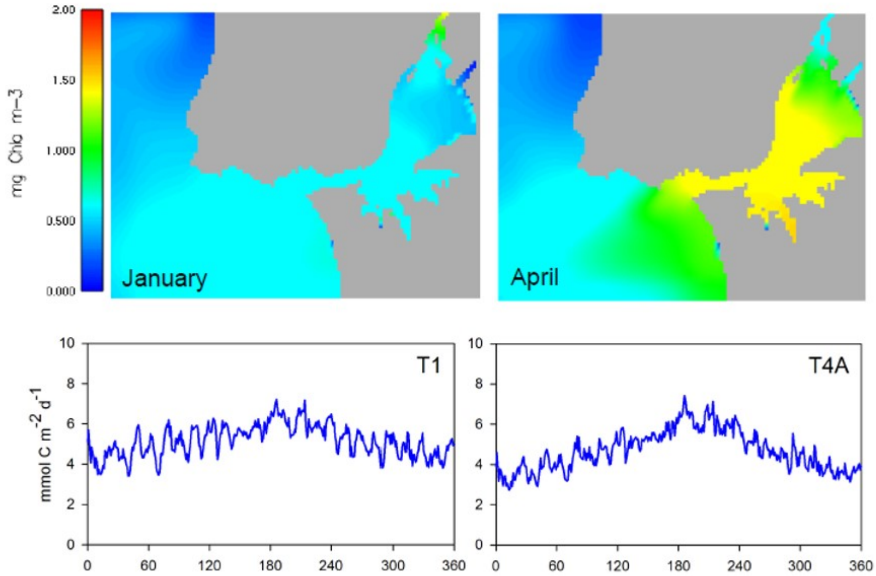


Figure 3.4: Model output for spatial distribution of chlorophyll concentration (top row) and temporal evolution of carbon dioxide fluxes at the water surface.

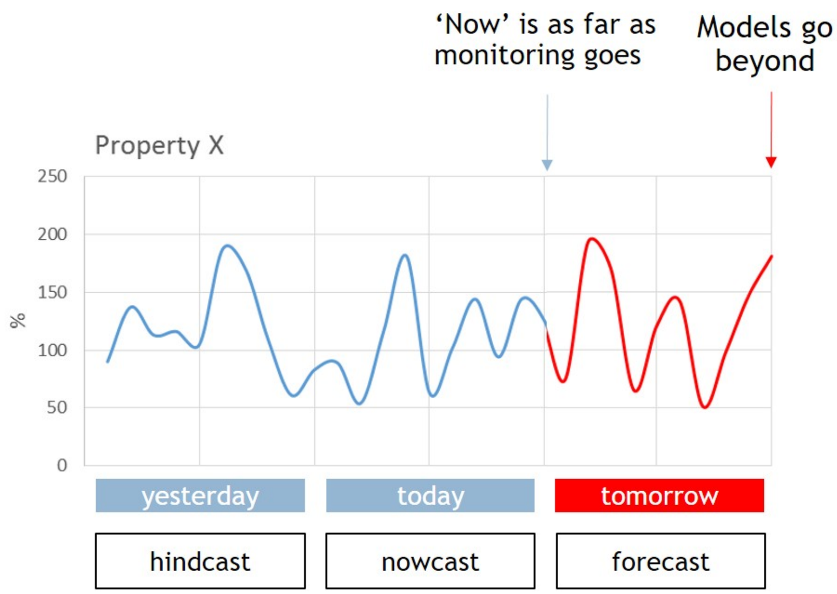


Figure 3.5: Numerical models allow us to 'predict' future behavior of the studied systems, an unique feature of these tools.

3.4 Which model should I use?

Making a model from scratch may be a demanding task, involving a serious of important tasks (see example in Figure 3.6). As such, some serious thought is needed before starting to develop such thing. For the argument sake, let's assume that most students or even professional will never get to develop their own models, but instead will work with models that are already available. If so, choosing a model from a batch of rather similar models may become a necessary first step. But before we can make an informed choice, we need to compare the models in their components, complexity, data requirements, etc.

Comparing models means that we look at the same features. However, descriptions of mathematical models in scientific literature often use lots of different terms. These terms can have one of two important purposes: (1) determine the type of model, and (2) describe the type of mathematics used in the model.

The **type of model** provides information on what the model does, i.e., on the sort of inputs used and outputs produced, the limitations of its applications, on how to use the model, etc. The **type of mathematics**, on the other hand, tells us how the model does that, more specifically how the model translates the inputs into the outputs.

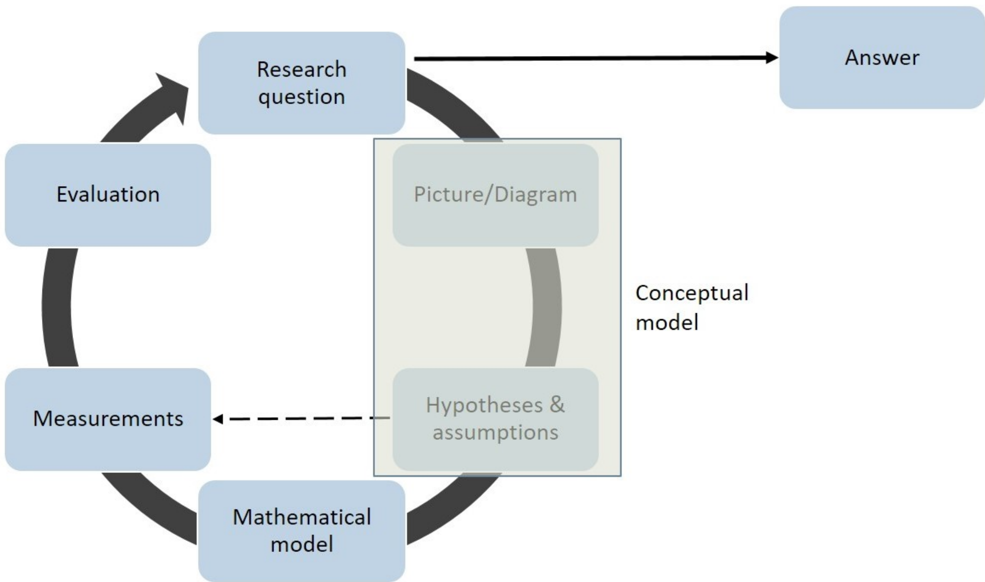


Figure 3.6: Making a model involves a series of important stages.

3.4.1 Assessing the type of model to use

The choice of a particular model must follow some basic criteria. First, we need to define the type of results that we expect. This is determinant to choose the type of model needed to achieve those results. The classification of a model focus on what the model can do, which is extremely helpful when choosing the right type of model.

Models can vary significantly in complexity, not just in their algorithms, but also on their underlying assumptions. This means that they can be based on anything from a simple hypothesis to a complex collection of hypotheses. Even the simplest hypothesis may include all the characteristics needed to classify a model. As such, by examining the characteristics of the underlying hypotheses, all of the major features of models can be characterized, even before we start using or changing them (e.g., by including new equations).

So, the classification of a model can be done according to its underlying hypotheses. From this we can then have a better idea on the type of model to use, by considering the following features: outputs, inputs, scope and application (Figure 3.7):

- **Outputs**: information produced by the model;
- **Inputs**: information needed by the model to run;

- **Scope**: defines whether the model is to be used to explain the results of the experiment, or to extrapolate to new situations. For instance, can the model be used outside the experiment used to develop it?
- **Application**: tells what the model will be used for. For example, is the model used to explain processes?

Model output can be **qualitative**, describing the nature of the output, or **quantitative**, if it provides a numerical measurement or count. Quantitative output provides a specific (or **deterministic**) value, or as a range (**stochastic**), specifying the probability (%) that the results falls within that given range. Input values can be fixed (**static**) or they can change over a series of measurements (**dynamic**).

Model can be termed **descriptive** or **predictive** in their scope, which is related to the hypothesis. Descriptive models are used to describe observations within the conditions of the current experiment. Predictive models, on the other hand, can extrapolate beyond the scope of the experiment. They can be predictive with respect to time, space, or any other input variable, and usually require some degree of understanding of the processes causing change (process-oriented).

The application can be based on a **functional** or **mechanistic** hypothesis. According to the mechanistic approach, the purpose of the hypothesis is to explain the underlying processes responsible for the overall results. Functional hypothesis, however, merely aims to represent or predict experimental observations. Consequently, the inputs needed to drive a functional model are usually less complex than those needed to drive a mechanistic model.

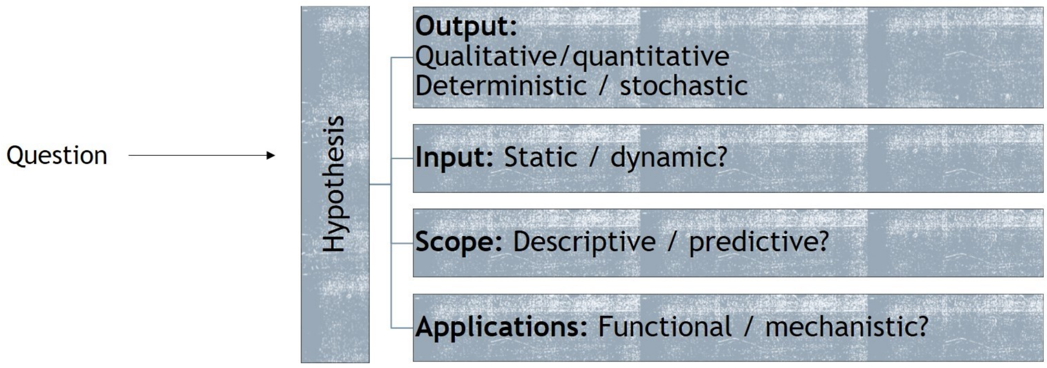


Figure 3.7: Model can be classified according to their outputs, inputs, scope and application. These features will determine if the model is adequate to address a specific hypothesis.

R **Some additional words on models and hypothesis**

Carefully and exactly state the questions to be addressed by the model. From this starting point you can then develop the null hypothesis and its alternatives. This is the basis of our model.

– **Null hypothesis** (H_0): a negative statement of the thing you want to test;

– **Alternative hypotheses** (H_1, H_2,..., H_n): list all possible alternatives to the null hypothesis.

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

3.4.2 Determining what type of mathematics

Different mathematical approaches can be used to address the general research questions associated with a specific hypothesis. Let's assume that a researcher wants to evaluate the effect of acid rain on plant growth. Suppose that his research hypothesis is: *The size of fern plants will decrease by 15% with each decrease of pH unit of the rain water at a chosen site*. To address this hypothesis, the model would be quantitative (the output, size, is a quantitative measure), deterministic (there is no interest in risk, so simpler single-value results is sufficient), functional (the model describes what the effects of pH are, not how pH has this effect) and descriptive (not intended 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. So, on this particular case, a statistical model is a good approach. In this approach the mathematical equations are directly derived from measured data using statistical techniques. Also, the model contains no understanding of the processes involved. Consequently, the model is unlikely to be accurate if used to predict results outside the conditions for which it was first developed.

It might be the risk to fern of the acid rain that are of most interest. Then, the hypothesis could become: *the size of fern plants will decrease by 10-20% with each decrease in pH unit of rain the water at a chosen site*. In this case, the model will be quantitative, functional and descriptive, but the outputs will be stochastic (providing an idea of risk).

Different mathematical approaches can be used for constructing stochastic models. **Bayesian statistics** [12, 13, 14], for example, incorporate prior knowledge and accumulated experience into probability calculations. However, it contains no understanding of the processes. A **neural network** (Figure 3.8) has a similar function [15, 16, 17], and can be trained by presenting it with examples of input and the corresponding desired output. The model works by mimicking the vertebrate central nervous system, to develop rules about relationships between inputs and outputs. If sufficient data is used to train the model, it can then be used in a predictive mode.

If an approach to describe spatial changes (movement across a region) in variables is necessary, then **cellular automata** (Figure 3.8) can be used [18]. These approach separates continuous space into discrete cells, each reacting according to a set of rules or relationships, to the local conditions around it. These models require a large number of calculations, so the relationships tend to be simple statistical relationships (provide process understanding of movements only).

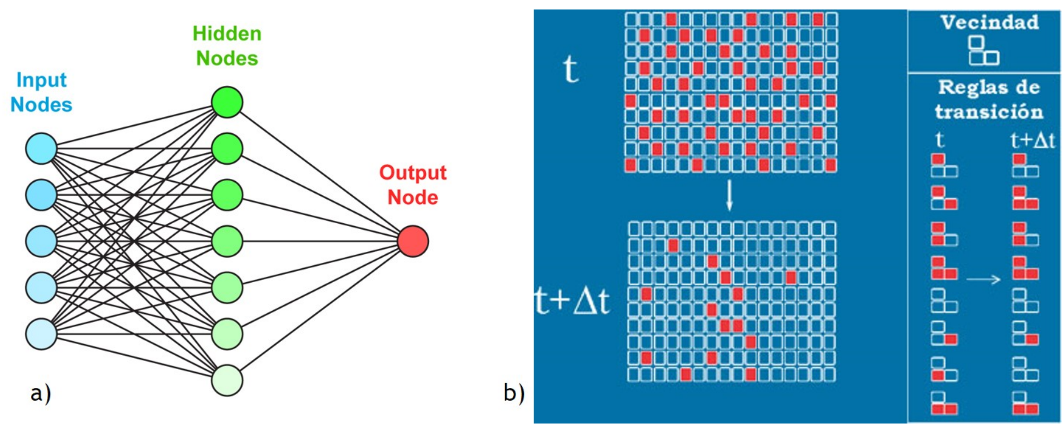


Figure 3.8: Neural network (a) and cellular automata (b) models schematics.

Considering the example provided above, a rather different approach must be used if we want to know why the acid rain affects the plant in the way it does. The research question must be restated, this time as *'why does acid rain affect fern?'* The initial hypothesis must be replaced accordingly, by a series of hypotheses describing the processes causing the plant to be affected by acid rain water, for example: *aluminum is released from clay minerals according to the equilibrium constant for the acid reaction*, or, *aluminum is toxic to plants at concentrations over 20 ppm*.

To address these hypotheses the model is no longer functional, but needs to be **mechanistic**, because it explicitly contains information about the processes in the system. Such models are usually described as **process-based** or *process-oriented*, and can be used to understand the mechanisms affecting the results.

3.5 Choosing an existing model

Finally, model classification will be a helpful preparation to search databases and scientific literature for existing models. This will lead you to ascertain if an appropriate model already exists. This process follows the paths depicted in Figure 3.9.

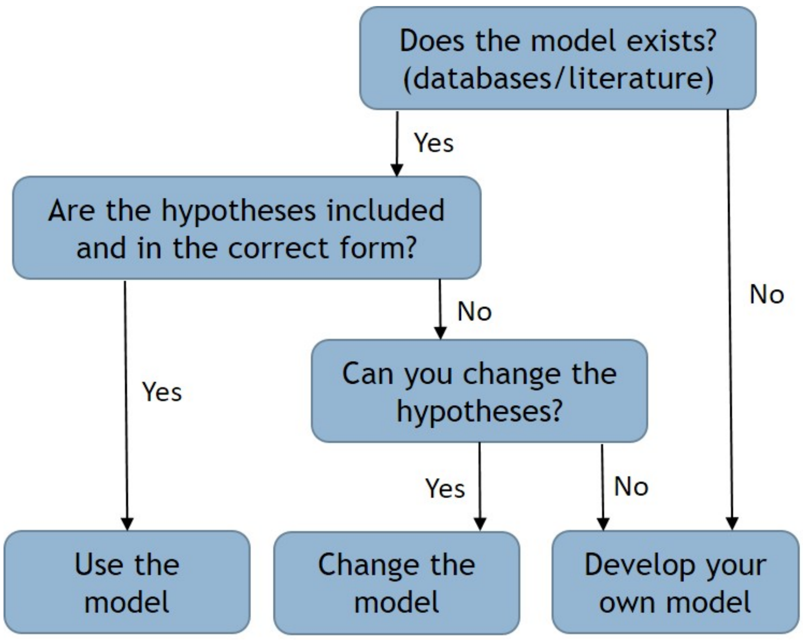


Figure 3.9: Decision tree for developing a new model or using an existing one.