4. Building a model from scratch

4.1 The essential steps

Developing a model from scratch can be a difficult task, if some essential steps are not followed in this effort. Before we start dealing with the equations that account for our state-variables (the mathematical model) or programming (the computer model), we need to make sure that we understand our problem, that we have identified the relevant variables, know the main processes involved and have a clear idea on the expected outputs. Simply put, we need to know what we want to achieve with the model. Consequently, when our aim is to develop a model, the following steps must be considered:

i. Start by choosing the type of model you want/need;
ii. Draw a conceptual model that broadly describes or illustrates the variables you want to address and the relationship between them (processes);
iii. Attach a mathematical model to your conceptual model by using equations to describe the evolution of your state-variables over time and the processes that control them;
iv. Finally, construct a computer model using the conceptual model as a blueprint and the mathematical model as the basis for all the calculus.

4.2 Basic terminology

Model structure relies on a rather small number of elements, each with a specific function and meaning in the overall scheme. Before we start developing our conceptual model, these elements must be understood, as well as the relationship between them. Only then can we proceed with the task of model construction in a logic and efficient way. Below there's a list of the elements that compose a model, their broad definition and some examples.

- **Constant**: Quantity whose value does not vary in the target system (e.g., speed of light);
- **Parameter**: Quantity whose value is constant in the case considered, but may vary in different cases (e.g., total solar radiation at top of atmosphere);
- **Variable**: Quantity whose value may change freely in response to the functioning of system (e.g., amount of precipitation);
- **Relation**: Functional connection or correspondence between two or more system elements

(e.g., rainfall, run-off and soil erosion);

- **Process**: Operation or event, operating over time and/or space, which changes a quantity in the target system (e.g., evapotranspiration);
- **Scale**: Relative dimension, in space and time, over which processes operate and measurements are made (e.g., local, regional, global; diurnal, seasonal, annual);
- **Structure**: Manner in which component parts of system are organized;
- **System**: Set of related elements, the relations between them, the functions or processes that govern these relations and the structure by which they are organized (e.g., forest ecosystem, drainage basin, global carbon cycle, earth's climate).

4.3 Developing a conceptual model

Conceptual models are expressed verbally, in written or diagrammatic forms (concepts), unlike the mathematical formulae that make a mathematical model, or the physical materials that compose a physical model. The conceptual model lays the foundations for numerical models and, consequently, the proper amount of time must be dedicated to them and their development. This process should involve an inclusive and detailed analysis of the *target phenomenon*, *process* or *system*, so that its component parts are identified, their respective *inputs* and *outputs*, the *relationships* between them and, finally, *processes* and structures that govern their interaction

There are no rigid rules for drawing a conceptual models, but there is a broad acceptance of a protocol that relies on the use of Forrester diagram symbols (Figure 4.1). These symbols are designed to represent any dynamic system in which a measurable quantity flows between system components. They are, nonetheless, an abstraction of the basic concepts of system components and material flows, but they make a general tool for qualitative modeling of systems.

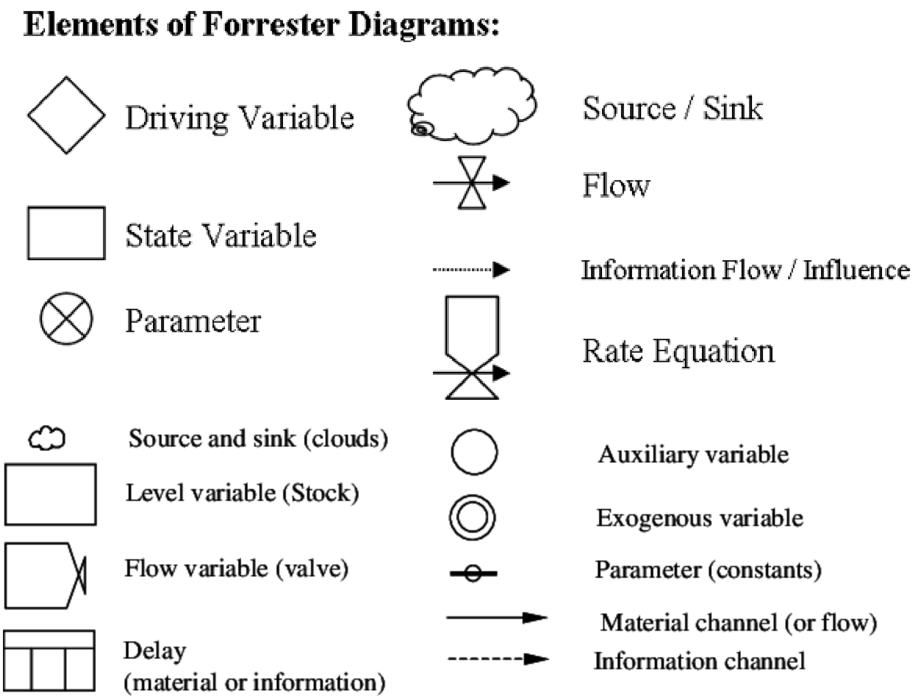

Figure 4.1: Some examples of the pictorial elements used in the Forrester diagrams.

4.3.1 Stating the assumptions

Every model is based on a set of assumptions, the number of which depending on the degree of complexity that the developer wants to achieve. These assumptions aim at simplifying the system enough to produce a working model and, at the same time, they should reflect the limits to knowledge of the target system. This means that the assumptions must be based on state-of-the-art knowledge.

It is important for the modeler to be aware that the validity and scope of the assumptions determine the value of the model. Ultimately, this means that the assumptions must be realistic. The choice of assumptions can be plastic, because some (or all) may later be found to be incorrect, in which case they can be revised in subsequent versions of model. However, for simplicity sake, the modeler may intentionally use wrong assumptions because they have an insignificant effect on model output or are necessary for reasons of simplicity or efficiency.

All assumptions must be identified, understood and stated explicitly. This will clarify the nature, purpose and limitations of the model. At the same time, it will help the user to understand the scope of model, challenge the assumptions on which it is based and, afterwards, to develop an improved versions of the model in the future.

4.3.2 List all hypotheses

The arrows denoting connections (fluxes) between state-variables on the conceptual model are usually built based on hypotheses. There may be one or a list of hypotheses to be incorporated in the model, and they represent the second important component of the conceptual model. The list determines which numerical equations must be derived when the mathematical components of the model are constructed. This implies that each hypothesis translates into at least one numerical equation.

4.3.3 List all assumptions

Models are simplification of the real world, which means that when we develop a conceptual model, that model is based on a number of assumptions. The list of these assumptions is the third important component of the conceptual model.

4.3.4 The principle of parsimony

When deciding what type of model to develop, we should rely on the principle of parsimony, also known as the Ockham's razor. This axiom states that 'of two competing theories, both of which are consistent with observed facts, we regard it as right and obligatory to prefer the simpler.' When applied, the principle of parsimony will prevent modelers to develop extremely complex models if simpler approaches yields similar results.

Also, a simpler model can be a more useful model because it provides greater understanding of the system, a greater explanation is afforded to each individual unit, and because individual units are less likely to be dependent on each other. From a computational perspective, simpler approaches will reduce the potential of error propagation through the model.

4.4 Types of models

4.4.1 Empirical models

An empirical model is primarily, or solely, based on observations, as distinct from a model that is derived from theory. The relationships between the component elements of the system are established by examining measurements of the variables concerned. The form of each relationship is defined by a mathematical function chosen from a set of candidate functions.

These models are not informed by theory, but derived from observation using data analysis techniques. A compromise must be achieved between how well the function fits data (e.g., via regression analysis) and the relative simplicity of its mathematical form. While valuable for making predictions in the cases for which they are developed, these models typically lack generality (difficult to employ them at other spatial locations or points in time). Hence, their use is rather limited or restrictive.

4.4.2 Deterministic models

A defining characteristic of deterministic model is that the outputs (results) are uniquely and consistently determined by the inputs (values used to drive model). Deterministic models, therefore, function in same way, and produce same output, each time they are run using a fixed set of input values. The state (i.e., the value) of a variable in a deterministic model is uniquely and consistently determined by the initial conditions of the model (i.e., the values of the constants and parameters input to the model) and, subsequently, the previous states of the variable itself.

These model are based on assumptions, theory or knowledge of the nature and form of the relations between the variables in the target system. Unlike empirical models, deterministic model tend to offer greater generality.

4.4.3 Stochastic (probabilistic) models

A stochastic (or probabilistic) model is one in which random events and effects play important role; the state of the model variables is, therefore, described by probability distributions, rather than single values. As a result, the model output varies from run to run, even when a fixed set of input values is used. This class of models is suitable where apparently random fluctuations in the system processes render deterministic models inappropriate. The observed random fluctuation may be due to processes and events that are truly random or they may be pseudo-random (i.e., where knowledge of a potentially deterministic process is inadequate or incomplete, such that it has to be treated as though it is random).

4.4.4 Static and dynamic models

Static models deal with systems that do not change, or at least are thought not to change, appreciably with respect to time. The focus of these models is on the processes or forces that keep the system in a state of equilibrium. Dynamic models, on the other hand, deal with systems that change over time and are typically constructed from difference equations or differential equations. The need to understand the behavior of an environmental system with respect to time leads the modeler to consider the issue of system stability.

The relative stability of an environmental system is partly controlled by feedback mechanisms. Feedback is the process by which a fraction of the output from the system, or part thereof (i.e., sub-system), is returned (fed back) as input to the system or sub-system. Many environmental systems also exhibit branching or splitting events, which make the behavior of the system more complex and, hence, more difficult to predict.

4.5 Analytical and numerical solutions

It is sometimes possible to construct a mathematical model by purely analytical means. The result is known as analytical or closed-form solution. These models are usually formulated so that they are expressed concisely in terms of mathematical equations, functions, variables and constants. Not all problems can be tackled with this approach, though; some must be solved numerically. The numerical solution relies heavily on computation power to "solve" the model by performing a sequence of operations, rapidly over and over again, each time applying operations to the result of the previous iteration. This procedure is expected to converge on the correct solution.

4.6 Black box and white box models

Mathematical models can also be classified according to the extent to which the composition, structure and operation of the target system is known and represented in the model. Models can be loosely classified as white or black box models, depending on what is known of its structure, even though these terms express the limitation of the user and not so much a property of the model. In white box models, the internal workings of the target system are known, completely understood and clearly stated, while on black box models the target system is treated as sealed unit, and is usually defined empirically. Most models, however, fall somewhere between these two extremes (gray box model).

4.7 Formulating the mathematical model

The next stage involves translating the conceptual model into mathematical formulae. This process, also known as formulating the mathematical model, is frequently the most challenging stage in the development of a model. The reason for this can be that the solution may demand the use of advanced mathematical techniques. Sometimes it may be that there are several ways in which system can be represented mathematically, and it is not immediately evident which is best. As such, deriving the mathematical formulation is often a trial-and-error process, and a skill that improves with practice.

4.7.1 Implementing the computational model

After having formulated the mathematical model, the next step is to program it in a computer language. In other words, the next step is to convert the model into software that can be run on computer. This step is labeled as implementation of the computational model.

At this stage, an enormous range of options are available, from spreadsheet packages (excel or equivalent), specialized modeling software or high-level computer programming and scripting languages (FORTRAN, Phyton, Matlab, etc.). Each has its own advantages and disadvantages, so their choice must be adequately pondered, and must take into account a number of issues, such as the cost (when applicable), personal preferences, prior experiment of the user with the software or computer language, ease which the software is learned (user-friendly, intuitive, etc.) and, the most relevant criterion, the suitability for the task at hand.