5. Model Derivation and Classification

In Chapter 4 we have used the method of least squares fit to arrive at a model that fits a given set of data. We will now look at a second approach, namely deriving a mathematical model by making assumptions about how the dependent (model) variable changes over time. This approach provides greater flexibility and gives insight into the meaning of the resulting model parameters (=constants). Section 5.1 compares the two approaches and outlines the process for developing a model from assumptions. In Section 5.2 we will briefly discuss criteria for the classification of mathematical models.

5.1. The Model Building Process

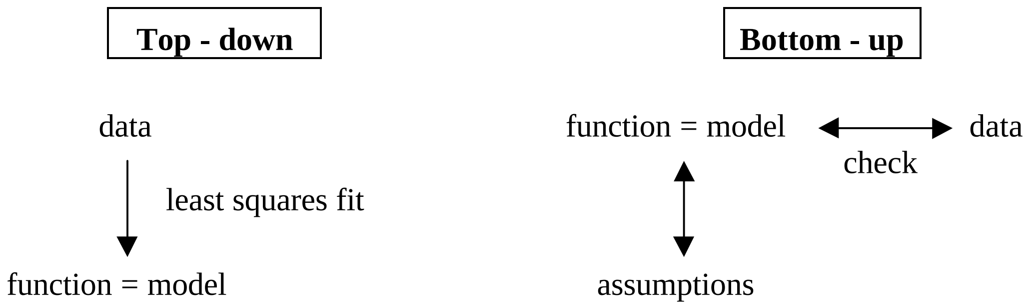
In the previous sections we have derived a model by using the method of least squares to fit a function to a given set of data. This method of starting from the result (= data) and trying to discover the underlying mechanism or process has several disadvantages:

- We are limited to the function types in our library - what if the data comes from another type of function?
- We do not gain insight into the mechanism connecting the input and output - the derivation does not involve any assumptions about the underlying process or context.
- There is no meaning associated with the parameters (= constants) of the function.

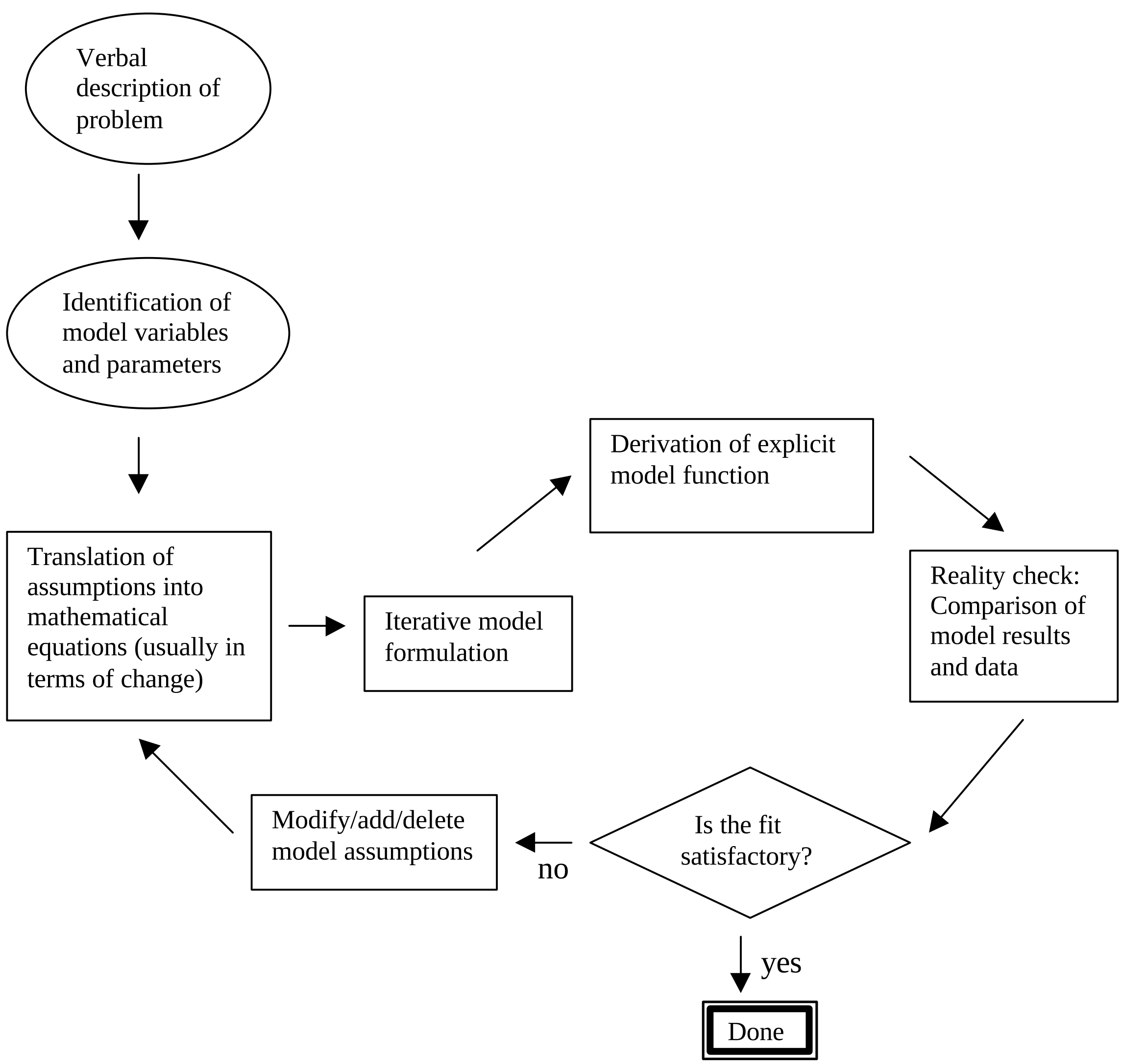
The first drawback can be addressed by using a larger library of function types. However, we are never guaranteed that the "true" function type is part of the library. All we can do is increase the likelihood that the appropriate function type is included.

The other two disadvantages cannot be fixed as easily. Even though certain function types such as the exponential or logistic functions result from very specific assumptions about change, this is not true for cubic and quadratic functions. Nevertheless, looking at the shape of the data may suggest some ideas about the assumptions for the model.

We will now approach the modeling process from a different angle, namely starting with assumptions about the model to derive a function whose input-output pairs are close to the data we have at hand. If this is the case, then the assumptions made when deriving the model are validated by the data. Here is a comparison of the two approaches:



In the *top-down approach*, the least squares method is used to produce the best fitting function of a specific type, resulting in the model equation(s). In the *bottom-up approach*, a model is derived from assumptions. If the model predictions are not close to the data, then the assumptions have to be revised until reasonable agreement is achieved. Here is a more detailed description of this adaptive *modeling process*.



This diagram poses a major question, namely "How do we determine whether a fit is satisfactory?" We have already seen that usually there is no complete match between the model results and the data (even with a function resulting from the method of least squares fit). It is unrealistic to expect an exact match as every model is a simplification of an often quite complex process. Thus, we continue the process of modifying assumptions until the resulting model agrees with the main features of the data. Such a match indicates that the model has captured the primary features of the underlying dynamics, even though smaller effects may have been neglected. Incorporating additional assumptions may explain these secondary features, but may make the model equations too complicated for an analytic solution (= a formula). Thus, we have to balance the simplicity of the model and the degree of fit. Generally, the goal is to find as simple a model as possible that exhibits the main features of the data.