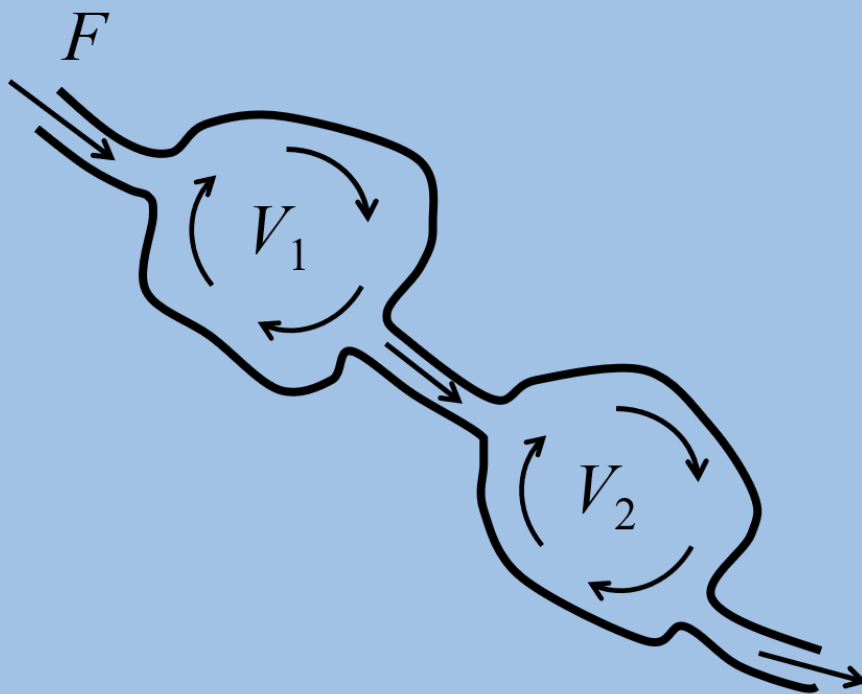


Mechanistic modelling essentials

PRINCIPLES, MATHEMATICS AND STATISTICS FOR
BUILDING AND APPLYING TK AND TKTD MODELS



Tjalling JAGER

Mechanistic modelling essentials

Principles, mathematics and statistics for building and applying TK and TKTD models

© 2021 Tjalling Jager
Version 0.2, Date: August 28, 2021

This e-book is published with Leanpub, and can be downloaded from https://leanpub.com/mechmod_book. This e-book will (hopefully) be regularly updated, polished and extended with new chapters. At this moment, this e-book is (unlike my other e-books) not available free of charge. Paying for it gives me the opportunity to continue working on this book, and on other ones. Furthermore, you will receive future updates of the book for free. Please note that you can always request a refund from Leanpub within 45 days, no questions asked, if you think the book was not worth the price. If the price is an obstacle for you, please send me a message.

You can cite this book as:

Jager T (2021). Mechanistic modelling essentials. Version 0.2. Leanpub: https://leanpub.com/mechmod_book.

If you spot errors (spelling, grammar, mathematical or conceptual), please notify me by email (see address below).

Tjalling Jager

tjalling@debtox.nl
DEBtox Research, Stevensweert, The Netherlands

Contents

Preface	v
About this book	v
Thanks	vi
Warnings	vi
1 General modelling principles	1
1.1 Introduction	1
1.2 The theoretician's modelling cycle	2
1.3 Systems and states	3
1.4 The role of assumptions	3
1.5 Model complexity and the need for generality	4
1.6 Mechanistic vs. descriptive models	5
1.7 Compartment modelling	5
1.8 Building blocks	6
1.9 Translation into mathematics and model testing	6
1.10 Confronting models with data	7
2 Basics of ODE's	11
2.1 Linear growth	11
2.2 Exponential growth	12
2.3 What if we cannot find a solution?	14
2.4 Notation issues	15
2.5 Dimensions and units	16
2.6 Why are differential equations so commonly used?	17
3 Compartment models	19
3.1 Unilake	19
3.2 Twin lakes	21
3.3 Unilake with a continuous input	22
3.4 Adding processes	22
3.5 What if the volume of the lake varies?	23
3.6 The role of assumptions	25
4 Error models and likelihood	27
4.1 Least squares	28
4.2 Defining the likelihood	29
4.3 Likelihood-ratio principle and confidence intervals	32

5	A journey into hyperspace	35
5.1	More than one parameter: parameter space	35
5.2	Profiling the likelihood function	39
5.3	What about CIs on model predictions?	43
6	Miscellaneous stuff	45
6.1	A word on Bayes	45
6.2	Bi- and multinomial likelihood	46
6.3	Normal likelihood	49
6.4	Evaluation	50
	Bibliography	52

Preface

About this book

This book is a re-worked version of the ‘refresher’ as used for the TKTD modelling courses (<http://www.debttox.info/dynmodtox.html>), amended with parts of my book chapter from Marine Ecotoxicology [9]. In both these works, I borrowed ideas from the teaching materials we used at the department of Theoretical Biology at the VU University, and from the textbook by Doucet and Sloep [4]. Chapter 5 also contains elements from the supporting information of the algorithm paper [13], plus elements from the GUTS book [15]. As you see, quite a bit of recycling went into the development of this book. Nevertheless, I think it is good to have all this information together in a concise and consistent form. Over time, I will attempt to add more topics and polish the contents.

In this book, I present the basic principles of modelling, and the bare essentials from mathematics and statistics that you need to build and apply simple models. I am writing from the perspective of TK and TKTD modelling in ecotoxicology, but these basics are applicable to many other scientific fields as well. Furthermore, I am specifically writing about models as tools to answer (applied) scientific questions. I am not so much interested in modelling for the sake of modelling. For me, models are the quantitative tools that make ideas (or theory) applicable to real-world questions. Therefore, this book is not just about mathematics (though there will be equations). It contains elements from philosophy of science, and the intellectual challenge of simplifying real-world complexity into a more manageable representation. Furthermore, statistics will be treated alongside the process models, whereas these topics are usually kept well separated in textbooks and teaching. As soon as we want to compare our model to data, we inevitably have to deal with statistics, so a solid statistical background is essential for practical modelling.

All of these elements are treated in a somewhat cursory and selective manner. I am focussing on the essential things that you need to get started with TK and TKTD modelling. I assume that the reader only has a very limited background in applied modelling, mathematics and statistics. That is why I will spend an entire chapter to familiarise the reader with the concept of differential equations. However, I assume that you are familiar with mathematical functions, powers and logarithms, derivatives and integration, and basic probability distributions such as the binomial and the normal. The chapters on likelihood functions and parameter hyperspace are probably less basic for most readers, as courses and textbooks on statistics seldom deal with this topic (in my experience, statistics is mostly taught from the perspective of hypothesis testing, which at best provides the correct answer to uninteresting questions).

Thanks

I would like to thank my former colleagues at the Department of Theoretical Biology of the VU University in Amsterdam, and specifically Bob Kooi, Bas Kooijman, Jacques Bedaux and Paul Doucet. Teaching courses with them, and discussing models and modelling in general, has shaped the thoughts expressed in this book.

Warnings

I noticed that some of my e-books are offered through other websites, after removal of the title page. Please download this e-book from Leanpub, which is the only place where it is legitimately offered. That will ensure that you have a complete and correct version, and that you are notified as soon as a new version is available.

The text in this e-book will contain errors and short cuts. If you spot an error or a short cut that is too short, please let me know and I'll address it in an update. I cannot accept liability or responsibility for any damage or costs incurred as a result of these errors or short cuts.

Chapter 1

General modelling principles

1.1 Introduction

A model is a simplified representation of a part of the real world. In the natural sciences, these models are generally presented in a mathematical form so that their performance can be evaluated quantitatively. Building models is not an aim in itself, but rather a means to obtain a deeper understanding of the mechanisms underlying processes observed in the real world. Without models, it would be impossible to interpret the integrated (often non-linear) effect of multiple factors that act on our system of interest. Furthermore, models are essential to make predictions for untested situations. In ecotoxicology, this can be useful for the optimal design of experimental work, but model predictions are particularly beneficial for a science-based risk assessment. For example, to evaluate the impacts of releasing a new chemical into the environment, or to evaluate the effectiveness of different mitigation strategies. The usefulness of models for regulatory purposes is demonstrated in environmental chemistry, as fate models are an integral part of virtually all frameworks for environmental risk assessment around the world. These models integrate quantitative knowledge about the transport and transformation of chemicals. Subsequently, they are used to understand chemical fate in the environment, and to predict the environmental concentrations (generally over time and space) from chemical properties and emission scenarios. In this way, fate models support risk assessments for new chemicals (before they are emitted into the environment) and new situations (e.g., an oil spill in a particular location). Even though a model is always a simplification of reality, and therefore always ‘wrong’, modelling should be an integral part of our quest to mechanistically understand and effectively manage the world around us.

In ecotoxicology, models are commonly applied to explain the uptake of chemicals in individual organisms over time (toxicokinetic or TK models). For the interpretation of toxic effects on individuals, hypothesis testing and dose-response curves are traditionally used. Even though these approaches can be viewed as crude models, there is no attempt to explain the observed effects from underlying principles (the purpose is to test for significant effects, or to interpolate, in a given set of data). Therefore, the two most important aims of modelling (understanding and prediction) are not served, and hence such descriptive approaches are not addressed in this book. More mechanistic models for explaining the effects on individuals over time (toxicodynamic or TD models) are gaining interest but are not as commonly applied yet in ecotoxicology as

TK models. Ecotoxicological modelling also increasingly takes place at lower levels of biological organisation (the molecular and cellular level) as well as higher levels (populations, food chains and ecosystems), but these models are not discussed in this book. The individual level is a central level of biological organisation for several reasons [10] but, for ecotoxicologists, the most important ones are twofold: effects at lower levels of organisation are hardly ecologically relevant unless they affect the life-history traits of individuals (i.e., growth, reproduction, survival), and effects at higher levels of organisation ultimately follow from changes in individual life-history traits by toxicant stress. In ecotoxicology, mechanistic models for toxic effects at the individual level are generally referred to as TKTD models, since they link a TK to a TD module.

This chapter presents the basic principles of model construction, the evaluation of their usefulness, and the confrontation to data. The focus is on concepts rather than on mathematical details; a firm grasp of the concepts is extremely important for biologists to be able to read and interpret modelling studies, and also essential before diving into the mathematics and coding. Furthermore, understanding the concepts is needed to design experimental tests in such a way that they can contribute to mechanistic modelling work. In this chapter, the amount of mathematics is therefore restricted to a minimum, and mainly used to illustrate the general principles. It is written from the perspective of TK and TKTD modelling, though these principles are more generally applicable to modelling.

1.2 The theoretician's modelling cycle

Before discussing modelling principles, it is good to consider the role of models in science in a more formal manner. For that purpose, I will borrow the empirical cycle as proposed by Bas Kooijman [17], as shown in Figure 1.1 in a very much simplified form. This cycle illustrates the interplay between the 'abstract world' of theory and modelling on the right hand side, and the 'real world' of experimentation on the left.

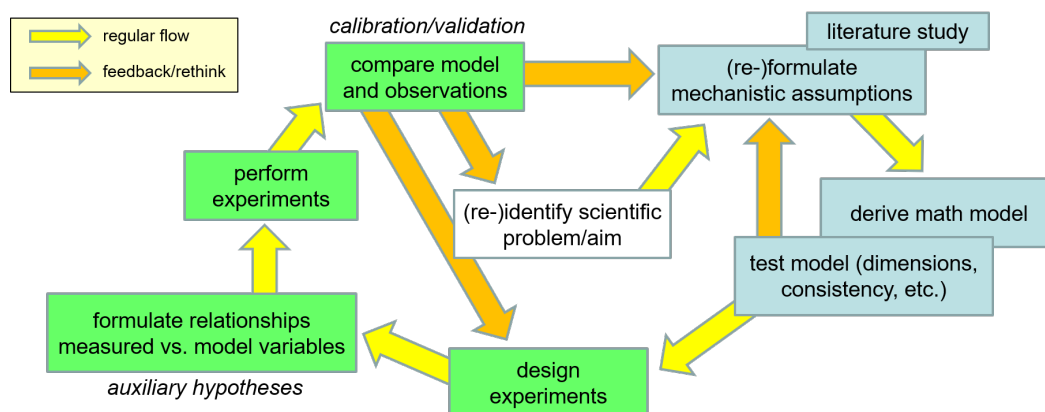


Figure 1.1: The empirical cycle from Bas Kooijman [17] (very much simplified).

I consider the cycle of Figure 1.1 to be an idealised view. In practice, model development and application will seldom fit such a nice and orderly cycle (see [14]). However, I do think it that this cycle is important to stress extremely valuable principles:

- Models are developed for a specific purpose. For the same system, a different

purpose or aim can lead to a very different model. A useful model for purpose A may be useless for purpose B.

- Useful models follow from mechanistically-inspired assumptions about the real world.
- Models should be extensively tested *before* comparing them to observation (this is discussed in much more detail in [17]).
- Experimental design should follow from the model.
- Useful models will often contain variables that cannot be directly observed. Linking models to measurements thus requires auxiliary hypotheses.
- Comparing models to observations is not the endpoint of the cycle. The results always lead to more questions that should lead to new experiments, or rethinking the model.

1.3 Systems and states

In modelling terms, a ‘system’ is basically a set of interacting components forming a unity, with boundaries separating it from the rest of the world. In biology, systems can be individual organisms, but also organs or cells within individuals, or populations of individuals, or entire ecosystems. For TKTD modelling, the individual organism is the system of choice, because the focus is on the uptake and effects of chemicals on individuals. However, the modelling principles from this chapter will be illustrated with other systems as well (such as the lake systems in Chapter 3).

The state of a system is specified by its state variables, relevant properties of the system that can change (generally over time). For example, the internal concentration is a state variable in TK models. To predict the future development of a system, we need to know the current value of the state variables, but we do not need to know their history. These variables thus fully capture the current state of the system, at least, as far as deemed necessary given the purpose of the model. The selection of appropriate state variables is thus a critical step in model design; for the same system, different research question may well lead to different sets of state variables.

In general, the change in the state variables over time depends on the current value of the states. Therefore, dynamic models are generally formulated in the form of Ordinary Differential Equations (ODE’s); equations where the derivative (the change in a state) is a function of the value of the state itself. The mathematical basis for ODE’s is explained in detail in Chapter 2.

1.4 The role of assumptions

Modelling does not start with mathematics, but with the identification of a scientific problem and the formulation of a research aim. Based on observations of the system to be modelled (e.g., from the literature), a set of simplifying assumptions is formulated. To be useful in a scientific setting, these assumptions should represent simplifications of the mechanisms that we assume to underlie the behaviour of the system. Clearly,

biological systems are complex, and simplifying this complexity in a useful manner is a considerable intellectual task.

Models thus follow from simplifying assumptions about (biological) reality; useful and well-described models follow uniquely from a clear and consistent set of assumptions. Unfortunately, models in biology are often presented in the form of equations only, and it is left to the reader to reconstruct the underlying assumptions that are made or implied. This unfortunate practice hampers the uptake of models (and their results) in scientific and regulatory settings. Therefore, I will try to make assumptions explicit for each model that is treated. In the dedicated e-books on GUTS [15], DEBtox [11] and DEBkiss [12] you will find explicit lists of assumptions.

1.5 Model complexity and the need for generality

Models are simplifications of complex systems. However, complex systems do not necessarily require complex models. Model complexity should be closely linked to the purpose of a model and the information available to parameterise it, and only to a much lesser extent by the (perceived) complexity of the system itself. Complex models are difficult to test (errors might easily go undetected), and require a lot of information to parameterise. And, more importantly, they will teach us very little about the system that we are modelling. The general strategy should thus be to start as simple as possible and only include more detail if absolutely necessary for the purpose at hand. Another aspect that should drive model design and model complexity is the degree of generality that is to be achieved. A model for the effects of chemical A on species B under the set of environmental conditions C could include a lot of detail on A, B, and C, and might thereby easily become very complex. However, developing a new model from scratch for each permutation of A, B, and C would be an inefficient use of time and resources; furthermore, the models (and their results) would be impossible to compare. Therefore, there is a lot to be gained by looking at what species and chemicals have in common, rather than focussing on their unique details. For this reason, I will focus on generic modelling principles and models here.

In Figure 1.2, I illustrate three distinct strategies for dealing with complexity. These strategies are discussed in more detail in an e-book [11]. A very similar categorisation was used by Hendriks [6], in a plea to focus on simple, parameter-sparse models for environmental risk assessment. Clearly, for risk assessment purposes, we cannot construct a new model for each chemical and for each species; we need generic models. Furthermore, we always need to extrapolate since toxicity tests are performed in the laboratory, under controlled conditions, while we need to say something about the impacts outside. This leaves the ‘simple box’ strategy as the most promising avenue. However, that is not to say that the other strategies do not have a role to play. Ultimately, the questions that we need to answer dictate the most appropriate modelling strategy. The most important take-home message is thus that there *are* different modelling strategies to consider.

An important guiding principle in model design is the selection of appropriate scales, both in terms of space and time. Combining very different scales into one model is inefficient and bound to lead to problems. For example, there is little to be gained by modelling processes at the molecular scale, which play out at the scale of milliseconds and nanometres, to explain effects on the life-history of a multicellular animal (playing

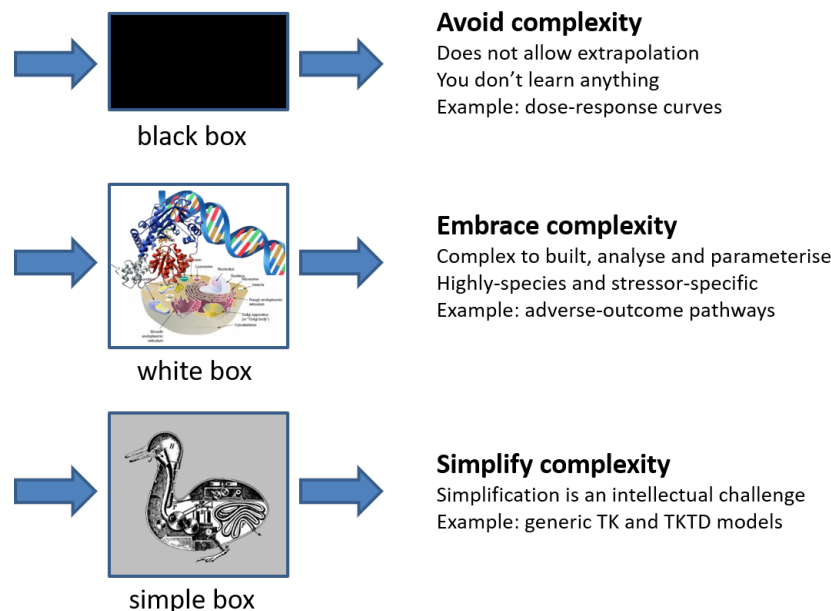


Figure 1.2: Three approaches for dealing with complexity in modelling biological systems.

out at a time scale of days to years, and a spatial scale of millimetres to meters).

1.6 Mechanistic vs. descriptive models

The distinction between mechanistic and descriptive models is not entirely straightforward. In ecotoxicology, dose-response curves and hypothesis testing are clearly descriptive. Such approaches are used to describe the data as they are, and we therefore do not learn much from their application about the underlying mechanisms. Furthermore, they do not allow for useful extrapolations beyond the conditions of the experimental test. For these reasons, summary statistics resulting from descriptive methods (such as the EC_x and NOEC) have very limited usefulness for science and risk assessment (see e.g., [20, 7]). A mechanistic model should be able to provide an explanation for the patterns, which should (at least in principle) provide a platform for extrapolation beyond the test conditions (e.g., from constant to time-varying exposure, and from *ad libitum* to limiting food availability). However, if we go deep enough, all mechanistic models will include descriptive elements, and the powers of extrapolation will have limitations in practice.

1.7 Compartment modelling

An important modelling principle that we can use to simplify biological complexity is the compartment approach. A compartment is a model component, generally with well-defined boundaries, that can be assumed homogeneous or 'well mixed'. This approach has a long history in modelling the environmental fate of chemicals (see [21]) and in toxicokinetics (see [1]). In toxicokinetics, we can often treat the organism (or more precisely: its internal concentration) as a single compartment. Compartments