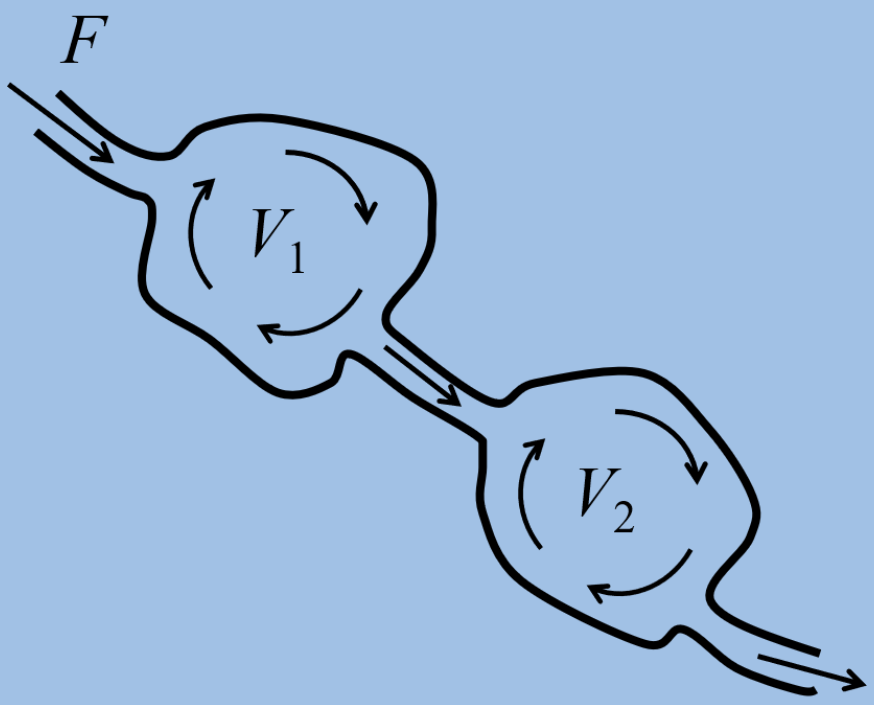
Mechanistic modelling essentials

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

Version 0.4, Date: June 14, 2023

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. 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 (2023). Mechanistic modelling essentials. Version 0.4. 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

About this book

The basis of this book is a re-worked version of the 'refresher' as used for the TKTD modelling courses (http://www.debtox.info/dynmodtox.html), amended with parts of my book chapter from Marine Ecotoxicology [16]. 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 [8]. Chapter 5 also contains elements from the supporting information of the algorithm paper [20], plus elements from the GUTS e-book [24]. Chapter 6 contains parts of the second chapter of my PhD thesis (also available from Leanpub at https://leanpub.com/jager_phd_thesis). Chapters 6 and 7 both borrowed from the GUTS e-book [24] and the 'DEBtox revisions' [18, 19]. 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 get started with building and applying simple models. I am writing from the perspective of toxicokinetic (TK) and toxicokinetic-toxicodynamic (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 statistical background is essential for practical modelling.

All of these elements are treated in a somewhat cursory and highly selective manner. I am focussing on the essential things that you need to get started with TK and TKTD modelling. I am hoping that this book is useful for readers that only have a very limited background in applied modelling, mathematics and statistics. However, practice teaches that reading a book (or watching on-line tutorials) is no substitute for a proper education. Therefore, I think this book works best as a 'refresher' for those that had math (especially differential equations) and stats (probability distributions and likelihood) in their education, but forgot most of it and/or are not too confident to

apply it. It could also serve as a starter, to set you up for more technical books or (on-line) courses.

I tried to write this book without specific models or modelling approaches in mind. However, it is inevitable that I have introduced a bias towards relatively simple compartment models for TK, and the generic GUTS and DEB-based models for TD (specifically in Chapters 6 and 7). This is not to criticise other modelling approaches, or to downplay their usefulness or importance; it merely reflects the areas where I have most experience and interest.

A proper understanding of (simple) differential equations is essential for TK and TKTD modelling. That is why I will spend an entire chapter to properly familiarise the reader with this concept. However, I assume that you are familiar enough with mathematical functions, powers and logarithms, derivatives and integration, and basic probability distributions such as the binomial and the normal. These concepts will not be explained in this book. The content of the chapters on likelihood functions and parameter hyperspace is probably new to most readers. 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). Therefore, don't worry if these chapters seem rather abstract on first reading. Applying these concepts in practice will be helpful to properly understand them.

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. I furthermore like to thank Andreas Focks for his feedback on a previous version of the book, which helped me to sharpen the text.

Warnings

I noticed that some of my e-books are offered through other websites, after removal of the title page (Page ii). 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 [17] 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 [33], 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.

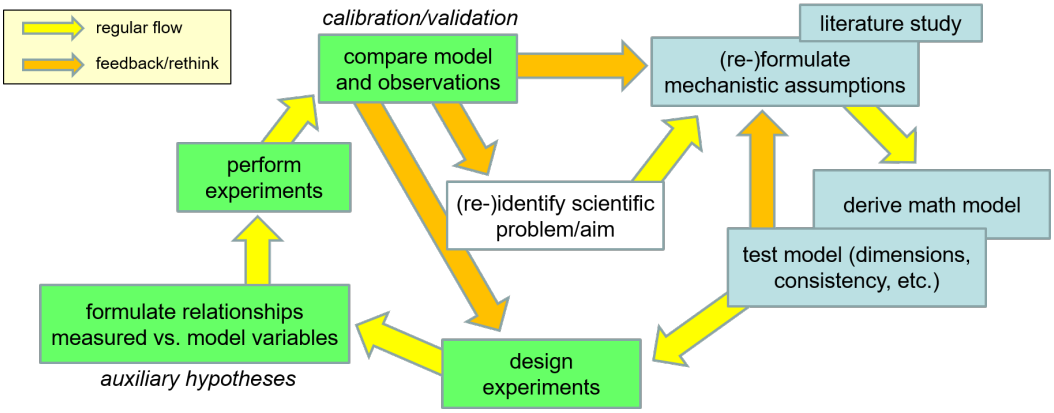


Figure 1.1: The empirical cycle from Bas Kooijman [33] (very much simplified).

We always enter the cycle by identifying the scientific problem that needs to be addressed 9the white box in the center of Fig. 1.1). In principle, we never leave the cycle: no model is ever perfect. When we leave the cycle, that would mean that this scientific problem is no longer relevant. Nevertheless, there will be outputs from the

cycle, which are not shown as they can occur at various points. For example, by publishing a consistent model formulation or the results of a validation.

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, orderly and isolated cycle (see [23]). However, I do think it that this cycle is important to stress extremely valuable priciples:

- 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 simplifying assumptions about the real world.
- Models should be extensively tested *before* comparing them to observation (this is discussed in much more detail in [33]).
- 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 first be illustrated with other systems (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, and their relationships, 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 or code, 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 recent years, I have tried to set an example for TKTD modelling: in the dedicated e-books on GUTS [24], DEBtox [18] and DEBkiss [21] 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 another e-book [18]. A very similar categorisation was used by Hendriks [13], 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.

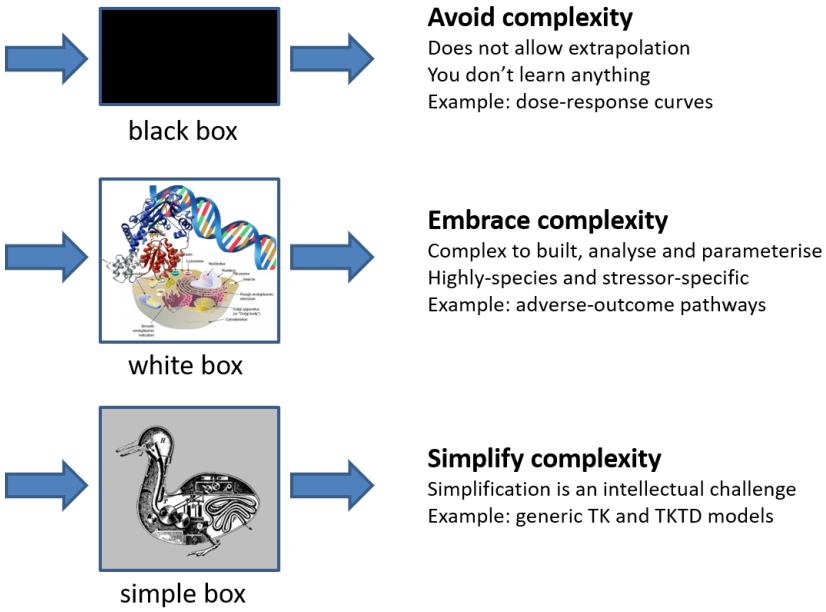


Figure 1.2: Three approaches for dealing with complexity in modelling biological systems.

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 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 ECx and NOEC) have very limited usefulness for science and risk assessment (see e.g., [39, 14]). 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.