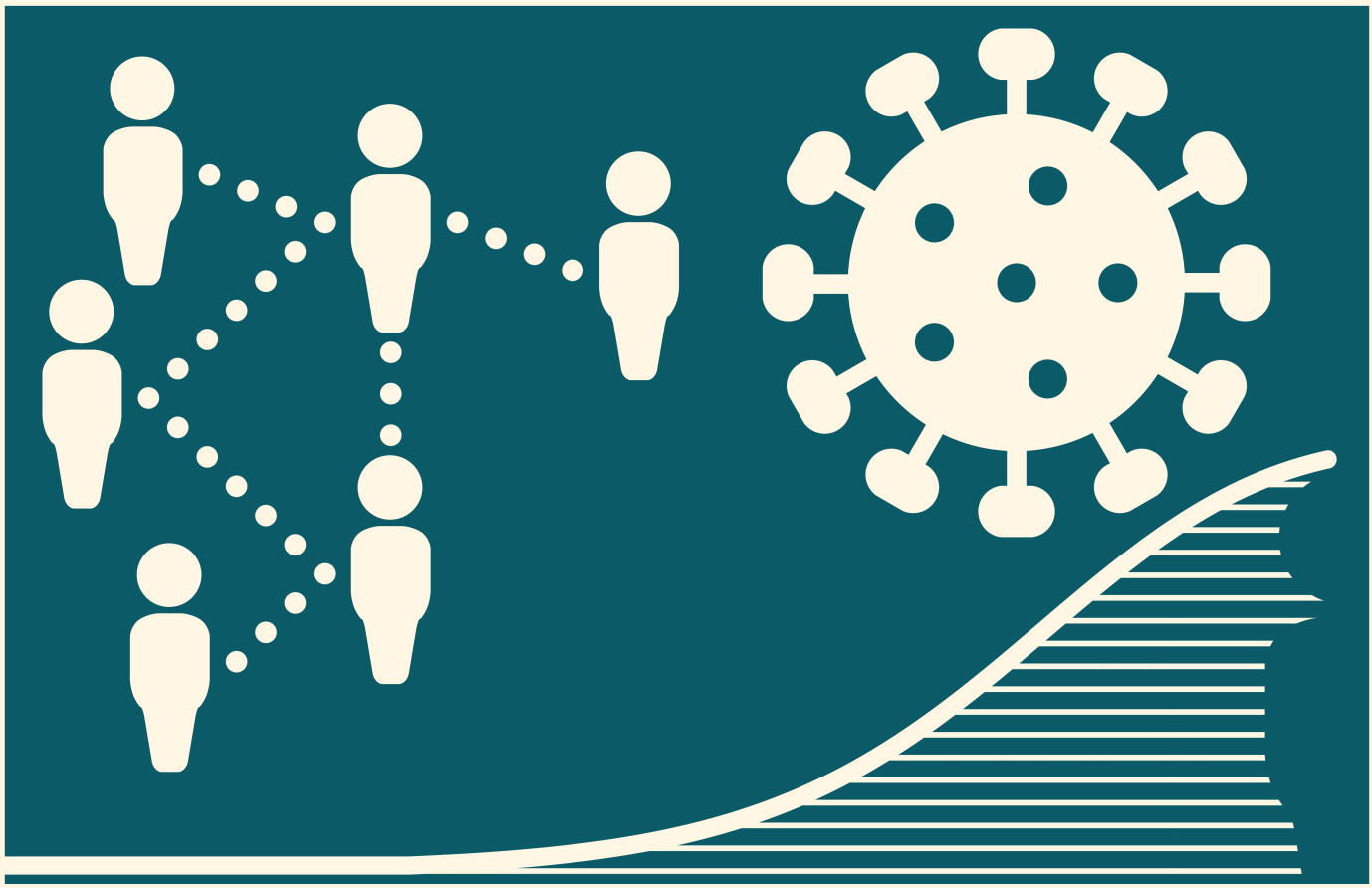
HANDBOOK OF MATHEMATICAL MODELLING OF INFECTIOUS DISEASES FOR DECISION-MAKING



Philip Gerlee, Torbjörn Lundh, Oscar Björnham
Lisa Brouwers & Anders Tegnell

Preface

This modelling handbook was developed as part of a project to strengthen the civil defence capacity in Sweden, run jointly by the Public Health agency of Sweden (PHAS), Chalmers University of Technology, the Swedish defence forces, and the Swedish Defense Research Agency (FOI). The project was financed by the Swedish Contingency Agency (MSB).

The COVID-19 pandemic highlighted the use of mathematical modelling as a tool for planning and for communication. It also demonstrated difficulties concerning communication of model results to a wider audience, especially when different groups present contradictory forecasts. A large number of models were rapidly developed and communicated in both academia and within public authorities in Sweden and internationally. In Sweden, the PHAS regularly developed national scenarios forecasting the potential development of the pandemic for the upcoming three months, and these scenarios were widely communicated and used. During the same period, on the regional level a number of universities supported relevant actors with models focusing on the local situation using local data, especially the need for hospital beds. Due to the urgency of the pandemic, there was no time nor possibility to initiate collaborations between the PHAS and the other groups producing mathematical models. Such collaborations would have been beneficial from many perspectives, and such collaboration is the aim of this project.

In the aftermath of the COVID-19 pandemic, the PHAS, Chalmers University, and FOI jointly wrote a proposal to the MSB to create a network of modelers in Sweden that would strengthen relations, trust, and communication in preparation for the next pandemic. The overall aim of this project is to increase the national capacity to model contagious diseases during a crisis like a pandemic by supporting wider collaboration regarding mathematical modelling. The network now includes mathematical modelers from both universities and national authorities and is engaged in several ongoing activities. In addition to the establishment of the network itself, this handbook was developed to present and discuss important modelling concepts and tasks. We believe that this handbook, developed by modelers, will support future collaboration between modelers by providing a common platform for collaboration and communication around mathematical modelling.

We would like to thank our external advisory board consisting of Lasse Engbo Christensen, Birgitte Freiesleben de Blasio, Helen Johnson, Bo Bernhardsson, and Mathew Biggerstaff for providing useful comments and suggestions on a preliminary version of this handbook.

Table of Contents

Chapter 1: Introduction

The increased use of models is not unique to epidemiology, but is a general trend in both the natural and social sciences. However, the term 'model' has different meanings in different disciplines, which can cause confusion in interdisciplinary projects where participants may have diverse viewpoints on what models are and how to best to utilize them. Two of the authors of this handbook have addressed this challenge previously (Gerlee and Lundh 2016), where the following definition of a scientific model was suggested: *Models are descriptions, abstract or material, that reflect or represent, and hence provide access to, selected parts of reality.* The scope of this handbook is narrower than that previous publication, and we will focus here on mathematical models of disease transmission and their connection to public decision-making.

Mathematical modelling plays a central role in infectious disease epidemiology both in terms of prospective studies, where future events are described, and retrospective studies in which past events inform us about disease characteristics. Mathematical and statistical models play an important role when crucial data are missing or when there are large uncertainties regarding key parameters of disease transmission because such models can extrapolate from existing data and provide decision-makers with an overview of possible outcomes.

The recent COVID-19 pandemic showcased both the benefits and shortcomings of mathematical modelling for decision-making in public health. The results from various models supported decision-makers during both the introduction and the removal of non-pharmaceutical interventions (NPIs), they informed healthcare providers about potential future needs of hospital and ICU beds, and they helped evaluate different strategies for vaccine rollout. They have also been used retrospectively for estimating the fraction of asymptomatic cases, inferring the serial interval, and assessing vaccine efficacy.

The pandemic put modelling groups in public health institutions and agencies under severe pressure because they were expected to deliver modelling results for a novel disease at an unprecedented pace. New constellations of infectious disease modelers also appeared when public health officials joined forces with academic research groups. However, some public health agencies were under too much pressure to be able to engage in new collaborations during the pandemic. This led to a situation where some academics directly joined the public debate by publishing their own modelling results. These factors, fuelled by the general sense of urgency during the pandemic, contributed in some cases to modelling processes that did not reach their full potential to support decision-making. Models constructed in haste, sometimes with unclear purposes, that were poorly documented and poorly communicated occasionally led to confusion among the public and sparked non-constructive debates between scientists.

We believe that the engagement of academic researchers with public health agencies during future pandemics is important and will provide added value. However, this cooperation would be enhanced if academics were to apply a well-defined modelling process and were to have a better understanding of how infectious disease modelling connects

to decision-making in contexts outside of academia. The aim of this handbook is to describe and discuss all of the relevant steps in the modelling process and how they are connected to decision-making – from the preparatory phase of conceptualization to the implementation, communication, and management of models.

The intended reader is someone who has a basic understanding of the technical aspects of modelling, including fundamental mathematical and programming skills, but who might have less experience in constructing models in the context of decision-making in public health. This task requires, among other things, knowledge about how the purpose of the model should guide its development, how uncertainty should be managed and visualized, and how the communication of the model's results should be tailored to the intended purpose of the model and to the intended recipients in the area of public health. These are skills that are usually less developed among academics who move into infectious disease modelling and among junior modelers at public health authorities. However, we also believe that the handbook could be a source of inspiration and discussion among senior modelers and managers. Furthermore, we also hope that this overview might have a harmonizing effect between different schools of modelers.

Mathematical modelling is a diverse field, and its practices are strongly influenced by the specific research area within which it is applied. Each research area that uses mathematical models has its own standards for how models are formulated, documented, and communicated. These standards are often tacit and are passed down from one generation of modelling practitioners to the next. However, in some contexts where models are linked to more formal decision-making processes there are clear guidelines. To give some examples from the health sector, the Food and Drug Administration (FDA) in the US regulates how computational models are to be documented if they are used in patent applications for medical technology (FDA 2016). Also, the International Society for Pharmacoeconomics and Outcomes Research is developing guidelines aimed at mathematical modelling studies (Weinstein et al. 2003), and the World Health Organization (WHO) is considering how evidence coming from mathematical models should be valued when it comes to guideline development (Egger et al. 2018). Another relevant document is The Aqua Book, which provides guidance on the development and use of mathematical models for the UK government (UK Government 2025). In relation to infectious disease modelling, a recent effort was made to develop guidelines for reporting items for epidemic forecasting and prediction research, called the EPIFORGE 2020 guidelines (Pollett et al. 2021). These guidelines provide a comprehensive checklist of items that should be reported in manuscripts, but they focus solely on academic publishing as a means of communication. In contrast, our scope is wider and we also consider the stages that precede publication and those that follow. In addition, we also consider communication with recipients other than fellow academics.

It is important to highlight that our aim is not normative in the sense that we do not want to prescribe a specific and definite recipe for modelling infectious diseases. On the contrary, we are aware that the modelling process often is adapted to local conditions and thus can take many shapes and forms. Independent of the path the modelling process takes, there will be a number of choices that have to be made, e.g. concerning choice of variables, model type, solution methods, visualization, and mode of communication. In

many cases these choices are never articulated, but rather are made unconsciously by the modeler or are based on traditions within institutions. The main aim of this handbook is to uncover these choices, helping modelers become aware of the multitude of paths that lead from the question posed to a modelling team to a finished project and helping them to reflect on the process. In order to avoid the connotations implicit in the term 'guidelines', we have opted for the more neutral term 'handbook' to describe this text.

Model development can roughly be divided into four distinct phases: Preparatory work, Implementation, Communication, and Management. These typically occur in chronological order, although iterations between the phases are common. This handbook is structured according to these phases, and in order to avoid abstract descriptions and to make the exposition more concrete we have included three different modelling cases that are used for illustrative purposes throughout the handbook.

In the chapter on preparatory work, we start by discussing how a specific purpose or question is central to successful model construction and how this question is formed in dialogue with decision-makers. We then move on to how the purpose of the model guides its conceptualization in terms of isolation and simplification. This process is also influenced by data availability. Lastly, we consider the range of different model types that are available, including their pros and cons, and how to connect the model and the data. This is followed by the chapter on implementation, which concerns the choice of programming language and the choice of solution algorithm for the model. We then move on to discuss how to verify the code and the process of model calibration, which is followed by a section on model validation and lastly a section on how model output can be visualized and how uncertainty can be communicated. Once the model results are in place, we move to the act of communicating the output. Here, we consider how communication is shaped by the purpose of the model, the intended recipients (e.g. experts or the public), the nature of the content, and the mode of communication. In the last chapter on model management, we take an organizational perspective and consider the factors the affect the reuse of models, e.g. model documentation, and we discuss the advantages and drawbacks of reusing models. The handbook ends with a concise checklist that summarises the key steps in the model construction process. We recommend that modelers fill in the checklist during the modelling process in order to articulate their modelling choices and to make the process more transparent.

Lastly, a note about what this handbook does not contain. We do not provide detailed instructions for constructing an infectious disease model. We also skirt around the technicalities of programming. Nor do we give a detailed account of model types or techniques for model calibration and validation. However, where appropriate we provide key references where details concerning different aspect of model development can be found.

Chapter 2: Preparatory work

Ideally, a model should be devised to answer a specific question; in other words, the purpose of the model ought to be known to the modeler. It is, of course, imperative to know what type of disease is to be modelled and the time frame and spatial location that the model should cover, but it is also important for the modeler to be aware of the model's intended use, such as what types of decisions the model will inform and who the intended recipients are.

If the specific purpose of the model is unknown at the outset of model construction, choices concerning elements such as the model input/output, the model structure, and the selection of variables might not align with the model's intended use and might instead be determined by the preferences of the modeler or the organization, which might not be suitable considering the ultimate use of the model.

When applicable, it is also important already from the start to have open and direct communication between the decision-maker and the modeler. In a well-functioning initial dialogue, the decision-maker might explain their expectations regarding the output, as wells as what decision-making mechanism the model output will be fed into. The modeler might then ask what data and knowledge about disease transmission exist today and what data might be obtained in the future. From there, a series of clarifications from both decision-makers and modelers concerning what might be possible to obtain and what one could wish for regarding both data and model outcomes will drive the modelling process forward. It is also desirable that different sources of uncertainty are discussed from both perspectives at this stage, as well as rough estimates of their magnitude.

There is a natural pull, both from decision-makers and modelers, to construct models that can serve a multitude of purposes and can answer questions that are posed after the model has been constructed. It is, of course, possible to construct such models, but typically this would require a time frame of validation that is not available in a pandemic or epidemic setting. Instead, during an ongoing pandemic there will usually be too little time to discuss and develop new models, and it is more likely that the discussion will focus on how to alter an existing model to obtain urgently needed results.

Below are examples of cases where the question the model is meant to answer is different as well as its intended purpose:

Model case 1. *What is the effect of imported cases on the disease incidence within a country?*
The model results will inform decisions regarding travel recommendations.

Model case 2. *How severe will an outbreak be in terms of morbidity under different interventions?*
The model will be used to assess the effect of different vaccination strategies.

Model case 3. *What is the expected number of hospital admissions in the coming weeks?*
The results of the model will aid hospitals in allocating the appropriate number of hospital beds.

A basic model of disease transmission at a national level could, at least in theory, answer all of the above questions, but such a model will most likely not perform very well for any particular question. Rather, the questions that are posed should guide the choices that are made throughout the modelling process, and this will most likely lead to different models that are each aligned to their corresponding purpose.

Another important factor besides the purpose of the modelling exercise is the time available for modelling. Given the typically exponential growth of cases early in an epidemic there is an urgency with which decision-makers need to act. When this time pressure is shifted onto modelers, this will by necessity shape the modelling process. If model results need to be communicated within a week, then this will lead to an accelerated process, where some steps, such as model validation and documentation, will be less thorough than desired. Even if the documentation can be completed afterwards, some validation would be highly preferable before using the model for any decision-making support. Although it is impossible to escape the time pressure in certain cases, it will improve the quality of the modelling exercise if the modelers are aware of the necessary steps because this will help them to prioritize activities under time constraints.

Before moving on, we would like to distinguish between two distinct purposes that predictive models can be used for. On the one hand, we might be interested in the most probable outcome, or *what is expected to happen*. This type of prediction is known as a forecast. On the other hand, we might be interested in knowing what would happen if certain future conditions are met, or *what could happen*. The latter type of predictions are projections based on given scenarios. Typically, forecasts are short-term (on the order of days to weeks), whereas projections are usually made on the order of months.

In the remainder of this section, we will, with the help of the above Model cases, discuss the steps of the modelling process that occur prior to implementation.

2.1 Isolation & Simplification

Prior to the formulation of a mathematical model, one must form a conceptual model in which all unnecessary details are stripped away so that only the relevant components and processes remain. This step also involves isolating the system we are describing. This essentially involves drawing boundaries in space and time in order to show what is included in the model and what is left out. This does not imply that the system is completely isolated, but rather that we set up boundaries both in time and space and use information about the external physical boundary conditions in order to decide what initial state to begin with.

A conceptual model is a simplified description of the system at hand, and it is often verbal (or even just mental) but is sometimes depicted using pictures and diagrams. Luckily, we do not have to start from scratch when formulating a conceptual model, and prior models and theories from mathematical epidemiology provide us with appropriate concepts and frameworks (Diekmann et al. 2012). Although this accelerates the modelling process, it is good to be aware of the framing that the theory provides. For example, the concept of discrete disease states (e.g. Susceptible, Infected, and Recovered) is a simplification of a much more complex reality where the disease state of an individual is multi-faceted and has several characteristics.

Similarly, the processes that drive disease transmission are often highly idealized in mathematical models (see section 2.3 below for a more detailed description of different model types). Compartmental models make use of the law of mass action, stating that the rate of reaction is proportional to the product of the concentration of the constituent reactants, which in terms of disease transmission translates to the incidence being proportional to the product of the fraction of infected individuals and the number of susceptible individuals. Agent-based models represent disease transmission at a higher level of realism, requiring that individuals inhabit the same location in order for transmission to occur. But even in this case certain assumptions have to be made. For example, assumptions have to be made about the probability of disease transmission in different locations (e.g. home vs. school) and about who will interact with whom. Similar considerations also lead to the use of network models (Keeling and Eames 2005).

It is of course also possible to consider model types that lie between compartmental models and agent-based models (as discussed below). The possibilities are enormous and, time permitting, it could be illuminating to consider a set of models in order to learn from various types of models instead of looking for a single “optimal model”.

Lastly, we would like to mention the common assumption that the age structure and size of the population remains constant. This assumption is closely tied to the time span that the model is meant to accurately describe and is reasonable for most epidemics. However, in some cases such as recurring measles epidemics the inclusion of vital dynamics (i.e. including births and deaths in the model) is required for an accurate description of disease transmission (Bjornstad et al. 2002).

These are only a few of the most common simplifications made in epidemiological models, and each model tied to a specific purpose will contain simplifications associated

with the problem at hand. In relation to the modelling cases presented above, we would expect Model case 1 to account for imported cases, while the other two models might disregard this aspect. Model case 2 needs to account for disease transmission and must make additional assumptions concerning things like age-dependent risk, and this is probably not relevant for Model case 1. Lastly, Model case 3 needs to capture the delay from infection to hospitalization, which is less relevant for the other two Model cases.

The degree of simplification makes it possible to order models from the most simplified and unrealistic, but understandable, models to more complex and, hopefully, more realistic models. At first, it might appear as if this ordering also reflects the accuracy of the models, such that complex models, which reflect more of the actual processes related to disease transmission, should be more accurate. However, this is far from the case, and in many instances more complex models are actually less reliable. This stems from the fact that including more components and processes requires more parameters to describe those processes, and if accurate knowledge about those parameter values is lacking then estimating those (either from the literature, from the data, or based on expert opinion) necessarily introduces uncertainty into the model output. This phenomenon has been termed the 'uncertainty cascade' (REF, chapter 4).

2.2 Choice of data

The choice of conceptual model dictates the variables that are included, and often there is an interplay between the data that are available at the time and the model construction process. In a sense, model construction can be seen as a dialogue between the purpose of the model and the available data. More types of data and higher spatial and temporal resolution implies that the model can be made more detailed. For example, if we obtain data on hospital admission at a weekly resolution then it usually makes sense to define the model and its parameters on the same time-scale rather than on a scale of days or hours.

Furthermore, the modeler should have a dialogue, as described above in the Preparatory work section, with the data provider in order to make suggestions about what data to collect.

In order to connect the variables in the model to the data, additional assumptions are required. Even if the variable and the data are similar at the surface level, a closer examination can reveal discrepancies. For example, from a given model we can calculate the daily disease incidence, and we also have access to national daily incidence data. We then have to ask how such data were collected. Is every infected person featured in the data or is there a testing bias? How are the data reported and aggregated? Are there regional differences that induce a bias? When possible, one should try to reduce bias. However, in cases where this is impossible, one should acknowledge any potential bias that might affect the accuracy of the model. In other cases, we need to make assumptions about the relation between the data and the variables in order to make a connection. For instance, this occurs when our model outputs disease incidence but we have data for

hospital admissions (as in Model case 3 above). We then need to make assumptions regarding the fraction of the cases that become hospitalized and the time delay from infection to hospitalization.

Data that are connected to epidemiological models fall into four broad categories:

- Data that are input to the model and affect different parameters, such as vaccine coverage and the rate of under-reporting. This also includes data that determine the initial conditions such as the fraction of the population that initially is infected.
- Data that affect parameters in a time-dependent fashion. In Model case 3 it is possible to use data on mobility to model a time-dependent contact rate (see, for example, REF). Other types of input data include syndromic data from, for example, telenursing calls and weather data in terms of outdoor temperature and humidity.
- Data that correspond to the main model output. For our three Model cases presented above this corresponds to disease incidence, mortality, and hospitalization, respectively. Such data are usually used for model calibration.
- Data that are connected to the output from the model but are only used for model calibration. In Model case 2 we might consider seroprevalence data in order to estimate the actual number of infected people.

In addition to the quality and potential biases present in the data, the choice of data should also be informed by the data's stability over time. Testing strategies that are subject to changes during an epidemic imply that data on incidence will exhibit time-dependent biases that might be difficult to account for. Hospital admission data are considered more stable, but it is possible that criteria for hospital admission can change. This also applies to mortality data, where procedures for how causes of death are recorded might change abruptly. In general, real-time data need to be adjusted due to the time lag in the reporting using now-casting methods (Bergström et al. 2022). The emergence and spread of novel genetic variants of the pathogen might also impact the risk of hospitalization and mortality. The ease of data collection should also be considered. For example, some data might require extensive preprocessing by third parties, e.g. telecom data, thus making the modelling process more vulnerable to external partners. In terms of the reproducibility of model results, it is preferable to choose data that are publicly available rather than protected. However, that goal will stand in contrast to GDPR protections. Data availability might also influence public trust in the model results (see the chapter on Communication).

2.3 Choice of model type

Mathematical models that describe disease transmission and other related variables such as hospital admissions broadly fall into three categories: compartmental models, statistical/time-series models, and individual-level models (Grimm et al. 2006). The latter category in turn contains three types of models that are similar: agent-based models (ABMs), individual-based models (IBMs), and micro-simulations. The definitions of ABMs and IBMs overlap to a large degree, and the latter term is more popular in ecology and in biology in general, whereas the former is more common in epidemiology and the social sciences. The key differences between ABMs/IBMs and micro-simulations are that

ABMs/IBMs are more theory-oriented, while micro-simulations are more policy-oriented; ABMs often use hypothetical populations, while micro-simulations typically use representative samples of real populations; and ABMs often use ad-hoc or roughly calibrated parameters, while micro-simulations rely more on empirically estimated processes (Richiardi 2014).

In compartmental models, the population is divided into two or more compartments that represent different disease states (e.g. Susceptible, Infected, and Removed, as in the classic SIR-model). In these models, it is also specified how individuals move between different states (compartments), which can occur either with a certain influence of randomness or can occur deterministically. It is also common to stratify the disease states into age groups and to represent geographically distinct populations using separate subcompartments.

Statistical (or time-series) models create forecasts by assuming that future data (for example, the daily incidence or the ICU occupancy) will follow a certain pattern in time (and possibly in space). For example, in the early stages of an epidemic we might assume that the curve that describes the daily incidence is exponential, and by fitting the model to historical data on incidence we can make predictions about the incidence in the near future. Statistical models can also make use of syndromic data in order to make predictions (e.g. tele-nursing calls as in REF). In this category we also include machine learning and AI models, which despite their hype have so far featured to only a limited extent in informing decision-making in public health. One can speculate about the reasons for this, but two main candidates might be a lack of training data, especially during the early stages of an epidemic, and the fact that forecasts, no matter how good they are, in themselves do not always help in decision-making without a good understanding of the processes that drive disease transmission.

Individual-level models, including network models, are similar to compartmental models, but instead of aggregating all individuals with the same disease state into a homogeneous group, each individual is described separately. This makes it possible to capture the demographics of the population in more detail and to incorporate heterogeneity to a greater extent in terms of contact patterns, disease development, and disease severity. However, it is important to note that a greater level of detail requires a greater number of parameters, which typically far exceeds the capacity of a compartmental model and, as discussed above, introduces additional uncertainty. However, this might not always be the case, and if one does not assign individual parameter values, but let some parameters take common values, one can in fact construct an ABM with few parameters. An additional drawback with this model type is that the computation time increases.

All three model types have their pros and cons, and these depend on the context in which the model is used and the purpose of the modelling exercise. An important consideration is the extent of disease-specific information required to produce reliable model predictions. Typically, all parameters of statistical models are estimated from historical data, whereas for compartmental and agent-based models it is common to estimate at least a subset of the parameters from the literature and expert opinion (e.g. the incubation and recovery times). This implies that statistical models, in comparison to other model types,

are straight-forward to deploy early during a pandemic/epidemic when disease-specific data are scarce and often uncertain. However, statistical models impose strong limitations on how the epidemic will evolve (e.g. an exponential increase or a single peak), which makes them unsuitable for longer-term predictions or for scenario projections. This also points to another important distinction between statistical models and compartmental/agent-based models. Statistical models rely on data, whereas the other model types are mechanistic in the sense that they represent (albeit in simplified form) the processes by which disease transmission occur. This makes the variables and parameters of such models interpretable, which is an advantage both during model development and during coding and when communicating the model to decision-makers and the public. In other words, statistical models tend to be black boxes, whereas mechanistic models are more transparent.

The use of an ABM can serve as a starting point for a study in order to get a feeling for the dynamics involved if one is not convinced that a standard compartment model is adequate or to assist in finding more realistic initial data. In the best of worlds, this might then lead to setting up other types of models in order to further increase both the understanding and the predictive power of the models and thus set the stage for better decision-making.

In terms of our three modelling cases presented above one might opt for a stochastic compartmental model or an ABM for Model case 1 given the small number of imported cases (compared to the total population). The impact of vaccination strategies on morbidity in Model case 2 could be assessed with a compartmental model where projections based on different scenarios could be contrasted. Lastly, the problem of forecasting hospital admissions in Model case 3 could be tackled with a statistical model.

2.4 Model parameters

All mathematical models come with parameters whose numerical values influence the model output. The number of parameters can range from one (e.g. the growth rate in an exponential growth model) to hundreds in complex individual-level models that describe everything from age-dependent risks to rates of movement between cities, or even to billions of parameters in machine-learning models. In statistical models, parameters rarely have a physical meaning, but rather serve as a means to connect model input and output. In compartmental models and individual-level models, the parameters typically correspond to some specific process (e.g. the probability of disease transmission in a given context), but parameters in simpler models usually aggregate many physical phenomena into a single number. For example, age-dependent transmission rates account for both varying rates of contact and heterogeneous probabilities of transmission upon contact.

There is also a temporal dimension to model parameters. Do we assume that parameter values are constant in time, or do we allow for multiple values separated by break points, or even parameters that depend continuously on time? The latter might make it easier to fit the model to data, but this will introduce additional uncertainty in the parameters.

2.5 Connecting the model and the data – the error model

The goal of a predictive model is to describe future data as accurately as possible given some initial data or event, or in the case of a scenario model to accurately describe possible future outcomes. This is achieved by calibrating (or fitting/training) the model on historical data, if available (see section 3.5). However, irrespective of how accurate our model is there will always be a difference between model output and historical data. In order to bridge this difference, it is often assumed that the actual outcome is given by the model output plus some randomness,

$$y_i = \widehat{y_i} + e_i$$

where the actual outcome is denoted y_i, the model output is denoted by $\widehat{y_i}$, and e_i is the assumed noise in the data, sometimes referred to as measurement noise. A common assumption is that e_i is an independent normal random variable because this is compatible with the theory of linear regression and maximum likelihood estimation. However, if the error is observed to depend on the magnitude of the model output, then other choices for modelling the error are suitable, e.g. a negative binomial distribution for the error (Alleman et al. 2023). In that case the error is not additive as in the example above and is instead included in the likelihood function that is used for model calibration (see section 3.5 for more details).

Chapter 3: Implementation

At this stage of the process a conceptual model has been decided upon and translated into an appropriate framework, and a choice has been made concerning what variables to consider and how they connect to the available data. The next step of the process is to implement the model and generate results. This will be the topic of this chapter, but before delving into the details it is worth mentioning that it is common to step back to the preparatory phase at some point during implementation, e.g. due to problems with model calibration. After adjusting the model structure or the data used for model calibration one can then move forward to generating model results.

3.1 Choice of programming language

Although some highly simplified models can be solved by hand, e.g. exponential growth, we still need to rely on computers for calibration and for visualization of model results, tasks that are crucial for accurate modelling and effective communication. This implies that programming will be an integral part of the modelling process.

We will not delve into the technical details of software development, but instead mention a couple of factors that are important when deciding upon a programming language. As a modeler you probably master more than one language, and most likely have a favourite language that you prefer over others. Using this language facilitates development of the code, but there are other factors that are equally important to take in account:

- If generating model output requires averaging over many realisations of a stochastic model, it is advisable to choose a language that is fast to execute. One such language that has emerged in recent years is Julia.
- Agent-based models are often more efficient and easier to implement in object-oriented languages. Although most languages support some form of object-oriented programming (e.g. Matlab and R) it is advisable to use languages like C++ or Python for those cases.
- The ability to share the code across members of the modelling team is important because the code might be developed in parallel with others or might be inherited by others (see chapter 5 on Management of models).
- Interactions with other software or databases might be simpler for certain languages compared to others.

Another aspect of code development, which is helpful to consider prior to coding, is the use of development platforms such as GitHub. These facilitate development by providing, for example, version control, bug tracking, and task management. In addition, the use of a development platform facilitates public access to the code and the reproducibility of the results.

3.2 Choice of solution method

At the core of the code that implements the model is the algorithm that solves the model and generates a numerical solution. For a statistical model this is trivial because it only amounts to plotting a curve with parameter values that have been calibrated from the data, but for compartmental models this part of the code is a bit more involved and requires a number of choices to be made.

Compartmental models, given in terms of a set of coupled ordinary differential equations (ODEs), are integrated forward in time with small step-sizes in order to generate a numerical solution. The future state of the model is given by the current state plus a small change. The most straightforward way to do this is to use the simple method known as Euler forward. However, this method usually needs very small time steps so as not to introduce numerical errors that can grow exponentially. To overcome this, one can use a more sophisticated numerical scheme like the Runge–Kutta method. An alternative to the deterministic methods is to use tau-leaping[1] where the small change is chosen at random depending on the current state of the model. It has been argued that the latter method is advisable when the number of individuals in each compartment is so small that random effects become important (Kratz et al. 2015). This occurs when compartmental models are stratified both with respect to age and space (with a high resolution, e.g. into parishes). If a deterministic approach is chosen, it is advisable to use built-in ODE solvers (e.g. *odeint* in Python or *desolve* in R) because these are faster to implement, reduce computational time, and lower the risk of coding errors.

For ABMs, which also contain an element of chance, it is common to use either discretised time and advance the model-state, e.g. by one hour, where the action of each agent is determined by probabilities that depend on model parameters, or to use the Gillespie algorithm, which, based on the parameters of the model, calculates the time to the next event and performs this event, e.g. moving an agent from home to workplace or advancing the disease state of an agent from exposed to infectious. The latter approach is more accurate but can be considerably more computationally expensive.

It should be noted that the difference between a highly stratified compartmental model solved using tau-leaping and an IBM is not crystal clear. Both models consider individuals that are shuffled (with an element of chance) between different disease states, and sometimes it is the structure of the code that makes a model individual-based (e.g. representing individuals using objects and actions with methods) rather than compartmental.

3.3 Coding

We will not go into details concerning the coding of the model because this activity is not specific to epidemiological models nor to models used for decision-making. Instead, we refer the reader to existing guidelines for scientific computing, e.g. (Wilson et al. 2014).

[1] A similar division into deterministic and stochastic solution methods can be made for time-discrete compartmental models.

However, we would like to mention that optimising code for speed within a given language is advisable.

3.4 Verification of the code

The verification step of the modelling process ensures that the model is correctly implemented, which is crucial for models that are to be used for decision-making. In other words, it is imperative that the code works as intended. Passing this step does not guarantee that the model is accurate or reasonable; it is only a formal check that the mathematical model is accurately implemented and solved by the code. The underlying conceptual model with its corresponding simplifications and limitations could still be "wrong" even though the code is verified.

There are no generally accepted common standards for code verification within the field of mathematical epidemiology. Although central for the later steps in the modelling process, it is rarely acknowledged as an independent activity and is rather lumped together with coding and is performed in parallel with the development of the code. This is, of course, sensible, but we would still like to highlight the importance of this step and point out that it should be carried out on the completed code.

There are no fool-proof methods for verifying epidemiological models, in contrast to finite-state models that can be formally verified (Weiss et al. 2001). Instead, we recommend applying several methods in parallel:

- Having the code reviewed by experts who are not involved in the development process.
- Investigating "expected" model behaviour using typical parameter values.
- If possible, comparing model output with existing models for the same system or co-developing multiple models that can be tested against one another.
- Ensuring that the solution is insensitive to minor changes in the time step.
- Checking that the solution conserves quantities that should remain constant (for example, in compartmental models without vital dynamics the sum of all compartments should always equal the population size).
- Investigating "extreme" model behaviour when one or more parameters are set to zero or tend to infinity. For example, if the infectivity in Model case 2 (see Preparatory work) is set to zero there should be no outbreak, whereas if it tends to infinity everyone should become infected.
- Using interactive debugging that makes it possible to track how variables in the code change during execution.
- Using pull requests on GitHub (when making incremental changes).
- Using unit testing where appropriate, i.e. testing the smallest functional unit of the code.

3.5 Calibration of the model

The act of making the model fit the data has many names, including training, calibration, fitting, parameter estimation, regression, and parameter inference. We will in the remainder of the handbook refer to this as (model) calibration, which refers to the act of adjusting the parameter values of the model to make the model output fit the existing data. This does not include adjusting or changing the structure of the model, in which case we are temporarily back in the Preparatory phase of modelling. However, we would like to mention that it is not uncommon to adjust the model structure in the light of data that the model cannot describe. Model development is thus typically not linear but loops back and forth between the different stages.

Typically, not all parameters of the model are adjusted, and the values of a subset are often taken from the literature instead. For example, in Model case 2, which concerns vaccination strategies, where we suggested a compartmental model, the serial interval or rate of recovery might be taken from previous studies whereas the infectivity can be estimated from existing data. In some cases there is no calibration carried out at all. For example, in Model case 1, on the importance of importations, we might not have any appropriate data to compare with and might have to rely solely on literature values and estimates from experts.

Model calibration can be carried out either using formal methods, informal methods, or a combination thereof. By informal methods we mean adjusting parameter values by hand in order to obtain a good visual resemblance between the data and the model output. Formal methods, on the other hand, make use of a quantitative measure of model fit, e.g. model error, which is then minimized (or in the case of a likelihood model is maximized, see below) using appropriate computational methods. In some cases, a subset of parameters are adjusted by hand while the remaining parameters are subject to formal calibration.

Two common methods for formal calibration are least squares estimation (LSE) and maximum likelihood estimation (MLE)[2]. In LSE the measure of model fit is the sum of the squared deviations between the N data points and the model output:

$$E(\theta) = \sum_{i=1}^{N} \left(D_i - M_i(\theta) \right)^2$$

where D_i is the ith data point, M_i is the corresponding model output, and θ is a vector containing all parameters of the model. The goal is to minimize the error $E(\theta)$ by adjusting the model parameters. This can be achieved using numerical optimisation (e.g. gradient-descent) methods or by brute force grid search, which evaluates the error for all possible combinations of parameter values. This can be very costly from a computational point of view but has the advantage of finding the optimal set of parameter values (Ugray et al. 2007). Numerical optimisation can, on the other hand, get stuck in local optima. In addi-

[2] If the error model is Gaussian (see below) the MLE and LSE of the parameters agree (García-Portugués 2025).

tion, a combination of the two approaches is also possible. Note that both methods suffer from the 'curse of dimensionality' (Saltelli and Di Fiore 2023) and require more computation as the number of parameters to calibrate grows.

In the MLE approach, the quantitative measure of model fit is the likelihood of observing the data. In order to find the optimal set of parameters, we view the likelihood as a function of the parameters and try to maximize the likelihood. The likelihood is essentially the probability of observing a certain sequence of data points given the model parameters and is calculated by taking the product of the individual probabilities:

$$L(\theta) = \prod_{i=1}^{N} P(D_i, \theta)$$

where again D_i are the data points and θ is a vector containing all parameters of the model. This time the model output defines the probability $P(D_i, \theta)$, which is given by the error model (see section 2.4 Model error). In a Gaussian error model, the model output M_i would appear as the mean while the variance would be unknown and would be determined by the deviations between the model and the data.

MLE is tightly connected to Bayesian methods for model calibration. In a Bayesian framework we do not associate a parameter with a single numerical value, but rather we think of parameters as random variables with certain distributions. Here, we distinguish between prior distributions, which are our best guess before the model is confronted with the data, and posterior distributions that are obtained by multiplying the prior distribution with the likelihood function. In many cases there is no closed expression for the likelihood, as is the case for most ODE models and ABMs, and one has to resort to approximation methods for obtaining posterior distributions, e.g. Markov chain Monte Carlo methods (Qian et al. 2003) or Approximate Bayesian Computation (Sunnåker et al. 2013).

Note that both LSE and MLE can be adapted to include multiple data sources and model outputs. This is achieved by adding additional terms in the sum of squared errors/likelihood to create a single error/likelihood function. These can also be given different weights depending on how we value model agreement with different data sources. When in doubt as to which method to use, it is often good to run both methods and compare the results. If the results differ significantly, further investigations will be needed, but as a rule of thumb LSE will be more reliable for smaller sample sizes.

Other techniques for model calibration are also available, e.g. history matching using emulators/surrogate models, which aims to discard implausible regions of the parameter space (Andrianakis et al. 2015), and approximate Bayesian computation, which makes use of numerical solutions and summary statistics in order to construct posterior distributions of the parameters (Minter and Retkute 2019).

When the model has been calibrated to the data, the remaining deviation between model and data, often termed the residual, is in theory attributed to the measurement error that is introduced via the error model. However, in reality the error is mainly due to the fact

that the model is a simplified representation of the system at hand, and therefore is usually biased in some manner.

3.6 Model validation

If the model is to be used for making predictions about future events and possibly in other locations, it is important to make sure that the model not only captures the current data well, but also captures future unseen data sufficiently well. This property is often evaluated by splitting the available data into one part that is used for calibration and a second, usually smaller, part on which model accuracy is evaluated. It is also common to calibrate the model using increasing time frames for the data in order to determine whether the error in the validation set changes and to see if the parameter estimates change as more data are used for calibration. If the latter occurs this is an indication that the model is unstable and might not perform well on unseen data.

Both the error used in LSE and the likelihood in MLE are unintuitive measures of model error. In the interest of transparency, it is common to consider measures that are easier to interpret for the validation step. Examples of this are the mean absolute percentage error (MAPE), which is a relative measure of error, and the root mean square error (RMSE), which is given by the square root of $E(\theta)$ and is an absolute measure of error. If the model output is probabilistic, accuracy can be evaluated using various scoring rules, e.g. the weighted interval score (WIS), the coverage probability, or the continuous ranked probability score (CRPS).

Evaluating model performance on data that are adjacent in time to the calibration data and on the same target population is sometimes called primary validation. This term is also used when the evaluation is carried out retrospectively. If the model evaluation is carried out prospectively, such that the model developers are blinded to the evaluation data or the evaluation is carried out in a different geographical region/population, it carries more weight and is often referred to as external or secondary validation (Ramspek et al. 2021). The possibility to carry out detailed model validation naturally depends on the available time frame for the modelling exercise and is also constrained by the available data, which might be scarce during the initial stages of an epidemic. As such, secondary validation should be viewed as an ideal rather than a standard.

In addition, we would also like to get an understanding on how sensitive our model is with respect to both variation in the data and variation in the parameter values. Such knowledge might provide better feedback to the data providers on what data need to be harvested with extra care. There are several ways to perform a sensitivity analysis (see e.g. (Wu et al. 2013)), and there are even ready-made packages in several languages, such as Matlab. We would also like to underscore that a sensitivity analysis not only provides a measure of how sensitive the model is, but that this procedure also provides information on what parts of the model need to be scrutinised or even adjusted.

3.7 Visualisation of model predictions

For the epidemiological models considered here, the model outputs are typically a time series of one or more disease-related variables. For example, for Model case 1 the output would be disease incidence in the absence or presence of one or more travel recommendations, whereas for Model case 3 it would be the most likely number of hospital admissions/day for a given hospital. When possible, it is common to plot the data on which the model was trained together with the forecast/scenario because this makes it possible for the recipient to gauge the accuracy of the model.

3.8 Handling uncertainty in model output

In order to communicate the uncertainty associated with model calibration it is common to not only plot the prediction given by the optimal parameter values, but also to include confidence intervals for the model output. In the case of simple statistical models (e.g. regression models) these can be calculated explicitly using built-in functions, and for non-linear regression the Delta-method can be utilized (Liu 2023). Compartmental models and ABMs require a bit more work, and usually confidence intervals are obtained by using the fact that the Hessian of $E(\theta)$ with respect to the parameters can be interpreted as the inverse of the covariance matrix of the parameters (Dalitz 2018). Sampling parameter values with the obtained covariance and solving the model provides an envelope of prediction from which a confidence interval can be calculated. A similar approach works for MLE where the covariance of the parameters is given by the inverse of the Fisher information matrix (Dalitz 2018). In a Bayesian setting, uncertainty can be visualized by sampling parameter values from the posterior distribution and solving the model to produce an envelope from which so-called credible intervals can be calculated. Please note that these approaches can underestimate extremes in model output (e.g. epidemic peak height), which are better accounted for using curve boxplots (Juul et al. 2021).

It should be noted that the uncertainty captured in confidence/credible intervals only reflects the difficulty of pinning down the correct parameter values and does not serve as quantification of the actual uncertainty of the predictions (Briggs et al. 2012). In other words, confidence intervals are conditioned on model choice and only reflect *parametric* uncertainty. Uncertainty regarding the implications of conceptualization, simplification, and model structure, often termed *structural* uncertainty, cannot be captured in visual representations of model output unless the results from several models with the same purpose are shown in tandem. Together with knowledge about the structure of the models it is possible for the recipient to understand how the model structure impacts the results. Even so, it is advisable to discuss structural uncertainty in the accompanying documentation, and this is even more important if a single model result is presented. Uncertainty in model output induced by uncertainty in the data should also be mentioned in the documentation.

To communicate uncertainty in scenario projections is a delicate task. If the model is calibrated using standard methods, then the projections could be visualised using confi-

dence intervals. However, this might signal that the uncertainty is bounded when in reality the projections rest on assumptions for that particular scenario, which might be highly uncertain or even speculative.

Another type of uncertainty, which is present in stochastic models, arises due to randomness in model dynamics. This implies that every simulation will generate a unique realization that differs slightly from previous ones, and therefore the typical model behaviour has to be assessed by compiling the distribution from a large number of simulations and has to consider the average value the model output. The variation in model output distribution is often referred to as *stochastic* uncertainty and is typically shown with error bars that correspond to, e.g., the standard error of the mean or some other quantification of the variability. This type of uncertainty should not be confused with parametric uncertainty and must be clearly labelled as stochastic uncertainty. If both stochastic and parametric uncertainty feature in the output, they should be summed to visualise the entire range of uncertainty. For deterministic models this uncertainty is expressed by adding a random error (see Section 2.4), which represents the underlying noise in the data. This is achieved using an error model. The uncertainty that results from both parametric uncertainty and noise in the data is described with a prediction interval, which equals the confidence interval of the parametric uncertainty *plus* the uncertainty induced by the unknown variance of the error model.

In summary, we have discussed five types of uncertainty related to model output that might be present:

- Structural uncertainty: arises from the model construction
- Parametric uncertainty: arises during model calibration
- Error model uncertainty: arises from estimated noise in the data
- Stochastic uncertainty: arises due to model simulations
- Scenario uncertainty: arises from assumption about future developments

Note that biases and assumptions concerning the relationship between input data and model variables propagate to the model output and thus should be included in the overall discussion of uncertainty.

Chapter 4: Communication

Mathematical models of infectious diseases can be used in many contexts and for a wide range of purposes. Despite this diversity, a common factor is that the results of the model, and to some extent the model itself, have to be communicated to recipients external to the group that has developed the model, e.g. the decision-makers.

As has been expressed above, communication with decision-makers is important to ensure a high quality of the entire modelling process, and in particular to create an understanding between the modelers and decision-makers concerning the quality and limitations of the data. However, the communication of the concrete modelling results deserves extra care.

The shape and form of the communication will depend on the purpose of the model, who the recipients are, the nature of the content, and lastly the numerical result itself. Communication of model results is related to the concept of transparency in science, and the present chapter is inspired by a taxonomy of transparency introduced by the philosopher of science Kevin C. Elliot (Elliott 2022). We will first introduce the different dimensions of communication and then illustrate these with three examples.

4.1 The purpose

In order to effectively communicate the results from a model one has to be clear as to why they need to be communicated. This is tightly connected to the purpose of the model or the question that the modelers have been tasked to answer. As suggested in the chapter on Prepatory work, it is important for modelers to have a clear picture of the purpose prior to model construction because this will guide the work going forwards. However, the purpose should also guide communication.

4.2 The audience

The second dimension that should influence the communication are the recipients. These can include other modelers, scientists, decision-makers, politicians, and the general public, which all have different needs, expectations, and competencies. For example, a presentation for the general public will be less efficient if the main focus is on technical details, whereas other modelers expect the model results to be communicated at a level of detail such that they become reproducible. In some instances, the purpose might be to educate the public about mathematical modelling, which requires yet another mode of communication.

A detailed discussion concerning all of the model assumptions and associated uncertainties might not be appropriate to communicate to the public due to the risk of causing misunderstandings because people might not be aware that all models require assumptions and simplifications. However, this information is critical if decision-makers are the intended audience.

As described above there is often a primary audience and a secondary audience, and both should be considered.

4.3 The content

With the purpose and intended audience fixed, the task of determining *what* to communicate becomes easier. As a bare minimum we consider that model results in terms of plots or tables or some other visual representations should be communicated. Although rare, there might be situations where visual aids are not necessary, e.g. if the question posed is a simple yes/no question (e.g. Should we close national borders to reduce disease transmission given some set of criteria?). But even in such cases it is probably of interest to the recipient to understand how the conclusion was reached.

At the other end of the spectrum, we have a full-fledged technical report accompanied by an online repository that contains fully documented code and all of the raw data required to reproduce the model results. Of course, it is also possible to communicate model results in a multi-modal fashion where an oral presentation of the key results is accompanied by a detailed report. In terms of the contents of technical reports, the EPIFORGE guidelines provide excellent guidance as to the necessary components (Pollett et al. 2021). Another option is to present the primary results and conclusions in the main body of the report and relegate the technical details to an appendix.

4.4 The mode of communication

The three dimensions discussed above should guide the mode of communication, but it is likely that the communication is also influenced by routines specific to the organization carrying out the modelling. However, there are a couple of aspects that are important to keep in mind:

- Communication of pandemic model results that contain forecasts in the short term (e.g. 1–2 weeks) are likely to change rapidly as more data become available. In such cases online dashboards that can easily be updated might be preferable.
- Making model code available to the public opens up for criticism, which might appear time-consuming and unpleasant to digest, but is important for transparency and also has the potential to improve and facilitate model development.
- If scientists external to the organization have co-developed the model, it should be clear from the outset who is responsible for communicating the results.
- Presenting model results in a forum where the recipients are comfortable to ask question facilitates the uptake of the results. This is particularly important when recipients lack technical skills and are expected to pass on the information to local decision-makers.

4.5 Dangers of communication

Although a well-planned strategy for communicating model results is likely to increase the uptake and usefulness of the modelling exercise, communication also comes with

dangers. The most obvious one is that careful documentation and communication might waste scarce resources and thus might slow down further model development. This problem is particularly relevant in the early stages of an epidemic when there may be a lack of time and the data could be especially unreliable and are being updated at a rapid pace. Another important aspect is the art of finding the right level of detail to describe the models for a general audience – not too over-simplified and thus underestimating the reader, but on the other hand not using too much technical jargon and details so as to make the model virtually inaccessible. Lastly, public access to data might violate privacy, in particular if it is of high spatial and/or temporal resolution.

Another thing to consider is how one should act if model results clearly conflict with other reported modelling results. This situation is of course not uncommon in academia, but in the case of a public health crisis diverging results may be utterly confusing for both decision-makers and the public, and such situations have created significant tension between different research groups in the past. Here, it is important to be clear about the discrepancies and to suggest possible reasons for them. However, in a crisis situation the process is considerably sped up and is played out on a public stage for any interested parties to observe.

4.6 Examples

We now present three different examples of communication that involve the Model cases presented above. These were chosen to inhabit different positions in the space spanned by the dimensions described above.

Model case 1, Travel importations: The purpose of this modelling exercise is to provide national decision-makers (i.e. politicians) with advice concerning the effect of travel recommendations on disease incidence. Because the purpose and audience are fixed, we now turn to the question of content. One possible way of illustrating the effect of travel recommendations is to contrast curves of disease incidence in the presence and absence of recommendations for international travellers (see e.g. (Godin et al. 2021)). Such a comparison will naturally rest on assumptions concerning the number of imported cases and reproduction number in the future and the current local incidence that the scenarios concern. These are therefore important to communicate alongside the results.

Model case 2, Vaccination strategies: Here the purpose is to assess the impact of different vaccination strategies in terms of the morbidity of an epidemic. We assume that the model results will be communicated to the public prior to implementation, with the intention of increasing vaccine uptake. This problem can of course be tackled in many ways, and here we provide one possible course of action. As in the previous example it is useful to contrast the effect of no vaccination with different strategies, preferably in some visual form. Additionally, it is important to stress that the model output is a scenario that corresponds to possible future outcomes and is not necessarily the single most likely outcome. If the results are presented at a press conference, then they preferably should be accompanied by a technical report, published prior to the event, that details both the preparatory phase of model development and the implementation phase. The code should also be made publicly available. In this case, the timing of communication could

also be of importance. If the results are communicated too early, before the epidemic is perceived as a threat, the public might not be interested, and if they are communicated after the implementation of the vaccination programme the media and the public will likely not understand the basis for the decision.

Model case 3, Forecasting hospital admissions: In this case the purpose is to inform local decision-makers in healthcare regarding future hospital admissions. Here we assume that the intended audience is the decision-makers and that the model results will not be made public. Because decision-makers are often concerned with how much the model results can be trusted, it is useful to communicate such forecasts by illustrating the most likely outcome together with 95% confidence/credible intervals. This parametric uncertainty should be accompanied by a discussion of the uncertainty that stems from model assumptions (including future changes of the underlying conditions), the model structure, and the data used for model calibration. This more complete view will aid decision-makers in their balancing of different facts and help them to reach well-informed decisions. If the model is continuously updated, it is advisable to keep track of model accuracy, e.g. in terms of the coverage probability (Cramer et al. 2022), which is a relevant metric for the recipient.

Chapter 5: Management of models

The modelling process is, as we have discussed, an interplay between policy questions posed by decision-makers and a modelling group with knowledge in mathematics, data science, and epidemiology. To be successful, this process requires skilled management, but it is also relevant to discuss what happens with a model after it has fulfilled its immediate need.

To set the stage, we have now reached a point where a model has been constructed with a specific purpose in mind. It has been implemented and the result have been communicated to the intended recipient. This chapter will deal with the subsequent step, namely how the model will be managed within the organization for potential further development and reuse. Although effective management of models might not impact the quality of models at the time, it has the potential to impact the efficiency of the modelling group in the longer term thus leading to overall improved quality of the model.

5.1 Changes in needs

Certain modelling tasks stretch over long periods of time, whereas other have a short lifespan. For example, the task of producing forecasts of hospital admissions during an epidemic (Model case 3) is likely to be ongoing as long as the epidemic puts pressure on the healthcare system. This contrasts with the purpose of estimating the effect of travel importations (Model case 1), which is likely to be of interest to decision-makers only during a short time window. A long-lasting task requires that more attention is paid to model development, and if the rough time span of the task is known from the outset then this should impact the entire modelling exercise. If we know that the model will be calibrated multiple times on a growing time-series, and that many similar figures and reports will be communicated, then it makes more sense to invest time and resources in automated pipelines for data and model output. This will also make the reader feel at home and thus be able to more easily compare the graphical outcome between the different consecutive reports. On the other hand, if we know beforehand that the model will only be used once, then less attention can be paid to automating the workflow.

5.2 Model documentation

We have already mentioned documentation in relation to communication, and that discussion focused on external recipients and their needs and desires. The potential for model development and reuse also depends on communication, but this time communication is internal to the modelling group. For example, the structure of data pipelines and other workflows connected to producing model results are not of interest to external users, but this knowledge facilitates continuous use, or reuse, of the model, be it by the same modeler or by someone who is new to the project. We would also like to underline the easily underestimated importance of the proper documentation of the code, especially when time is limited. Descriptive names for variables and functions, removal of unused code, and detailed comments within the code are aspects that accelerate reuse. However, it is also worth considering the trade-off between the effort it takes for someone

to understand inherited code and the time it takes to develop new code. If the life span of the model is expected to be substantial, one must consider the durability of its storage, e.g. backed-up hard-drives or other media that is insensitive to organizational changes. Also worth considering is the risk that the code can become unusable due to dependencies on methods and functions that become deprecated, although this problem can be tackled by creating a reproducible computational environment (e.g. using *renv* in R).

5.3 Continuity among staff

Because models are commonly developed by single individuals, they will not persist within an organization unless they are known and properly documented. Turnover of staff poses a serious threat to organisational learning and is an important factor to consider when managing a modelling group. In exceptional situations, such as during a pandemic, when modelling groups are under pressure to deliver above normal capacity, the introduction of new members of staff can be particularly challenging. A possible solution to this problem is to engage academics in model development during non-pandemic time. This can, for example, be achieved through practical exercises where modelers from public health agencies collaborate with academics in scenarios that describe future epidemics. Such activities can foster a sense of community and personal relationships that hopefully will aid collaborations during public health crises.

5.4 Privacy and security issues

In most cases, infectious disease models rely on data that are not subject to privacy or security concerns. However, in some cases these matters need to be considered. For example:

- When the model uses data that contain sensitive personal information.
- When the model makes use of classified data or spatial data of high spatial resolution where single individuals might be identifiable.
- When model results are sensitive and are related to national security.

Chapter 6: Conclusions

In this brief handbook we have tried to describe the central elements of modelling infectious diseases, with an emphasis on the use of modelling for decision-making purposes. The exposition has a chronological structure, moving from the preparatory phase, to implementation, communication, and lastly the management of models. However, as most people who have been involved in mathematical modelling of some form know, the path from idea to finished project is seldom straight and is rather cyclical in nature. Nevertheless, we believe that we have captured all the relevant parts of the process.

In this short format it is impossible to cover all methods and techniques used in infectious disease modelling, and this handbook is therefore far from comprehensive and should only serve as an introduction to the topic. However, we hope that the references provided in the text can serve as a means of learning more about the subject.

We hope that this handbook will stand the test of time and that it will turn out to be useful during coming public health crises. At the time of writing, AI and machine learning techniques are yet to be used at an appreciable scale in infectious disease modelling. One obvious explanation for this is that at the beginning of a new pandemic there will always be a lack of data for model calibration and validation. This is, of course, even more critical for machine learning models because they require massive amounts of data for calibration. An additional possible explanation for the low usage of machine learning techniques is the lack of transparency afforded by such methods. Given the current developments in AI, this might change soon with advances in so-called Explainable AI. However, even in the case of new modelling types or numerical methods the strategic methodology presented in this handbook will be applicable.

Mathematical modelling remains an important tool for understanding and providing a basis for decision-making during epidemiological crises. We believe that the knowledge and advice collected in this handbook will serve as useful guidelines in all aspects of such endeavours and, most importantly, will assist society in better managing the threat of future epidemics.

Chapter 7: Checklist

In order to condense the material presented in this handbook and to make the steps of the modelling process described herein applicable and useful during actual modelling for decision-making, we have formulated the following checklist. It is not a list of reporting items, such as the EPIFORGE guidelines (Pollett et al. 2021), rather it is meant to serve as a tool during model development in order to ensure that all the necessary steps and modelling choices have been carried out and considered. Like the handbook, they are presented in chronological order. Each item roughly corresponds to a subsection of the handbook and contains a question and a textbox for the answer.

Item	Section in handbook	Answer
1.1 What data does the model utilize and in what way?	2.2	
1.2 What uncertainty is associated with the data?	2.2	
1.3 What type of model is it?	2.3	
1.4 What parameters does the model contain?	2.4	
1.5 Which error model is used?	2.5	
2.1 In what programming language is the model coded?	3.1	
2.2 How is the model solved or simulated?	3.2	
2.3 What methods have been used to verify the code?	3.4	
2.4 What method is used for model calibration?	3.5	
2.5 How was the model output validated?	3.6	
2.6 What types of uncertainty are present in the model?	3.8	
2.7 How is the model uncertainty communicated in the documentation and visualization?	3.8	
3.1 Is the purpose of the model stated in the communication?	4.1	
3.2 What is the intended audience of the communication?	4.2	

3.3 Is the level of detail in the communication appropriate for the audience?	4.3	
3.4 What are the potential dangers of communicating the model's results?	4.5	
3.5 If there are dangers, how are they handled?	4.5	
4.1 What is the expected lifespan of the model?	5.1	
4.2 Is the model documentation appropriate for the lifespan?	5.2	
4.3 Are there any issues with privacy or (national) security concerning the data or the model?	5.4	

Chapter 8: References

Alleman, Tijs W., Michiel Rollier, Jenna Vergeynst, and Jan M. Baetens. 2023. 'A Stochastic Mobility-Driven Spatially Explicit SEIQRD COVID-19 Model with VOCs, Seasonality, and Vaccines'. *Applied Mathematical Modelling* 123 (November): 507–25. https://doi.org/10.1016/j.apm.2023.06.027.

Andrianakis, Ioannis, Ian R. Vernon, Nicky McCreesh, et al. 2015. 'Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda'. *PLoS Computational Biology* 11 (1): e1003968. https://doi.org/10.1371/journal.pcbi.1003968.

Bergström, Fanny, Felix Günther, Michael Höhle, and Tom Britton. 2022. 'Bayesian Nowcasting with Leading Indicators Applied to COVID-19 Fatalities in Sweden'. *PLOS Computational Biology* 18 (12): e1010767. https://doi.org/10.1371/journal.pcbi.1010767.

Bjornstad, Ottar N., Barbel F. Finkenstadt, and Bryan T. Grenfell. 2002. 'Dynamics of Measles Epidemics: Estimating Scaling of Transmission Rates Using a Time Series SIR Model'. *Ecological Monographs* 72 (2): 169. https://doi.org/10.2307/3100023.

Briggs, Andrew H., Milton C. Weinstein, Elisabeth A.L. Fenwick, Jonathan Karnon, Mark J. Sculpher, and A. David Paltiel. 2012. 'Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6'. *Value in Health* 15 (6): 835–42. https://doi.org/10.1016/j.jval.2012.04.014.

Cramer, Estee Y., Evan L. Ray, Velma K. Lopez, et al. 2022. 'Evaluation of Individual and Ensemble Probabilistic Forecasts of COVID-19 Mortality in the United States'. *Proceedings of the National Academy of Sciences* 119 (15): e2113561119. https://doi.org/10.1073/pnas.2113561119.

Dalitz, Christoph. 2018. 'Construction of Confidence Intervals'. arXiv:1807.03582. Preprint, arXiv, July 10. https://doi.org/10.48550/arXiv.1807.03582.

Diekmann, Odo, Hans Heesterbeek, and Tom Britton. 2012. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press.

Egger, Matthias, Leigh Johnson, Christian Althaus, et al. 2018. 'Developing WHO Guidelines: Time to Formally Include Evidence from Mathematical Modelling Studies'. *F1000Research* 6 (February): 1584. https://doi.org/10.12688/f1000research.12367.2.

Elliott, Kevin C. 2022. 'A Taxonomy of Transparency in Science'. *Canadian Journal of Philosophy* 52 (3): 342–55. https://doi.org/10.1017/can.2020.21.

FDA. 2016. ‘Reporting of Computational Modeling Studies in Medical Device Submissions’. *Https://Www.Fda.Gov/Regulatory-Information/Search-Fda-Guidance-Documents/Reporting-Computational-Modeling-Studies-Medical-Device-Submissions*.

García-Portugués, Eduardo. 2025. *Notes for Predictive Modeling*. https://bookdown.org/egarpor/PM-UC3M/app-ext-mle.html.

Gerlee, Philip, and Torbjörn Lundh. 2016. *Scientific Models*. Springer International Publishing. https://doi.org/10.1007/978-3-319-27081-4.

Godin, Arnaud, Yiqing Xia, David L Buckeridge, et al. 2021. ‘The Role of Case Importation in Explaining Differences in Early SARS-CoV-2 Transmission Dynamics in Canada—A Mathematical Modeling Study of Surveillance Data’. *International Journal of Infectious Diseases* 102 (January): 254–59. https://doi.org/10.1016/j.ijid.2020.10.046.

Grimm, Volker, Uta Berger, Finn Bastiansen, et al. 2006. ‘A Standard Protocol for Describing Individual-Based and Agent-Based Models’. *Ecological Modelling* 198 (1–2): 115–26. https://doi.org/10.1016/j.ecolmodel.2006.04.023.

Juul, Jonas L., Kaare Græsbøll, Lasse Engbo Christiansen, and Sune Lehmann. 2021. ‘Fixed-Time Descriptive Statistics Underestimate Extremes of Epidemic Curve Ensembles’. *Nature Physics* 17 (1): 5–8. https://doi.org/10.1038/s41567-020-01121-y.

Keeling, Matt J, and Ken T.D Eames. 2005. ‘Networks and Epidemic Models’. *Journal of The Royal Society Interface* 2 (4): 295–307. https://doi.org/10.1098/rsif.2005.0051.

Kratz, Peter, Etienne Pardoux, and Brice Samegni Kepgnou. 2015. *Numerical Methods in the Context of Compartmental Models in Epidemiology*. Version 1. https://doi.org/10.48550/ARXIV.1501.03645.

Liu, Enwu. 2023. *Delta Method for Predictive Models*. https://rpubs.com/enwuliu/1103572.

Minter, Amanda, and Renata Retkute. 2019. ‘Approximate Bayesian Computation for Infectious Disease Modelling’. *Epidemics* 29 (December): 100368. https://doi.org/10.1016/j.epidem.2019.100368.

Pollett, Simon, Michael A. Johansson, Nicholas G. Reich, et al. 2021. ‘Recommended Reporting Items for Epidemic Forecasting and Prediction Research: The EPIFORGE 2020 Guidelines’. *PLOS Medicine* 18 (10): e1003793. https://doi.org/10.1371/journal.pmed.1003793.

Qian, Song S., Craig A. Stow, and Mark E. Borsuk. 2003. ‘On Monte Carlo Methods for Bayesian Inference’. *Ecological Modelling* 159 (2–3): 269–77. https://doi.org/10.1016/S0304-3800(02)00299-5.

Ramspek, Chava L, Kitty J Jager, Friedo W Dekker, Carmine Zoccali, and Merel Van Diepen. 2021. 'External Validation of Prognostic Models: What, Why, How, When and Where?' *Clinical Kidney Journal* 14 (1): 49–58. https://doi.org/10.1093/ckj/sfaa188.

Richiardi, Matteo. 2014. 'The Missing Link: AB Models and Dynamic Microsimulation'. In *Artificial Economics and Self Organization*, edited by Stephan Leitner and Friederike Wall, vol. 669. Lecture Notes in Economics and Mathematical Systems. Springer International Publishing. https://doi.org/10.1007/978-3-319-00912-4_1.

Saltelli, Andrea, and Monica Di Fiore, eds. 2023. *The Politics of Modelling: Numbers Between Science and Policy*. 1st edn. Oxford University PressOxford. https://doi.org/10.1093/oso/9780198872412.001.0001.

Sunnåker, Mikael, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. 2013. 'Approximate Bayesian Computation'. *PLoS Computational Biology* 9 (1): e1002803. https://doi.org/10.1371/journal.pcbi.1002803.

Ugray, Zsolt, Leon Lasdon, John Plummer, Fred Glover, James Kelly, and Rafael Martí. 2007. 'Scatter Search and Local NLP Solvers: A Multistart Framework for Global Optimization'. *INFORMS Journal on Computing* 19 (3): 328–40. https://doi.org/10.1287/ijoc.1060.0175.

UK Government. 2025. 'The Aqua Book'. *Https://Www.Gov.Uk/Guidance/the-Aqua-Book*.

Weinstein, Milton C., Bernie O'Brien, John Hornberger, et al. 2003. 'Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies'. *Value in Health* 6 (1): 9–17. https://doi.org/10.1046/j.1524-4733.2003.00234.x.

Weiss, M.W., S.C. Seth, S.K. Mehta, and K.L. Einspahr. 2001. 'Design Verification and Functional Testing of Finite State Machines'. *VLSI Design 2001. Fourteenth International Conference on VLSI Design*, 189–95. https://doi.org/10.1109/ICVD.2001.902659.

Wilson, Greg, D. A. Aruliah, C. Titus Brown, et al. 2014. 'Best Practices for Scientific Computing'. *PLoS Biology* 12 (1): e1001745. https://doi.org/10.1371/journal.pbio.1001745.

Wu, Jianyong, Radhika Dhingra, Manoj Gambhir, and Justin V. Remais. 2013. 'Sensitivity Analysis of Infectious Disease Models: Methods, Advances and Their Application'. *Journal of The Royal Society Interface* 10 (86): 20121018. https://doi.org/10.1098/rsif.2012.1018.

Handbook of mathematical modelling of infectious diseases for decision-making

This handbook offers a clear and practical guide to how models can support public-health decision-making before, during, and after an epidemic. Developed through a national collaboration between the Public Health Agency of Sweden, Chalmers University of Technology, the Swedish Defence Research Agency, and the Swedish Armed Forces, the handbook distils lessons learned from the COVID-19 pandemic and highlights the importance of transparent, well-communicated modelling.

The book provides a structured walk-through of the full modelling process: from defining the purpose of a model and choosing data and model types, to implementation, calibration, validation, communication, and long-term model management. It emphasises the interplay between modellers and decision-makers, showing how clear objectives, realistic assumptions, and awareness of uncertainty are essential for models to be useful in real-world settings.

Rather than prescribing a single modelling approach, the handbook illuminates the choices, trade-offs, and practical considerations that shape every modelling project. It is intended for early-career modellers, academics entering applied infectious-disease modelling, and practitioners in public health who rely on model-based evidence. Senior modellers will also find it a valuable foundation for harmonising both practices and nomenclature, and strengthening collaboration across institutions.