**A practical introduction to**

**Mapprochesveterinary disease transmission modeling**

**Carsten Kirkeby1\*, Victoria Brookes2,3, Michael P. Ward4, Salome Dürr5, Tariq Halasa1**

1Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, 1870 Frederiksberg, Denmark

2School of Animal and Veterinary Sciences, Faculty of Science, Charles Sturt University, Wagga Wagga, NSW 2678, Australia

3Graham Centre for Agricultural Innovation (Charles Sturt University and NSW Department of Primary Industries), Wagga Wagga, NSW 2678, Australia

4Sydney School of Veterinary Science, University of Sydney, Camden, NSW, Australia

5Veterinary Public Health Institute, University of Bern, Liebefeld, Switzerland

**\*Correspondence:**

Carsten Kirkeby

Ckir@sund.ku.dk

Keywords: Simulation model, transmission model, disease dynamics, equation-based model, mechanistic model.

**Abstract**

Disease transmission models are frequently used in veterinary science to simulate disease spread, predict impact of different surveillance or control strategies, and to provide insights about disease causality via comparison of model outputs with real life data. We here list some of the most popular approaches and go further into detail with mechanistic simulation models (MSMs). We describe the modeling procedure before, during and after the programming stage. We focus on mechanistic modeling in the programming stage and include code examples. We describe how the disease transmission model should be verified and validated after it has been built, and important concepts such as convergence and sensitivity analysis. Lastly we provide an overview of recent developments and concluding remarks.

# Introduction

## 1.1 The purpose of disease transmission models

A disease transmission model is a simplified representation of a real-life system of disease dynamics. Disease transmission models can be more broadly termed simulation models, because they simulate a disease system. Because these systems are usually highly complex, the often used quote during discussions of models, ‘all models are wrong, but some are useful’ (Box, 1976) is of particular relevance. Disease transmission models are motivated by a need to 1. better understand a system’s dynamics; 2. predict a system’s outputs (including the evaluation of different strategies to improve surveillance and control of epidemics or endemic diseases); and 3. study how the outputs of a system can be influenced. They can represent diverse disease ecologies (bacterial and viral infections, parasites and vector-borne diseases) in a range of hosts (Mancy et al., 2017). They are useful when experiments on disease transmission and control are not possible – for example due to high resource requirements (such as time and monetary costs) or logistical and ethical constraints (such as the investigation of exotic diseases) – or would not sufficiently represent real-life disease ecology.

In the veterinary field, models of disease transmission are developed for both theoretical insight and practical application, depending on the end-user and the intended purpose of the information generated by the model. Researchers might want a model to extend our theoretical understanding of a particular disease ecology; for example a model could be used to reconstruct a system using limited available knowledge, to identify critical elements and knowledge gaps (Singer et al., 2011). Once knowledge gaps in the system are identified, empirical data can be collected – for example, by fieldwork or *in vitro* experiments – to further develop the model to inform practical recommendations. Practical applications of models might include advising decision makers on surveillance or control of a specific disease by comparing strategies in contexts such as outbreak situations (Keeling et al., 2001; Zinsstag et al., 2009), outbreak preparedness (e.g. Bates et al., 2003; Halasa et al., 2019; Martinez-Lopez et al., 2010; Tildesley and Keeling, 2008; Backer et al., 2012; Szmaragd et al., 2009; Ward et al., 2009; Dürr and Ward, 2015; Dürr et al., 2013), and the control of endemic pathogens (e.g. Allore et al., 1998; Østergård et al., 2005; Kudahl et al., 2007; Ezanno et al., 2007; Marce et al., 2010; Lu et al., 2010; Groenendaal et al., 2002; Brookes et al., 2015; Kirkeby et al., 2016; Gussmann et al., 2018; Zingg et al, 2017). Consequently, there are different types of disease transmission models, with different aims and structures. An often-used model type is mechanistic models, which can be used to mimic a disease system in great detail. These models are process-based, usually comprise stochastic events and are useful for exploring disease systems via simulation of complex mechanisms that are considered *a priori* important for disease spread. An example of complex mechanisms is heterogeneity of contacts between individuals (Brookes et al., 2019; Laager et al., 2018). As a consequence, these models often have a high level of complexity and therefore are computationally intensive (hence slow to implement and simulate). Another type is equation-based models (deterministic or stochastic), useful for exploring general patterns within a system. Overall, the type of disease transmission model used depends on the requirements of the end-user, the ecology of the system, the disease epidemiology, the technical expertise of the modeler and the data available.

Mancy et al. (2017) provide a discussion of the different motivations for developing models in ecology and present a conceptual framework to guide model construction, focusing on the pre-modeling stage (model selection, establishing and testing the theory). In building on Mancy et al. (2017) our objectives are three-fold; 1) to provide a general guide to developing a mechanistic model of animal disease transmission; 2) to describe important concepts before, during and after developing a model of animal disease transmission; and 3) to provide practical examples of the most commonly used models in veterinary science. We focus on modelling disease transmission and control in the veterinary field, and specifically, we provide simple examples (including code) of mechanistic simulation models (MSMs).

# Methods

## 2.1 Definitions and concepts

We describe approaches applied in disease transmission modeling, including essential definitions and concepts for each approach. Our intention is not to make a comprehensive review, but mention some often used approaches, and give practical introduction on how to start modeling with MSM.

### 2.1.1 Terms used in disease transmission modeling

Disease transmission models in the veterinary field represent the dynamics of transmission of an infectious disease between the modeled *units of interest*. This unit is the smallest entity of the model and could be an individual animal (or part of it; for example, a quarter of the udder in a mastitis model), a group of animals, herds, or populations in regions or countries. Most commonly, the simulated system includes time as a variable, making it a *dynamic* model. The way in which time is modeled in the dynamic system can be *continuous* or *discrete* (time-steps). In the latter case, the length of a time step depends on the disease and purpose of the model. Daily time-steps are typical for most infectious disease models (e.g. Bates et al., 2003; Backer et al., 2012; Kirkeby et al., 2016), but weekly (e.g. Østergård et al., 2005; Kudahl et al., 2007) or biweekly (e.g. Halasa et al., 2010), or biannual (e.g. Groenendaal et al., 2002) or even yearly time steps can be used (for example, when simulating long duration control programs, such as Zingg et al., 2017).

A model can either be *deterministic* or *stochastic*. A model is *deterministic* if there is no variation within the model (no random processes and model parameters can only take fixed values) and consequently, no variation in the model outputs. A model is *stochastic* when there is randomness arising from the use of distributions − rather than fixed values − to describe input parameters, or by allowing model events to occur as random processes (inherent randomness).

Disease transmission models usually represent the progression of the modeled unit of interest through *disease states*, for instance *Susceptible (S)*, *Infectious* (*I*) and *Recovered (R)* states (an SIR model). The modeled states are dependent on the disease (for diseases with a longer subclinical than clinical period it is recommended to add a subclinical, infectious state) and the purpose of the model (for example, differentiation of clinical and subclinical infectious state is mainly important when disease transmission differs between the two states or when clinical detection of the disease is an essential aspect in the model).

*Homogeneous* or *heterogeneous* contacts (mixing) between the units of interest can be modeled. *Homogeneous* contact means that all the modeled units have equal probability of contact with each other (no clustering). *Heterogeneous* contact means that the probability of contact between units of interest is not equal, hence clustering (spatial or related to other contact characteristics) exists in the population.

### 2.1.2 A variety of modeling approaches

Since Kermack and McKendrick first formulated the basic compartmental SIR model using differential equations in 1927 (Kermack and McKendrick, 1927), numerous approaches to modeling disease transmission have been developed. One popular approach is mechanistic modeling, which is described in detail below. Other approaches include equation-based models (EBM), or mathematical models, that use either *differential* or *difference* equations to simulate transition between disease states, possibly using a Gillespie or tau-leap algorithm (Keeling 2008; Smith et al.,2015; Hunter et al., 2018). These models are often not stochastic, although differential equations can also be extended to provide stochastic output (Protter, 2005).

Another approach is using next generation matrix (NGM) models, which are a modification of the classic compartmental model, to deduce the basic reproductive number (Diekmann et al., 1990). For a practical introduction to the use of NGMs, see Diekmann et al., (2010).

Some models are spatially explicit, mimicking a system where the transmission varies with spatial location (Alkhamis et al. 2016; Mur et al. 2018). For instance, the transmission of disease could rely on spatial features (see also section “Modeling disease transmission” below).

Recently, Bayesian (network) model frameworks have been developed to simulate diseases in populations (Green et al. 2010; Hedell et al., 2019). In these models, the parameterization of the model can be more relaxed because important estimates like sensitivity and specificity of diagnostic tests can be estimated through the model. However, they then rely on accurate datasets similar to the modeled system to be available (Engblom et al. 2019).

### 2.1.3 Mechanistic simulation models (MSM)

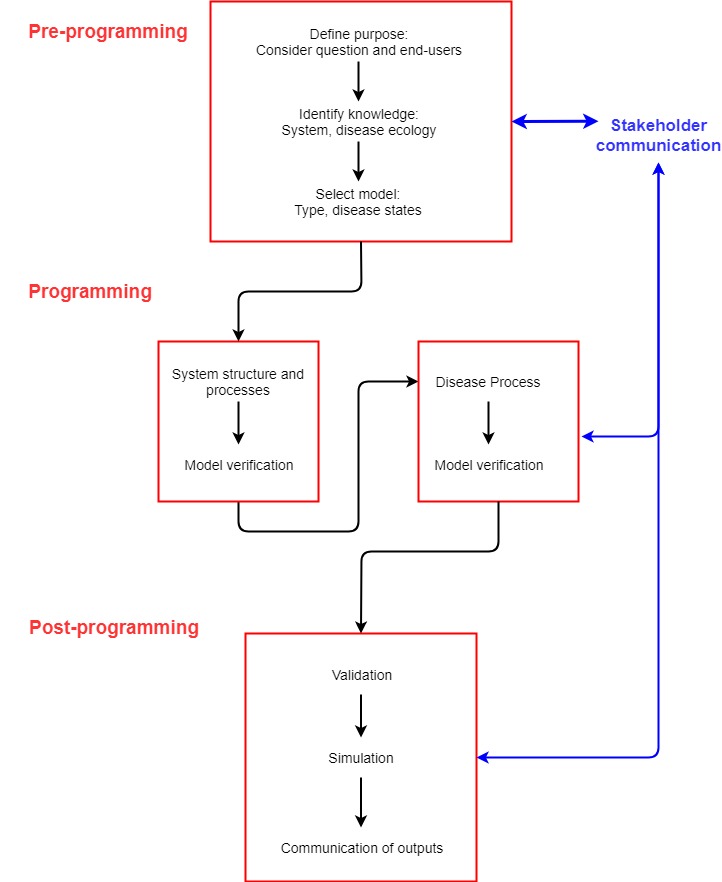
Mechanistic models (also known as process-driven models) use discrete time steps. “Mechanistic” can be defined as “thinking of living things as if they were machines” (<https://dictionary.cambridge.org/dictionary/english/mechanistic>, accessed 16/09/2019). In biology, and in the context of disease transmission modeling, these models describe the processes and mechanisms of disease transmission between units (Smieszek, 2009). Mechanistic models are useful in the veterinary field because they simulate disease transmission heterogeneity between pairs of units of interest. The units of interest in MSMs represent discrete entities, such as herds or individual animals. The advantage of such models is that the individuals can have their own properties that can influence disease transmission, detection or control. For example, in a model of foot-and-mouth disease (FMD), an individual herd might be predominantly sheep or cattle, which might influence disease susceptibility and transmission at the herd level and be important to incorporate in a model (Mardones et al., 2010; Ward et al., 2009). In another example in which rabies transmission is modeled, individual dogs were assigned roaming characteristics that influenced their contacts with other dogs (Hudson et al., 2019). These models can all be referred to as *individual-based models* *(IBMs*, or Individual-level models, ILM (Mahsin et al. 2019)*),* but some are also *agent-based models (ABMs)*. Both are mechanistic and model the individuals within the system; however, ABMs simulate contact − and hence disease transmission − between explicit pairs of individuals. Thus the agents (individuals) in the model interact with each other, which includes the inherent heterogeneity between each pair by representing the individuals in both space and time. ABMs often include explicit movement of animals and can therefore model heterogeneity in the population contact structure in great detail (Tang and Bennet, 2010). If the individuals are modeled, but the model does not explicitly consider interactions between them, the model is an IBM.

In a herd-level MSM, the whole herd or fractions of the population (rather than individuals) can be associated with specific disease states. Thus, there is no need to keep track of individuals within the population. For example, using the herd as the unit of interest, once a herd is infected then animals within that herd can be divided into compartments reflecting specific disease states (for example, *SIR*) and disease transmission within the herd is modeled using EBMs. Although within-herd disease spread elements are not mechanistically modeled, disease spread between herds is modeled using an IBM or even an ABM approach, by simulating the actual animal movements from one herd to another. Such models are called nested models in ecological modeling (Mideo et al., 2008).

In Appendix 1 (and <https://github.com/ckirkeby/MDT>), code examples for some of the models described above are shown; a *difference equation model,* a *differential equation model,* a *mechanistic stochastic model (at herd level),* and a *mechanistic stochastic model (at individual animal level).* The examples are written as functional programming, but other styles are available, such as object-oriented programming. For an introductory distinction between these two styles, see Chambers (2014).

## 2.2 Building a mechanistic model

Figure 1. Stages, and steps within each stage, in building a disease transmission model.



In this section, we focus on building an MSM. Model building can be divided into three stages: pre-programming, programming and post-programming. Each stage includes different elements that should be considered (Figure 1).

### 2.2.1 Pre-programming stage

#### Purpose

When using a model, it is important to consider the research question to be investigated, which will drive the requirements of the model output for the end-user (Garner and Hamilton, 2011). The purpose could be to estimate the epidemiological consequences of the disease (number of infected herds and epidemic duration). In the case of exotic diseases, the outputs could be used for contingency planning to improve surveillance and control; for example, identifying sentinel herds, culling capacity or laboratory capacity (for example, Garner et al., 2016; Boklund et al., 2017). In this case, it is essential to generate capacity-related data that are included as parameter values into the model. Similarly, if the purpose is to compare different surveillance strategies, sensitivity and specificity of tests used to detect disease could be included (Nusser et al., 2008; Halasa and Boklund, 2014).

Secondly, the purpose could be to evaluate and identify optimal control strategies given a particular set of circumstances and constraints (for example Allore et al., 1998; Backer et al., 2012; Østergaard et al., 2005;Ward et al., 2009; Lu et al., 2010; Dürr et al., 2014; Gussmann et al., 2018). This requires additional knowledge and data to inform the model. If the intention is to investigate control strategies for a specific disease, then all variables and mechanisms necessary to simulate those control strategies should be included. For example, to simulate vaccination as a control strategy, parameters related to vaccination-specific parameters (such as the number of individuals or herds vaccinated per day, vaccine efficacy, time required to order vaccine and perform vaccination) should be included, dependent on the required level of detail (Backer et al., 2012; Dürr, 2014; Dürr and Ward, 2015). In EBMs, this is typically modeled by including a specific compartment in the model that accounts for vaccinated animals (Tildesley et al., 2006; Backer et al., 2012; Rodrigues et al., 2014). The optimal strategies could be defined according to epidemiological and economic outputs (if economic outputs are included, the model can be called a bioeconomic simulation model; for example, Kirkeby et al., 2016).

Thirdly, if there are gaps in knowledge about the disease system, modeling can be used to simulate various transmission patterns and compare model output with data from real situations, to deduce the most likely transmission pathways and understand disease dynamics, or to derive values for parameters such as transmission rates (for example Zinsstag et al, 2009; Marcé et al., 2010; Schulz et al., 2018).

The structure and outputs of the model must reflect the purpose of the model. For example, if the purpose is to estimate the prevalence of a disease over time, the output should include the number of infected animals and the total population at-risk of that disease during the relevant time periods of the simulated outbreak. For other research questions, the time to detect the disease or the economic benefit generated for a specific scenario might be of interest

#### Unit of interest

The largest unit of interest is selected such that the simulation model is sufficiently representative of the system so that outputs are meaningful and useful according to the purpose of the model. This is the epidemiological unit of the model, and can range from individuals (e.g. Allore et al., 1998; Groenendaal et al., 2002) or their parts (e.g. Gussmann et al., 2018) to sub- or entire populations (Kopec et al., 2007). In some systems there might be more than one unit of interest to be modeled, as in the case of vector-borne diseases − both the vector and the animal can be units of interest (Græsbøll et al., 2012). It then becomes critical to consider how these units interact so that disease transmission is modeled. In Appendix 1 we provide code examples of MSM using different units of interest (also available online at <https://github.com/ckirkeby/MDT>).

The choice of epidemiologic unit of interest is highly dependent on the purpose of building the model, the disease of concern and the data available to parameterize the model. Whereas for some disease transmission models disease spread at the individual level needs to be captured (e.g. because disease detection or control is performed on individual level), in the case of modeling the spread of an exotic disease in animals aggregated in herds in a country or a region, the herd might be a more realistic unit to model, because surveillance and decisions occur at the herd-level, and also it would be computationally challenging to model each individual animal in the country. This also raises another consideration: the availability of computational resources. This is discussed in Section *Programming stage*.

#### System knowledge, complexity and data availability

Model outputs reflect the input parameters of the model; if the input parameters are highly uncertain, the results will also be highly uncertain. There will likely be more variance in real life than in a model, and if the model is parameterized using data from just one subpopulation of the target population, bias in model output could be introduced. Furthermore, all processes are modeled with the assumption that they correctly represent the system. This is not always the case. Following the principle of parsimony, a model should only be as complex as necessary to achieve the model purpose, thereby requiring the minimum number of assumptions (Lilien, 1975). In all models, decisions need to be made about which of the known processes to include and which to exclude.

To create a model that is a sufficient representation of a real-life system, good knowledge of that system is needed. This includes the population dynamics of the unit of interest in the system, including the population structure in terms of births, deaths and lifespan (this is usually based on age, but in the case of a livestock production system, this could be parity), migration of individual units in and out of the system, the contact patterns of units and the production system of the modelled population (for example, milk or beef production). It also includes knowledge of the epidemiology of the disease to be modeled, such as the relevant disease states and their durations, the modes of transmission of the causative pathogen (for example, whether or not airborne spread is an essential pathway of transmission) and how the disease develops in the individuals. This information is important to gather prior to model building to assess how much data about the system is available, the level of uncertainty that is due to limited knowledge, and the feasibility of delivering outputs that fulfil the purpose of the model. If essential data are missing to fulfil the designated purpose, options include collecting more data before modeling is initiated, re-specifying model complexity or re-evaluating the model purpose.

#### Model type selection

Once the unit of interest has been established, and knowledge about the system to be modeled and the available data obtained, an informed decision can be made about the type of model required (for example, whether an EBM or an MSM is appropriate). In this way, model specification (structure and type) and data gathering for parameterization is typically an iterative process (Figure 1, section 2.4).

### 2.2.2 Programming stage

***Programming language and coding***

Programming languages can be classified in many ways − such as whether interpreted directly or compiled (all the code has to be run together, rather than running one single line of code at a time; for example, Python and R *versus* C++ and Fortran, respectively); and whether they are ‘high’ or ‘low’ level languages. This latter classification refers to the machine-readability of the language; many languages used in the context of disease modeling can be considered high-level (for example, Pascal, Java, C++, R and Python). In general, high-level languages require more memory space but are more readable by a human, and therefore more accessible to people without detailed programming training. Low level languages are more efficiently machine readable and traditionally considered to run faster. However, language classifications are now blurred and many languages have features consistent across classifications (consequently, once-compiled then interpreted languages can run quickly). In addition, the increasing computational power associated with high-performance computing enables horizontal scaling of models by multi-threading a model and running it across multiple cores to reduce ‘walltime’ (the time taken to complete a simulation).

Focusing on final run speed also ignores the concept of overall programming productivity. Programming in some languages is more challenging and less accessible to the research team, which increases the time required for programming. An increasing number of teams in the veterinary field use the free software R (R Core Team, 2019), which is a statistical programming language suitable for building both EBMs (for example, Frasso et al., 2016) and MSMs (for example, Bates et al., 2003; Kirkeby et al., 2016; Dürr et al., 2014;). There are many packages available for languages such as R, and they are well-supported and maintained by R’s open-source community. These packages remove the need to program basic functions, which allows the team to focus on modeling the system and the disease, thus increasing modeling productivity. Overall, the level of modeling productivity also depends on the context of model development; for example longer lasting explorative research projects (such as a PhD study) versus fast acting near real-time decision making (such as an exotic disease incursion that demands immediately decisions).

In regards to code programming, we highly recommend that modelers annotate their code during modeling with detailed descriptions of each part of the code. This assists the modeler to remember what each line of code does, and also facilitates use of the model by others. Following publication of a study, it is a requirement of many journals that the code be made available to readers. Version control is also valuable, especially when more than one modeler is involved in the project or when published code is used by other researchers. This can be facilitated by version control tools such as git (<https://git-scm.com>, accessed 10/09/2019) so that modelers can easily track changes in the code, and view previous versions (branches) of the model. We also highly recommend that each line or chunk of co­de is executed with fictitious inputs to check for errors (debugging) during modeling and for model verification (see Section *Model Verification and Validation* for more details).

#### Modeling the population structure

During this step, the ‘background’ structure of the system is modeled. For example, a model of canine rabies spread requires a population of dogs. In an individual- or agent-based MSM, each dog might be assigned characteristics such as age and sex, and the geographic location of its home (Dürr and Ward, 2015). The population structure often includes the life cycle of the individuals, so that over time, some individuals die and some are born. This is linked to the disease model; for example, the newborns can be susceptible, infected or immune (see section *Modeling Disease Progression*). In an example of Johne’s disease (paratuberculosis) transmission, individual cattle or herds could be modeled, and characteristics might include individuals’ milk production and lactation duration (e.g. Groenendaal et al., 2002; Lu et al., 2010; Kirkeby et al., 2016).

Commonly in veterinary science models, it is crucial to include the spatial component to allow spatio-temporal modeling of disease transmission. This can be realized by using geolocations of the units of interest, e.g. farms, as a feature of the population structure (e.g. Bates et al., 2003; Tildesly and Keeling, 2008; Backer et al., 2012). Spatio-temporal modeling could also represent other population structures than farms, as in the case of modeling spatio-temporal distributions of vectors that transmit blue tongue virus (Kelso and Milne, 2014).

Once the background structure of the system in which the disease exists has been modeled, this should be verified and tested (see sections *Model verification and validation*) before infection transmission is modeled, to ensure that the model simulates the system with sufficient accuracy, as well as to determine computing requirements such as the number of iterations required for burn-in (see section *Modeling disease dynamics*).

#### Modeling disease dynamics

A disease transmission model must simulate disease dynamics. For each stage of disease, there should be a corresponding ­state in the model if it is important in terms of disease dynamics and model output. The simplest framework is the *SI* model with two, mutually exclusive disease states; *Susceptible (S)* and *Infectious (I)*. All individuals in the model are assigned to either *S* or *I*. For each simulated time step, each individual has a probability of acquiring infection − depending on the contact pattern between individuals and the disease transmission rate given a contact − and thus transitioning from *S* to *I*. We describe below how the transmission events can occur. In the case of the *SI* model, there is no probability of individuals returning to the *S* state. If animals can recover from the modeled disease, the model becomes an *SIS* model in which infectious individuals can return to the *S* state. This represents recovered animals that are re-susceptible and the transmission from *I* to *S* is quantified by the recovery rate which can be influenced by self-recovery or by treatment, and is essentially modeled the same way as infection events using the transmission rate. Another approach to modeling recovery could be that after a fixed number (or drawn from a distribution) of time steps an individual transitions from the *I* to the *S* state. This would be useful if an infection usually has a well-defined time span. It is also useful if there is a minimum infectious period, in which case a recovery rate is not suitable because some individuals could, by chance, recover in the model immediately following infection.

Another common framework is the SIR model, in which the infectious individuals can ‘recover’ and enter the *Recovered* (*R*) state (and be resistant to infection) or be ‘removed’ from the population. For example, in the case of a rabies model, infected animals always die and therefore are removed. The transmission from *I* to the *R* state are also modelled (such as for the *SIS* model) either via recovery rate or fixed duration infectious period. Following this logic, the disease transmission framework can be expanded in numerous ways dependent on the disease; for example, by introducing an *Exposed (E)* state before infection that reflects an individual who is latently infected but not yet infectious. It is noteworthy that even if some disease states occur in reality, it is not always useful or necessary to represent them in the model; it will also depend on the purpose of the model. Other parameters − such as mortality or behavior traits associated with disease states − can be made dependent on disease states for individuals or groups. Thus, infected individuals can behave differently – for example, in terms of contact rates – than non-infected individuals, as investigated in Brookes et al. (2019).

In the case of modeling endemic diseases, once the population structure and disease propagation framework is modeled, the model might need to be simulated to reach a stable prevalence (‘burn-in’ period; the number of time steps for the population characteristics and the disease prevalence to stabilise). When such a model is used to assess control strategies, these strategies are usually implemented after the burn-in period, when a stable situation has been reached.

#### Modeling disease transmission

Disease transmission can be modeled in a large number of ways, and here, we describe some commonly used methods. Generally, transmission can be either direct or indirect (Anderson and May, 1991). In the case of direct transmission (used to model contagious transmission), a contact rate (usually represented by β and also called the effective contact rate) can be used to represent disease transmission (Kean et al., 1999) as the rate at which two individuals come into effective contact (sufficient enough for disease transmission to occur) per unit time. This effective contact rate consists of two parts; a physical contact rate between individuals (*C*), and the probability of transmission per contact per time-step (*PT*):

|  |  |
| --- | --- |
|  | eq. 1 |

Also it is necessary to differentiate between density-dependent or frequency-dependent transmitted diseases (Begon et al., 2002; Ryder et al., 2007). Density-dependent transmission assumes that the more individuals there are, the greater the probability of infection, for a fixed effective contact rate (McCallum et al., 2001):

|  |  |
| --- | --- |
|  | eq. 2 |

where *P(S)* is the probability of infection for each susceptible individual, is the effective contact rate and *S* and *I* are the number of susceptible and infected individuals, respectively.

Frequency-dependence means that the probability of infection is independent of the number of individuals in the population (*N*):

|  |  |
| --- | --- |
|  | eq. 3 |

The effective contact rate can be fixed or variable (drawn from a distribution). In mechanistic models, each individual has a probability of an effective contact. This probability can be equal for all individuals in the model or it can be defined for each individual separately. One option is to convert the contact rate to a probability by using an exponential function (which ensures that the probability never exceeds 1):

|  |  |
| --- | --- |
|  | eq. 4 |

where *P(S)* is the probability of infection for a susceptible individual *S*, *β* is the effective contact rate (scaled in case of discrete-time models to reflect the same time interval as used in the model), *I* is the number of infectious individuals in the model at this time step, *N* is the total number of individuals in the model. A simple R code example of this type of model is given in Appendix 1. In this way, the infection pressure is scaled to the proportion of the population that are infected within each time step, i.e. *I* changes over time, whereas and *N* (within a closed system) remain constant. The infection process is dynamic because the *P(S)* changes over time with changing numbers of *I* in the population (assuming a fixed *N* and ).

Another option is to consider the spatial structure of the underlying demography and define the probability of effective contact per time step for a susceptible unit of interest dependent on its distance from infectious units in the model. For this approach, distance kernels are built from which the probability of effective contact can be drawn (such as used in Dürr & Ward, 2015; Hudson et al., 2019). This spatially dependent contact rate can be combined with information on the frequency of contacts between herds. For example, the frequency of potential contacts between herds may not only depend on the distance between them, but also on the frequency of movements from and to herds, which in turn may depend on the herd types (Bates et al 2003;s Ferguson et al., 2001).

When appropriate knowledge and data are available, the contact structure of a population can be based on a social network (Reynolds et al., 2015; Hirsch et al., 2016; Van der Waal et al., 2017; Brookes et al., 2019). A heterogeneous herd contact structure between groups of animals (for example, calves and heifers) and homogenous contacts within animal groups might also be described (Østergaard et al., 2005; Ezanno et al., 2007; Kirkeby et al, 2016; Gussmann et al., 2018).

There are also several ways to simulate indirect (environmental) disease transmission. It can be similarly spatially dependent as described for the direct transmission, or simulated as a fixed transmission probability:

|  |  |
| --- | --- |
|  | eq. 5 |

Here, *P(S)* is the probability of infection of a susceptible individual *S*, and is the indirect disease transmission rate. This fixed transmission rate can be based on a stable baseline infection pressure, or more variable, such as bacteria from infected individuals shed over time in the environment (Kirkeby et al., 2016).

When disease transmission occurs through both direct and indirect contacts, a combination of both of these pathways (direct and indirect) might be more appropriate (Gussmann et al., 2018).

### 2.2.3 Post-programming stage

#### Model verification and validation

Model verification and validation is essential to ensure that model concepts, programming and outputs are reliable, accurate and reasonably representative for the modeled system for the intended purpose of the model (Klügl, 2008’ Garner and Hamilton, 2011; Sargent, 2013). Model verification ensures that model code and the conceptual framework are implemented correctly. Verification is also called computerized model verification, internal validation or conceptual validation (Sargent, 2013). Several methods can be used for model verification, and we list here a few of the most important ones: 1) The rationalism method, in which several scenarios are simulated with different inputs, and outputs are compared to determine whether the changes are rational given the changes in the inputs (sensitivity analysis, see below); 2) The tracing method, in which individuals or other units of interest are followed through the different time steps; their progress within one or more process steps is observed and then the outputs of interest are verified (checking that they behave as expected); and 3) The face validity method, in which an expert is asked to check the outputs or even the code to verify the credibility of the model.

Model validation (also called external or operational validation) ensures that the model predictions have a satisfactory range of accuracy in relation to the actual behavior of the modeled system in real life (adapted from Sargent, 2013). Real-life data (i.e. empirical outbreak data) is needed to execute this process correctly; to our knowledge, few models in veterinary science have been externally validated (e.g. Foddai et al., 2014; Zinsstag et al., 2017; Zingg et al., 2015). This is usually due to the high associated costs or ethics of obtaining such data, and the complexity of the modeled systems. If empirical outbreak data are lacking from the setting in which the model was built and applied − such as in the case of exotic diseases and regions with historical disease freedom − then validation options might include either adapting the model to a region where data are available, or using previous outbreak data; for example, historical data from the last Swiss FMD outbreak was used to validate an FMD model for Switzerland (Zingg et al., 2015).

#### Convergence analysis

Convergence analysis assesses the stability of the results based on the number of iterations (repetitions) the model is simulated. It must be conducted before final model simulations for predictions, to ensure that the outputs are stable. Above a given threshold of simulations, the outputs should be independent of the number of model iterations. Stability can be checked by ensuring that the variance of the outputs of interest (for example, the number of infected individuals or epidemic duration) is stable. A commonly used approach is to visualize the change in the variance when increasing the number of iterations (e.g. Halasa et al., 2016), or to use thresholds of the coefficient of variance as a decision metric (Cowled et al., 2012, Dürr and Ward, 2015; Brookes et al., 2019).

We have included an example of how to determine convergence of a model in Appendix 1 (also available online at <https://github.com/ckirkeby/MDT>).

#### Sensitivity analysis

Sensitivity analysis is an essential process that is used to understand and examine the robustness of model predictions to changes in input parameter values, model structure and processes (e.g. Ezanno et al., 2007). In the sensitivity analysis, the behavior of the model and the outputs of interest are examined when the model or its parameters are varied. There are different ways to approach sensitivity analysis. Sensitivity analysis of input parameters can be assessed by changing the values of one (one-at-time sensitivity analysis) or more (global sensitivity analysis) parameters to examine the impact of these changes on model outputs. It can also be done by modeling a specific process in a different way to examine the impact of this process on model predictions (this is sometimes referred to as structural sensitivity analysis). Sensitivity analysis can be used to identify parameters and processes that have a major influence on model predictions; therefore, the values of these parameters − and the way in which processes are modeled − must be certain enough to produce model predictions acceptable to the end-user.

The simplest method is one-at-a-time perturbations (Norton, 2015). However, this does not allow assessment of the sensitivity of the model output to changes in combinations of parameter values (other than the default combination used). In most cases it is impractical to vary all parameters in all combinations and examine their influence on the model outputs. Another used method is the latin hypercube scheme in which a number of combinations of values for the parameter to be examined are chosen at random (Lord et al., 1996). This should represent the desired range of all parameter values. Thus, the latin hypercube is not an exhaustive analysis exploring all combination of parameter values, but rather can be a practical shortcut to gain insight into the effect of changes in parameter combinations on the model outputs. Many more and complex methods exist (Frey and Patil, 2002; Brookes et al., 2015; Norton, 2015), and a complete review is beyond the scope of this article.

The results of the sensitivity analysis can be presented in different ways. One option is to use a tornado chart (Abdulhameed et al., 2018). Another option is to show the full progression of disease over all iterations (Gussmann et al., 2018 – supplementary information).

We have included code in Appendix 1 to conduct a simple sensitivity analysis on a model parameter (also available online at <https://github.com/ckirkeby/MDT>).

#### Presentation of model outputs

Presentation of clear results that deliver project requirements is an important element for transparent communication of the model outputs. This should already be reflected and incorporated during the design stage. The type of output obtained from the model is dependent on the modeling approach used. For example, deterministic models provide single value outputs (without variation), whereas stochastic models provide distributions of outputs. Thus, when results from stochastic models are presented, it is essential to not only show median or mean values, but also the variation around these values; for example, using boxplots or histograms. From a disease transmission model, outputs usually include the number of infected units of interest and the epidemic duration. However, outputs can also include the number of units of interest under control (culled, vaccinated or banned in movements), economic outputs in case of a bioeconomic model, predicted changes in production (such as milk yield or growth rates), or maps from spatially-explicit models.

### 2.2.4 Documentation and Communication

Good documentation is essential to allow the reproducibility and repeatability of the model, communication of model outcome statements, and comparison between different models. Standardized protocols for disease transmission model documentation have been developed, such as the ODD (Overview, Design concepts, and Details) (Grimm et al., 2010) and TRACE (Grimm et al., 2014) and can be used to communicate models in scientific publications.

At all stages of model design, development and implementation, communication should be maintained with relevant stakeholders. These will include the end-users of the model, but can also include experts for the specific disease and system modeled, and those that are funding model development and implementation. Comprehensive communication at all stages ensures that the model focus remains to defined purpose, so that useful information is provided to the end-users, or that the end-user can even adapt the model according to specific needs.

**2.2.5 Recent developments**

Here we present some recent developments in disease transmission models used in veterinary medicine. First, more than one disease can be modeled within one model. Mostert et al. (2018) present a bio-economic stochastic dynamic model that simulates subclinical and clinical ketosis, mastitis, metritis, displaced abomasum and lameness in dairy cattle. In intense production systems, such as in the dairy sector, it is an advantage to evaluate the impact of several diseases concurrently, to optimize management strategies. Inclusion of economic impacts and the economics of disease mitigation in these models facilitates broader use, in addition to improving animal welfare. Second, many populations can be captured in one model. One example is the trend for models of vector-borne diseases (which we have not covered here, and introduces at least one more population, the vector, into the model).

Third, the ensemble modeling approach is relatively new in veterinary epidemiology (Webb et al., 2017). Decisions on how to respond to an incursion of FMD virus in a previously disease-free country are complex. Several models of FMD spread have been developed and applied. These vary in their approaches to modeling, assumptions made, the disease processes modeled, and parameterization. For any set of inputs, outputs from these various models are plausible. Variability in model outputs can be valuable because these are likely to include the range of realisations that could be observed during an FMD outbreak. A method of reconciling variability − borrowed from fields such as weather forecasting, climate-change science and medical science − has recently been applied to this situation. Using outputs from six different models which simulated the spread of FMD in the Midlands and Wales areas of the United Kingdom in 2001, Webb et al. (2017) applied a Bayesian Reliability Ensemble Average (BREA) method to integrate outputs regarding outbreak duration and two control methods. The BREA method determines the weights applied to each model output based on agreement with observed data (bias criterion) and consensus between models (convergence criterion). The latter was used by Webb et al. (2017) and their case study highlights the potential of ensemble modelling to reduce the uncertainty of outputs from individual models, thus improving decision-making.

# Conclusions and recommendations

Disease transmission models are motivated to gain a better understanding of a system’s dynamics, predicting a system’s outputs, or to study how the outputs of a system can be influenced. They have application when disease transmission, surveillance and control experiments are not possible, would not sufficiently represent real-life disease ecology, or are unethical. In veterinary science, models of disease transmission have been applied to identify critical elements and knowledge gaps, and to support operational and policy decisions on disease control, prevention and surveillance. Throughout the description of model development in the current paper, it is clear that there are many decisions to be made during the process of designing and implementing mechanistic models. In conclusion, we emphasize two well-known, key axioms: 1. disease transmission models are simplified representations of real-life systems so that ‘all models are wrong, but some are useful’ (Box, 1976), and 2. model outputs can only be as accurate as model inputs allow. Model simplification is often driven by data availability; therefore, full use of any available data is recommended. However, when considering whether more data should be collected or how a process should be modeled, we note that highly detailed models (more complex processes with more parameters) can produce output that might be less generalizable than more simplified models. In addition, the output from more simplified models might adequately predict the essential components of disease transmission needed to achieve the end-users’ objectives. This presents modelers with dilemmas that – in the authors’ opinion – can be a potential pitfall for less experienced modelers; a highly detailed model is not necessarily less ‘wrong’ or more ‘useful’ than a simplified model. Whilst the steps of model verification, validation and sensitivity analysis can help avoid too much or too little simplification during the model building process, we recommend that particularly during the design phase, modelers focus on development of the simplest model to achieve useful output. Communication between end-users and modelers about the value and assumptions of a model is critical. We therefore recommend that modelers and end-users, wherever possible, establish a framework for communication about modeling objectives, the need for validation, and application of model outputs to ensure optimal use of simulation modeling, to improve animal health, welfare and production.

ACKNOWLEDGEMENTS

This article was initiated during a special course in simulation modeling held at the University of Sydney in June 2019, funded by the University of Sydney and University of Copenhagen Partnership Collaboration Awards 2019.

**REFERENCES**

Allore, H.G., Erb, H.N., 1998. Partial budget of the discounted annual benefits of mastitis control strategies. J. Dairy Sci. 81, 2280–2292.

Abdulhameed, M. F., Habib, I., Al-Azizz, S. A., and Robertson, I. (2018). Cystic echinococcosis in marketed offal of sheep in Basrah, Iraq: Abattoir-based survey and a probabilistic model estimation of the direct economic losses due to hydatid cyst. Parasite epidemiology and control, 3(1), 43-51.

Alkhamis, M., Hijmans, R. J., Al-Enezi, A., Martínez-López, B., & Perea, A. M. (2016). The use of spatial and spatiotemporal modeling for surveillance of H5N1 highly pathogenic avian influenza in poultry in the middle east. Avian diseases, 60(1s), 146-155.

Anderson, R. M., and May, R. M. (1991). Infectious diseases of humans. New York: Oxford Science Publications.

Backer JA, Hagenaars TJ, Nodelijk G, van Roermund HJW (2012) Vaccination against foot-and-mouth disease I: Epidemiological consequences. Prev Vet Med107: 27–40.

Bates, T. W., Thurmond, M. C., & Carpenter, T. E. (2003). Results of epidemic simulation modeling to evaluate strategies to control an outbreak of foot-and-mouth disease. American Journal of Veterinary Research, 64(2), 205-210.

Begon, M., M. Bennett, R. G. Bowers, N. P. French, S. M. Hazel and J. Turner (2002). "A Clarification of Transmission Terms in Host-Microparasite Models: Numbers, Densities and Areas." Epidemiology and Infection 129(1): 147-153.

Boklund, A., Halasa, T. H. B., Christiansen, L. E., Willeberg, P., & Enøe, C. (2012). Simulated Effects of Introducing Emergency Vaccination or Depopulation During FMD Outbreaks in Denmark. In EuFMD meeting 2012: Open Session.

Box, G. E. (1976). Science and statistics. Journal of the American Statistical Association, 71(356), 791-799.

Brookes, V. J., Jordan, D., Davis, S., Ward, M. P. and Heller, J. (2015) Saltelli Global Sensitivity Analysis and Simulation Modelling to Identify Intervention Strategies to Reduce the Prevalence of Escherichia coli O157 Contaminated Beef Carcasses. PLoS ONE, 10(12), e0146016. doi:10.1371/journal.pone.0146016

Brookes, V.J., Dürr, S., Ward, M.P. (2019) Rabies-induced behavioural changes are key to rabies persistence in dog populations: Investigation using a network-based model. PLoS Negl Trop Dis 13(9): e0007739. https://doi.org/10.1371/journal. pntd.0007739

Chambers, J. M. (2014). Object-oriented programming, functional programming and R. Statistical Science, 29(2), 167-180.

Cowled, B. D., Garner, M. G., Negus, K., and Ward, M. P. (2012). Controlling disease outbreaks in wildlife using limited culling: modelling classical swine fever incursions in wild pigs in Australia. Veterinary research, 43(1), 3.

sDiekmann, O., Heesterbeek, J. A. P. & Metz, J. A. J. 1990 On the definition and computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. 28, 365 –382. (doi:10.1007/ BF00178324)

Diekmann, O., Heesterbeek, J. A. P., & Roberts, M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. Journal of the Royal Society Interface, 7(47), 873-885.

Dürr, S., zu Dohna, H., Di Labio, E., Carpenter, T.E., Doherr, M.G., (2013). Evaluation of control and surveillance strategies for classical swine fever using a simulation model. Prev Vet Med 108(1):73-84.

Dürr, S., Fasel-Clemenz, C., Thür, B., Schwermer, H., Perler, L., Doherr, M.G., Hadorn, D., (2014). Evaluation of the benefit of emergency vaccination in a foot-and-mouth disease free country with low livestock density. Prev Vet Med 113(1):34-46

Dürr, S., and Ward, M. P. (2015). Development of a novel rabies simulation model for application in a non-endemic environment. PLoS neglected tropical diseases, 9(6), e0003876.

Engblom, S., Eriksson, R., & Widgren, S. (2019). Bayesian epidemiological modeling over high-resolution network data. arXiv preprint arXiv:1910.11720.

Ezanno, P., Fourichon, C., Viet, A.-F., Seegers, H., 2007. Sensitivity analysisto identify key-parameters in modelling the spread of bovine viraldiarrhoea virus in a dairy herd. Prev. Vet. Med. 80, 49–64.

Ferguson, N. M., Donnelly, C. A., and Anderson, R. M. (2001). The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. Science, 292(5519), 1155-1160.

Foddai, A., Enøe, C., Krogh, K., Stockmarr, A., and Halasa, T. (2014). Stochastic simulation modeling to determine time to detect bovine viral diarrhea antibodies in bulk tank milk. Preventive veterinary medicine, 117(1), 149-159.

Frasso, G., and Lambert, P. (2016). Bayesian inference in an extended SEIR model with nonparametric disease transmission rate: an application to the Ebola epidemic in Sierra Leone. Biostatistics, 17(4), 779-792.

Frey, C. H., and Patil, S. R. (2002). Identification and review of sensitivity analysis methods. Risk analysis, 22(3), 553-578.

Garner, M. G. and Hamilton, S. A. (2011). Principles of epidemiological modelling. Revue Scientifique et Technique-OIE, 30(2), 407.

Garner MG, Bombarderi N, Cozens M, Conway ML, Wright T, Paskin R, et al. (2016) Estimating resource requirements to staff a response to a medium to large outbreak of foot and mouth disease in Australia. Transbound Emerg Dis 63:e109–21. doi:10.1111/tbed.12239

Green, M. J., Medley, G. F., Bradley, A. J., & Browne, W. J. (2010). Management interventions in dairy herds: exploring within herd uncertainty using an integrated Bayesian model. Veterinary research, 41(2), 1-10.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: a review and first update. Ecological modelling, 221(23), 2760-2768.

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S., ... & Thorbek, P. (2014). Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. Ecological modelling, 280, 129-139.

Groenendaal, H., Nielen, M., Jalvingh, A.W., Horst, S.H., Galligan, D.T., Hesselink, J.W., 2002. A simulation of Johne’s disease control. Prev. Vet. Med. 54, 225–245.

Græsbøll, K., Bødker, R., Enøe, C., & Christiansen, L. E. (2012). Simulating spread of bluetongue virus by flying vectors between hosts on pasture. Scientific reports, 2, 863.

Gussmann, M., Kirkeby, C., Græsbøll, K., Farre, M. and Halasa, T. (2018). A strain-, cow-, and herd-specific bio-economic simulation model of intramammary infections in dairy cattle herds. Journal of theoretical biology, 449, 83-93.

Halasa, T. and Boklund, A. (2014). The impact of resources for clinical surveillance on the control of a hypothetical foot-and-mouth disease epidemic in Denmark. PLoS One, 9(7), e102480.Halasa, T., Nielen, M., Huirne, R. B. M. and Hogeveen, H. (2009). Stochastic bio-economic model of bovine intramammary infection. Livestock Science, 124(1-3), 295-305.

Halasa, T., Bøtner, A., Mortensen, S., Christensen, H., Toft, N. and Boklund, A. (2016). Simulating the epidemiological and economic effects of an African swine fever epidemic in industrialized swine populations. Veterinary microbiology, 193, 7-16.

Halasa, T., Boklund, A., Bøtner, A., Mortensen, S., & Kjær, L. J. (2019). Simulation of transmission and persistence of African swine fever in wild boar in Denmark. Preventive veterinary medicine, 167, 68-79.

Hedell, R., Andersson, M. G., Faverjon, C., Marcillaud-Pitel, C., Leblond, A., & Mostad, P. (2019). Surveillance of animal diseases through implementation of a Bayesian spatio-temporal model: A simulation example with neurological syndromes in horses and West Nile Virus. Preventive veterinary medicine, 162, 95-106.

Hirsch, B. T., Reynolds, J. J., Gehrt, S. D., & Craft, M. E. (2016). Which mechanisms drive seasonal rabies outbreaks in raccoons? A test using dynamic social network models. Journal of Applied Ecology, 53(3), 804-813.

Hudson, E. G., Brookes, V. J., Ward, M. P., & Dürr, S. (2019). Using roaming behaviours of dogs to estimate contact rates: the predicted effect on rabies spread. Epidemiology and Infection, 147, e135. doi:10.1017/S0950268819000189

Hunter, E., Mac Namee, B. and Kelleher, J. D. 2018. A Comparison of Agent-Based Models and Equation Based Models for Infectious Disease Epidemiology. 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science, 2018 . doi.org/ 10.21427/rtq2-hs52

Kean, J. M., Barlow, N. D., & Hickling, G. J. (1999). Evaluating potential sources of bovine tuberculosis infection in a New Zealand cattle herd. New Zealand Journal of Agricultural Research, 42(1), 101-106.

Keeling, M. J., Woolhouse, M. E., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., ... & Grenfell, B. T. (2001). Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science, 294(5543), 813-817.

Keeling, M. J. 2008. Modeling infectious diseases in humans and animals. Princeton: Princeton University Press.

Kelso, J. K., Milne, G. J. (2014) A spatial simulation model for the dispersal of the bluetongue vector Culicoides brevitarsis in Australia. Plos One 9(8): e104646 doi:[10.1371/journal.pone.0104646](https://dx.doi.org/10.1371%2Fjournal.pone.0104646" \t "pmc_ext)

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 115(772), 700-721.

Kirkeby, C., Græsbøll, K., Nielsen, S. S., Christiansen, L. E., Toft, N., Rattenborg, E. and Halasa, T. (2016). Simulating the epidemiological and economic impact of paratuberculosis control actions in dairy cattle. Frontiers in veterinary science, 3, 90.

Klügl, F. (2008). A validation methodology for agent-based simulations. In Proceedings of the 2008 ACM symposium on Applied computing (pp. 39-43). ACM.

Kopec, J. A., Finès, P., Manuel, D. G., Buckeridge, D. L., Flanagan, W. M., Oderkirk, J., Abrahamowicz, M, Harper, S. Sharif, B., Okhmatovskaia A., Sayre, E. C., Rahman, M. M., Kudahl, A. B., Østergaard, S., Sørensen, J. T., & Nielsen, S. S. (2007). A stochastic model simulating paratuberculosis in a dairy herd. Preventive veterinary medicine, 78(2), 97-117.

Kudahl, A. B., Østergaard, S., Sørensen, J. T., & Nielsen, S. S. (2007). A stochastic model simulating paratuberculosis in a dairy herd. Preventive veterinary medicine, 78(2), 97-117.

Laager, M., C. Mbilo, E. A. Madaye, A. Naminou, M. Léchenne, A. Tschopp, S. K. Naïssengar, T. Smieszek, J. Zinsstag and N. Chitnis (2018). "The importance of dog population contact network structures in rabies transmission." PLOS Neglected Tropical Diseases 12(8): e0006680.

Lilien, G. L. (1975). Model relativism: A situational approach to model building. Interfaces, 5(3), 11-18.

Lord, C. C., Woolhouse, M. E. J., Rawlings, P. and Mellor, P. S. (1996) Simulation studies of African horse sickness and Culicoides imicola (Diptera: Ceratopogonidae). Journal of medical entomology, 33(3), 328-338.

Lu, Z., Schukken, Y. H., Smith, R. L. & Grohn, Y. T. Stochastic simulations of a multi group compartmental model for Johne’s disease on US dairy herds with test-based culling intervention. *Journal of theoretical biology* **264**(4), 1190–1201 (2010).

Mahsin, M., Deardon, R., & Brown, P. (2019). Geographically-dependent individual-level models for infectious diseases transmission. arXiv preprint arXiv:1908.06822.

Mancy, R., Brock, P. M., and Kao, R. R. (2017). An integrated framework for process-driven model construction in disease ecology and animal health. Frontiers in veterinary science, 4, 155.

Marcé, C., Ezanno, P., Weber, M. F., Seegers, H., Pfeiffer, D. U. and Fourichon, C. 2010. Invited review: modeling within-herd transmission of Mycobacterium avium subspecies paratuberculosis in dairy cattle: a review. Journal of dairy science, 93(10), 4455-4470.

Mardones, F.; Perez, A. ; Sanchez, J. ; Alkhamis, M. ; Carpenter, T. (2010). Parameterization of the duration of infection stages of serotype O foot-and-mouth disease virus: an analytical review and meta-analysis with application to simulation models. Veterinary Research; 41(4): 45

Martinez-Lopez B, Perez AM, Sanchez-Vizcaıno JM. (2010). A simulation model for the potential spread of foot-and-mouth disease in the Castile and Leon region of Spain. Prev Vet Med 96:19–29. doi:10.1016/j. prevetmed.2010.05.015 12.

McCallum, H., Barlow, N., & Hone, J. (2001). How should pathogen transmission be modelled?. Trends in ecology & evolution, 16(6), 295-300.

Mideo, N., Alizon, S., and Day, T. (2008). Linking within-and between-host dynamics in the evolutionary epidemiology of infectious diseases. Trends in ecology & evolution, 23(9), 511-517.

Mostert, P. F., Bokkers, E. A. M., Van Middelaar, C. E., Hogeveen, H., and De Boer, I. J. M. (2018). Estimating the economic impact of subclinical ketosis in dairy cattle using a dynamic stochastic simulation model. animal, 12(1), 145-154.

Mur, L., Sánchez‐Vizcaíno, J. M., Fernández‐Carrión, E., Jurado, C., Rolesu, S., Feliziani, F., ... & Martínez‐López, B. (2018). Understanding African swine fever infection dynamics in Sardinia using a spatially explicit transmission model in domestic pig farms. Transboundary and emerging diseases, 65(1), 123-134.

Norton, J. (2015). An introduction to sensitivity assessment of simulation models. Environmental Modelling & Software, 69, 166-174.

Nusser, S. M., Clark, W. R., Otis, D. L. and Huang, L. (2008). Sampling considerations for disease surveillance in wildlife populations. The Journal of Wildlife Management, 72(1), 52-60.

Protter, P.E., (2005). Stochastic integration and differential equations. Springer, Berlin, Germany <https://doi.org/10.1007/978-3-662-10061-5>

R Core Team (2019). R: A language and environment for statistical computing, version 3.3. 1. Vienna, Austria: R Foundation for Statistical Computing; 2016.

Rahmandad, H., Sterman, J., (2008). Heterogenious and network structure in the dynamics of diffusion: comparing agent-based and differential equation models. Management Science, 54:5. <https://doi.org/10.1287/mnsc.1070.0787>

Reynolds, J. J., Hirsch, B. T., Gehrt, S. D., & Craft, M. E. (2015). Raccoon contact networks predict seasonal susceptibility to rabies outbreaks and limitations of vaccination. Journal of Animal Ecology, 84(6), 1720-1731.

Rodrigues, H. S., Monteiro, M. T. T. and Torres, D. F. (2014). Vaccination models and optimal control strategies to dengue. Mathematical biosciences, 247, 1-12.

Ryder, J. J., Miller, M. R., White, A., Knell, R. J., and Boots, M. (2007). Host‐parasite population dynamics under combined frequency‐and density‐dependent transmission. Oikos, 116(12), 2017-2026.

Sargent, R. G. (2013) Verification and validation of simulation models, Journal of Simulation, 7:1, 12-24, DOI: 10.1057/jos.2012.20

Schulz, J., Boklund, A., Toft, N. and Halasa, T. (2018). Drivers for Livestock-Associated Methicillin-Resistant Staphylococcus Aureus Spread Among Danish Pig Herds - A Simulation Study. Scientific Reports 1–11, doi:10.1038/s41598-018-34951-1*.*

Singer, A., Salman, M., and Thulke, H. H. (2011). Reviewing model application to support animal health decision making. Preventive Veterinary Medicine, 99(1), 60-67.

Smieszek, T., (2009). A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease transmission. Theoretical Biology and Medical Modelling 6:25 doi:10.1186/1742-4682-6-25

Smith, R. L., Schukken, Y. H., and Gröhn, Y. T. (2015). A new compartmental model of Mycobacterium avium subsp. paratuberculosis infection dynamics in cattle. Preventive veterinary medicine, 122(3), 298-305.

Szmaragd, C., Wilson, A. J., Carpenter, S., Wood, J. L. N., Mellor, P. S., Gubbins, S. (2009) A modeling framework to describe the transmission of blue tongue virus within and between farms in Great Britin. Plos One 4(11):27741 doi:[10.1371/journal.pone.0007741](https://dx.doi.org/10.1371%2Fjournal.pone.0007741" \t "pmc_ext)

Tang, W., and Bennett, D. A. (2010). Agent‐based modeling of animal movement: a review. Geography Compass, 4(7), 682-700.

Tildesley, M. J., Savill, N. J., Shaw, D. J., Deardon, R., Brooks, S. P., Woolhouse, M. E., ... & Keeling, M. J. (2006). Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. Nature, 440(7080), 83.

Tildesley MJ, Keeling MJ (2008) Foot-and-Mouth Disease - A Modelling Comparison between the UK and Denmark. Prev Vet Med 85: 107–124.

van der Waal, J., Daenekindt, S., & de Koster, W. (2017). Statistical challenges in modelling the health consequences of social mobility: the need for diagonal reference models. International journal of public health, 62(9), 1029-1037.

Ward, M. P., Highfield L. D., Vongseng, P., and M. G. Garner. (2009). Simulation of foot-and-mouth disease spread within an integrated livestock system in Texas, USA. Prev Vet Med (2009) 88:286–97. doi:10.1016/j.prevetmed.2008.12.006 11.

Webb, C.T., Carpenter, T.E., Dürr, S., Ferrari, M., Garner, M.G., Jewell, C., Lindström, T., Patyk, K., Stevenson, M., Tildesley, M., Ward, M.P. and Werkman, M. (2017). Ensemble modelling and structured decision-making to support emergency disease management. Preventive Veterinary Medicine, 138: 124−133.

Zingg D, Häsler S, Schuepbach- Regula G, Schwermer H and Dürr S (2015) Evidence for Emergency Vaccination Having Played a Crucial Role to Control the 1965/66 Foot-and-Mouth Disease Outbreak in Switzerland. Front. Vet. Sci. 2:72. doi: 10.3389/fvets.2015.00072

Zingg, D., Steinbach, S., Kuhlgatz, C., Rediger, M., Schüpbach-Regula, G., Aepli, M., Grøneng, G. M., Dürr, S. (2017).Epidemiological and Economic Evaluation of Alternative On-Farm Management Scenarios for Ovine Footrot in Switzerland, Front Vet Sci, 4: 1-13

Zinsstag, J., Dürr, S., Penny, M. A., Mindekem, R., Roth, F., Gonzalez, S. M., Naissengar, S. and Hattendorf, J. (2009). Transmission dynamics and economics of rabies control in dogs and humans in an African city. Proceedings of the National Academy of Sciences, 106(35), 14996-15001.

Zinsstag,J., Lechenne, M., Laager, M., Mindekem, R., Naïssengar, S., Oussiguéré, A., Bidjeh, K., Rives, G., Tessier, J., Madjaninan, S., Ouagal, M., Moto, D.D., Alfaroukh, I.O., Muthiani, Y., Traoré, A., Hattendorf, J., Lepelletier, A., Kergoat, L., Bourhy, H., Dacheux., L., Stadler, T., Chitnis, N. (2017) Vaccination of dogs in an African city interrupts rabies transmission and reduces human exposure. Science Translational Medicine, 9(421), eaaf6984

Zinsstag, J., Dürr, S., Penny, M. A., Mindekem, R., Roth, F., Gonzalez, S. M., Naissengar, S. and Hattendorf, J. (2009). Transmission dynamics and economics of rabies control in dogs and humans in an African city. Proceedings of the National Academy of Sciences, 106(35), 14996-15001.

Østergaard, S., Chagunda,M.G.G., Friggens, N.C., Bennedsgaard, T.W., Klaas, I.C., 2005. A stochastic model simulating pathogen-specific mastitis control in a dairy herd. J. Dairy Sci. 88, 4243–4257.

APPENDIX 1

This code is also available online at <https://github.com/ckirkeby/MDT>

# Code for showing simple examples of the different types of models created in R.

#### Difference equation model example ####

beta <- 0.5 # transmission rate

timesteps <- 1:10 # Time steps

S <- 9

I <- 1

N <- S + I

n.inf <- numeric()

n.sus <- numeric()

n.pop <- numeric()

for (i in timesteps)

{

dI <- beta \* S \* I / N

I <- I + dI

S <- S - dI

N <- S + I

n.inf[i] <- I

n.sus[i] <- S

n.pop[i] <- I + S

}

plot(timesteps, n.pop, type="b", ylab="Y", xlab="time", main="Difference equation model example", ylim=c(0,10))

points(n.inf, col="red", type="b")

points(n.sus, col="blue", type="b")

#### Differential equation model example ####

## Load deSolve package

library(deSolve)

## Create an SIR function

SI <- function(time, state, parameters) {

with(as.list(c(state, parameters)), {

dS <- -beta \* S \* I

dI <- beta \* S \* I

return(list(c(dS, dI)))

})

}

### Set parameters

# Proportion infected at start:

inf.start <- 0.1

## Proportion in each compartment: 10% infected and 90% susceptible

init <- c(S = 1-inf.start, I = inf.start)

## beta: infection parameter; gamma: recovery parameter

parameters <- c(beta = 0.5)

## Time frame

timesteps <- c(0:10)

## Solve using ode (General Solver for Ordinary Differential Equations)

out <- ode(y = init, times = timesteps, func = SI, parms = parameters)

## change to data frame

out <- as.data.frame(out)

out$N <- out$S + out$I

plot(timesteps, out$N, type="b", xlab="time",

main="Differential equation model example", ylim=c(0,1),

ylab="Proportion of population")

points(out$I, col="red", type="b")

points(out$S, col="blue", type="b")

#### Stochastic population-based mechanistic model example ####

beta <- 0.5 # transmission rate

timesteps <- 1:10 # Time steps

S <- 9

I <- 1

N <- S + I

n.inf <- numeric()

n.sus <- numeric()

n.pop <- numeric()

for(i in timesteps)

{

PI <- 1 - exp( - beta \* I / N)

dI <- sum( rbinom(S , 1 , prob=PI) )

I <- I + dI

S <- S - dI

N <- S + I

n.inf[i] <- I

n.sus[i] <- S

n.pop[i] <- I + S

}

plot(timesteps, n.pop, type="b", ylab="Y", xlab="time",

main="Stochastic population-based mechanistic model example", ylim=c(0,10))

points(n.inf, col="red", type="b")

points(n.sus, col="blue", type="b")

#### Stochastic individual-based mechanistic model example ####

beta <- 0.5 # transmission rate

timesteps <- 1:10 # Time steps

pop <- data.frame(ID = c(1:10), inf.status = c(1 , rep(0, 9)) )

n.inf <- numeric()

n.sus <- numeric()

n.pop <- numeric()

for(i in timesteps)

{

I <- length(pop$inf.status[pop$inf.status == 1])

S <- length(pop$inf.status[pop$inf.status == 0])

N <- length(pop$inf.status)

PI <- 1 - exp( - beta \* I / N)

new.inf <- rbinom(S , 1 , prob=PI)

pop$inf.status[pop$inf.status == 0] <- new.inf

n.inf[i] <- I

n.sus[i] <- S

n.pop[i] <- I + S

}

plot(timesteps, n.pop, type="b", ylab="Y", xlab="time",

main="Stochastic individual-based mechanistic model example", ylim=c(0,10))

points(n.inf, col="red", type="b")

points(n.sus, col="blue", type="b")

#### Stochastic individual-based mechanistic simulation model with multiple iterations example ####

MaxIterations <- 10

Output <- matrix(numeric(0),ncol=4)

timesteps <- 1:10 # Time steps

for(j in 1:MaxIterations)

{

beta <- runif(1,0.4,0.6) # transmission rate

pop <- data.frame(ID = c(1:10), inf.status = c(1 , rep(0, 9)) )

n.inf <- numeric()

n.sus <- numeric()

n.pop <- numeric()

for(i in timesteps)

{

I <- length(pop$inf.status[pop$inf.status == 1])

S <- length(pop$inf.status[pop$inf.status == 0])

N <- length(pop$inf.status)

PI <- 1 - exp( - beta \* I / N)

new.inf <- rbinom(S , 1 , prob=PI)

pop$inf.status[pop$inf.status == 0] <- new.inf

n.inf[i] <- I

n.sus[i] <- S

n.pop[i] <- I + S

Output <- rbind(Output,c(i,sum(pop$inf.status==0),sum(pop$inf.status==1)))

}

}

## If you run the model 1 iteration, you can observe the progress of the infection for that iteration using this code

plot(timesteps, n.pop, type="b", ylab="Y", xlab="time",

main="Stochastic individual-based mechanistic model example", ylim=c(0,10))

points(n.inf, col="red", type="b")

points(n.sus, col="blue", type="b")

## If you run the model for > 1 iteration, then you can observe the effect of randomness on infection using the following code

plot(Output[,1],Output[,2],xlab="Time", ylab="Number of animals",ylim=c(0,10),typ="l")

lines(Output[,1],Output[,3],col="red")

## To observe the progress of infection based on all iterations, median number can be ploted over time as follows

Sus <- sapply(unique(Output[,1]),function(x) median(Output[Output[,1]==x,2]))

Infect <- sapply(unique(Output[,1]),function(x) median(Output[Output[,1]==x,3]))

plot(unique(Output[,1]),Sus,xlab="Time", ylab="Number of animals",ylim=c(0,10),typ="l")

lines(unique(Output[,1]),Infect,col="red")

#### Sensitivity analysis example ####

# Make the model (except the beta definition) as a function of beta:

# Here we use the individual-based mechanistic model, but it could be any of those described above.

model <- function(beta)

{

pop <- data.frame(ID = c(1:10), inf.status = c(1 , rep(0, 9)) )

n.inf <- numeric()

n.sus <- numeric()

n.pop <- numeric()

for(i in timesteps)

{

I <- length(pop$inf.status[pop$inf.status == 1])

S <- length(pop$inf.status[pop$inf.status == 0])

N <- length(pop$inf.status)

PI <- 1 - exp( - beta \* I / N)

new.inf <- rbinom(S , 1 , prob=PI)

pop$inf.status[pop$inf.status == 0] <- new.inf

n.inf[i] <- I

n.sus[i] <- S

n.pop[i] <- I + S

}

return(data.frame(n.pop=n.pop, n.inf=n.inf, n.sus=n.sus))

}

beta <- 0.5 # transmission rate

timesteps <- 1:10 # Time steps

tmp <- model(beta)

plot(timesteps, tmp$n.pop, type="b", ylab="Y", xlab="time",

main="Stochastic individual-based mechanistic model example", ylim=c(0,10))

points(tmp$n.inf, col="red", type="b")

points(tmp$n.sus, col="blue", type="b")

# Now the model can be run several times with a new result because it is stochastic.

# We can then use the model to find the distribution of the number of infected individuals:

# We loop over a number of iterations:

iterations <- 1:1000

infected.end <- numeric()

for(j in iterations)

{

tmp <- model(beta)

infected.end[j] <- tmp$n.inf[10]

}

hist(infected.end, main="Number of infected individuals at the end of the simulation")

# In this case, most of the population is most frequently infected.

table(infected.end)

# We can then decrease the beta and look again:

beta <- 0.35 # transmission rate

# We loop over a number of iterations:

iterations <- 1:1000

infected.end <- numeric()

for(j in iterations)

{

tmp <- model(beta)

infected.end[j] <- tmp$n.inf[10]

}

hist(infected.end, main="Number of infected individuals at the end of the simulation")

# Now the most frequent number of infected animals are around 6. Notice that there

# are a large proportion of iterations where the only infected individual is the

# one that was infected from the beginning: the epidemic never took off.

# We can see this clearly in the table:

table(infected.end)

#### Convergence ####

beta <- 0.5 # transmission rate

timesteps <- 1:10 # Time steps

# We use the model defined above, in the loop.

iterations <- 1:1000

infected.end <- numeric()

for(j in iterations)

{

tmp <- model(beta)

infected.end[j] <- tmp$n.inf[10]

}

# Now find the variance for 1 to 100 iterations:

conv <- numeric()

for(u in 2:length(iterations))

{

conv[u] <- var(infected.end[1:u])

}

plot(conv)

# Generally, it looks like 400 iterations are sufficient for convergence with the default values used here

# Copyright (c) Carsten Kirkeby 2019