Mathematical Biosciences 243 (2013) 28–45

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

When can a deterministic model of a population system reveal what will happen on average?

Leif Gustafsson *, Mikael Sternad

Signals and Systems, Department of Engineering and Sciences, Uppsala University, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 3 June 2011
Received in revised form 4 January 2013
Accepted 11 January 2013
Available online 31 January 2013

Keywords:
Comparison of models
Consistent results
Poisson Simulation
Population model
Simulation methodology
Unbiased results

ABSTRACT

A dynamic population system is often modelled by a deterministic difference equation model to obtain average estimates. However, there is a risk of the results being distorted because unexplained (random) variations are left out and because entities in the population are described by continuous quantities of an infinitely divisible population so that irregularly occurring events are described by smooth flows.

These distortions have many aspects that cannot be understood by only regarding a deterministic approach. However, the reasons why a deterministic model may behave differently and produce biased results become visible when the deterministic model is compared with a stochastic model of the same structure.

This paper focuses first on *demographic stochasticity*, i.e. stochasticity that refers to random variations in the occurrence of events affecting the state of an individual, and investigates the consequences of omitting this by deterministic modelling. These investigations reveal that bias may be strongly influenced by the type of *question to be answered* and by the *stopping criterion* ending the analysis or simulation run. Two cases are identified where deterministic models produce unbiased state variables: (1) Dynamic systems with stable local linear dynamics produce unbiased state variables asymptotically, in the limit of large flows; and (2) linear dynamic systems produce unbiased state variables as long as all state variables remain non-negative in both the deterministic and the stochastic models. Both cases also require the question under study to be compatible with a solution over a fixed time interval.

Stochastic variability of initial values *between* simulation runs because of uncertainty or lack of information about the initial situation is denoted *initial value stochasticity*. Elimination of initial value stochasticity causes bias unless the model is linear. It may also considerably enlarge bias from other sources.

Unknown or unexplained variations from the environment (i.e. from outside the borders of the studied system) enter the model in the form of stochastic parameters. The omission of this *environmental stochasticity* almost always creates biased state variables.

Finally, even when a deterministic model produces unbiased state variables, the results will be biased if the output functions are not linear functions of the state variables.

1. Introduction

This paper examines the mathematical conditions for a deterministic difference equation model to produce unbiased results when used for population studies.

A *population* consists of entities that may have attributes of different kinds. These entities may perform different actions and they may interact with other entities and with the environment. A *population model* should therefore capture the nature of the population system under study as it evolves *dynamically* over *continuous time*, focusing on the aspects specified by the purpose of the study. Population models constitute a class of models with a *non-negative integer number of discrete entities* such as plants, animals, patients, vehicles, molecules, atoms, data packets, etc. Such models are fundamental in ecology, epidemiology, demography and queuing systems. They are also used in physics, chemistry, molecular biology, traffic planning, production planning and many other fields.

1.1. Stochastic population models – representation and time handling

In this paper the behaviour of stochastic population models is used as a reference in comparisons with that of deterministic models.

All irregularities from the behaviour of a real population, its environment and the system conditions at a given (initial) point in time are rarely known and cannot be modelled in detail. Instead, irregularities are often specified in probabilistic terms, leading to

---

* Corresponding author. Tel.: +46 18 500061.
E-mail address: Leif.Gunnar.Gustafsson@gmail.com (L. Gustafsson).

0025-5564/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mbs.2013.01.006

stochastic population models. A non-trivial question is then how such a probabilistic description of a dynamic population can, and should, be converted into an executable simulation model.

Three types of stochasticity may be implemented in a population model, namely demographic stochasticity, environmental stochasticity and initial value stochasticity. *Demographic stochasticity* [1,2] is the randomness about 'when the next event such as birth, death, migration or other change of state (or attribute) of an entity will occur'. Stochastic variability of initial values *between* simulation runs because of uncertainty or lack of information about the initial situation is denoted *initial value stochasticity*. This is conceptually related to demographic stocasticity, but will only intervene at the initial point of the study, while *environmental stochasticity* [1,2] comes from external, partly unknown factors such as temperature, precipitation, supply of food, etc. that can vary *within* the simulation run.

In a recent investigation of the foundations of stochastic population modelling and simulation [3], it was found that demographic stochasticity can be correctly handled by either of three different types of *representation*:

- First, the population model can use an *entity-based representation* where each entity is individually described with its attributes and conditional behaviour and can be individually monitored. This micro representation is used in e.g. Discrete Event Simulation and Individually Based Simulation.
- The second option is to use a *compartment-based representation* where all entities in the same situation (i.e. having the same set of attribute values) are located in the same compartment. Only the number of entities in each compartment is then recorded. This macro representation is used in e.g. compartment models, which can be numerically handled using tools from Continuous System Simulation.
- Third, a *state-based representation* exists, where every situation that the entire model can be in has to be represented as a unique separate state. This representation is used for e.g. Markov models and often leads to a model of gigantic size.

The *time handling*, i.e. the updating of the (demographic) stochastic model over time, is based on irregularly occurring *events* that will change the attribute values. The time handling can be performed in three different ways, irrespective of the representation used. The first way is to increment time in such small time-steps that zero or at most one event may occur during a time-step (using Bernoulli statistics). The second way is to use larger time-steps allowing several events per time-step (using Poisson statistics), and the third is to 'jump' to the point in time for the next event (using exponential statistics).

It has been shown [3] that irrespective of the representation (entity, compartment or state) and the time handling method selected, stochastic population models will produce mutually *consistent* (contradiction-free) results, provided that some simple rules are observed during the model building process.

1.2. Deterministic models of a population

Historically, there is a long tradition of using *deterministic* differential or difference equation models for analysis or simulation of populations.

Using a deterministic model simplifies model building. There is an overwhelming number of studies where populations are modelled in this way, using continuous state variables and ignoring the effects of stochasticity. Even in textbooks on e.g. epidemiology, biology and ecology [4–6] and in more general textbooks about dynamic systems [7,8], the presentations are often entirely based on deterministic models, or at least lack demographic stocasticity. It is then taken for granted that this is adequate and no concern is expressed about the validity.

On the other hand, there are early examples of understanding that stochastic and deterministic models of the same population system may produce profoundly different results. For example, in 1926 A.G. McKendrick [9] published a stochastic treatment of an epidemic process and in 1960 M.S. Bartlett [10] published his book '*Stochastic Population Models in Ecology and Epidemiology*' in which the necessity of stochastic population models was emphasised and different ways to construct them were discussed. Many, more recent, textbooks present both deterministic and stochastic modelling and stress the relative advantages of both types [11–15]. However, a general discussion of the circumstances where deterministic models are appropriate is lacking: "Full conditions for agreements between deterministic and expected stochastic solutions are at present unknown." [13]. Some studies have compared the results of different approaches to specific problems [16–22], but in general this is not done.

1.3. Objective and overview of the paper

"Understanding the dynamics of stochastic populations, and how they deviate from the deterministic ideal, is being viewed with increasing importance by ecologists and epidemiologists" [22].

It is not unusual that modellers assume that models with large numbers (or large flow rates) or that linear models will be sufficient to secure unbiased average results when stochastics are not included. Still worse, working within the deterministic tradition, this issue may not be raised at all.

Since entity-based, compartment-based and state-based models are three ways to produce a consistent realisation of a conceptual population model with demographic stochasticity [3,23], they provide a gold standard when compared with a simplified deterministic difference equation model.

The objective of this paper is to investigate when a deterministic model can reproduce the average quantities[1] of a corresponding stochastic population model and to demonstrate and explain what might otherwise happen. For this purpose, we require the quantities to be *unbiased*, in the sense that:

$$\textit{deterministic quantity} = E[\textit{stochastic quantity}],$$

where $E[\cdot]$ denotes an ensemble average for a corresponding stochastic model. In the following, this property is evaluated analytically when possible. Otherwise, a check is made of whether a quantity from the deterministic model falls within the 95% confidence interval of that from the stochastic model, based on 10,000 simulation runs.

A fundamental restriction implied by the use of a deterministic model is that all information about variations, confidence intervals, risks, extremes, correlations, etc. is lost and only deterministic results such as point estimates can be obtained instead of probability distribution functions.

Nevertheless, a deterministic model of a population may be useful in several circumstances if it can produce outcomes that are close to the average of the outcomes that would be produced by a corresponding stochastic model:

- One case is when average results are the main outcome of interest in a study. The deterministic model could then produce these average results in just one simulation run, instead of averaging over many simulation runs of a stochastic model.

---

[1] We need to distinguish between bias in the state variables and in the results (outcomes of the study). 'Quantity' is used to include both 'state variable' and 'result'.

- Furthermore, parameter estimation, optimisation and sensitivity analysis are difficult and problematic to perform with stochastic models. If it is known that a deterministic approximation will reproduce average results correctly, then model parameters of the deterministic model can be adjusted/optimised with respect to a data set. They can then also be used in a corresponding stochastic model, which reveals additional features such as the variability around the average outcomes.

Investigation of the quality of deterministic model behaviour is complicated in several ways. First, omission of different types of stochastics may create bias in the state variables representing the subpopulations. Even when the state variables are unbiased, a result, reflecting the objective of the study, can be a non-linear function of the state variables and thus creates biased results.

Furthermore, the objective of the study also affects the stopping criterion of the solution. The problem then is that the ensemble of stochastic solutions may end at different points in time, causing state variable values of a deterministic model to deviate from the averages of the state variables of the stochastic solutions.

To clarify the situation, this investigation is conducted in steps. First, we investigate when state variables of deterministic compartment models equal the average of the state variables of corresponding compartment models with demographic stochasticity. Next, we investigate what happens when initial value stochasticity is omitted, and thereafter the consequences of omitting environmental stochasticity. Finally, we investigate the conditions for unbiased state variables to generate unbiased results, expressed in terms of an output function.

This paper mainly focuses on qualitative aspects of using a deterministic model – when will this create bias and of what reasons? However, in the examples the presentation is complemented with a more pragmatic view – when can a deterministic model be used to reproduce average results with a tolerable bias?

The paper is structured as follows. Section 2 presents the method for investigating possible bias for deterministic models. In Section 3 we discuss various implications of removing the stochastics of a population model. In Section 4 the consequences of omitting demographic stochasticity are demonstrated by three simple examples. The two following sections deal with special cases where unbiased state variables might be obtained when demographic stochasticity is dropped. Section 5 considers the case where the flows between compartments stay very large and the model has asymptotically stable linearised dynamics, while Section 6 deals with positive linear systems and demonstrates the necessity of a limitation to a fixed stopping time criterion to obtain unbiased results. Section 7 treats the consequences of omitting initial value stochasticity, and Section 8 the consequences of omitting environmental stochasticity. Section 9 discusses unbiased results expressed by an output function of the state variables. In Section 10 we take a pragmatic approach, focusing on the specific model and discuss how to test the consequences of dropping the stochastics of the specific model. Finally, Section 11 provides a discussion and conclusions.

2. Methods

2.1. Introduction

There is a great advantage in dividing the modelling process into a *conceptualisation phase* and a *realisation phase*. The conceptualisation phase deals with whether the conceptual model, with all its simplifications, boundaries, assumptions, etc., is a good enough description of the studied system to serve the given purpose of the study. The realisation phase, on the other hand, only deals with constructing an executable dynamic model that should be consistent (contradiction-free) with the conceptual model. The conceptualisation phase is outside the scope of this paper. We start from an ideal case where a conceptual stochastic population model is completely specified. Fig. 1 presents a general overview of the studied problem structure.

The method used in this investigation is to start with a well-defined stochastic population model and compare its average behaviour with the behaviour of a deterministic model of identical structure. Note here that demographic stochasticity is intrinsic in entity-based, compartment-based and state-based models, while initial value stochasticity and environmental stochasticity have to be added by specification.

Investigation of when a deterministic model of a population will produce unbiased results is strongly simplified by the fact that any of the three stochastic representations using any of the three ways of time handling will produce mutually consistent results. The deterministic compartment model is here usually compared with a stochastic compartment model where time is handled by Poisson statistics, a so-called *Poisson Simulation model* [24]. This method is also called the 'tau-leap method' (the tau is the time-step Δt leaping over many events) [25]. The two models to be compared are then both represented by systems of difference equations, so the structural and mathematical similarities become obvious.

2.2. The models to be compared

A general stochastic population model, realised in Poisson Simulation form, is given by:

$$x_{s,i}(t+\Delta t) = x_{s,i}(t) + In_i(t) - Out_i(t) \text{ and } x_{s,i} \geqslant 0; \quad i = 1\ldots n \tag{1A}$$

$$x_{s,i}(t+\Delta t) = algebraic\_fcn_i(\boldsymbol{x}_s(t+\Delta t), t+\Delta t); \quad i = n+1\ldots r \tag{1B}$$

$$In_i(t+\Delta t) = Po[\Delta t \cdot G_i(\boldsymbol{x}_S(t+\Delta t), t+\Delta t)]; \quad i = 1\ldots n \tag{1C-1}$$

$$Out_i(t+\Delta t) = Po[\Delta t \cdot H_i(\boldsymbol{x}_S(t+\Delta t), t+\Delta t)]; \quad i = 1\ldots n \tag{1C-2}$$

$$y_{s,k}(t) = C_k(\boldsymbol{x}_S(t), t), k = 1\ldots m \tag{1D}$$

$$x_{s,i}(0) = xo_{s,i}; \ \ i = 1\ldots n. \text{ Initial values, possibly from statistical distributions.} \tag{1E}$$

$$p_{s,j}(t); \ \ j = 1\ldots q. \text{ Assignment of parameters, possibly fromstatistical distributions.} \tag{1F}$$

$$\textit{Stopping criterion} \tag{1G}$$

$$\textit{Integration rule, size of } \Delta t. \tag{1H}$$

Here, $\boldsymbol{x}_S = (x_{s,1}, x_{s,2}, \ldots, x_{s,r})^T = (\boldsymbol{x}_s^T, x_{s,n+1}, \ldots x_{s,r})^T$ denotes an extended state vector. It is composed of the vector of integer-valued state variables (compartments) $\boldsymbol{x}_s = (x_{s,1}, x_{s,2}, \ldots, x_{s,n})^T$, and possibly also of auxiliary real-valued variables $x_{s,n+1}, x_{s,n+2}, \ldots, x_{s,r}$, defined by the algebraic equations (1B).

The difference equations (1A) are the kernel of the model, generating the dynamics of the state variables. They are expressed here in Euler form. Inflows to and outflows from each compartment are explicitly separated, which is necessary in the stochastic case. Note that $x_{s,i} \geqslant 0$ is required for all the n state variables. This sometimes requires the inclusion of a conditional clause into an appropriate equation.

In (1B), the $r - n$ algebraic equations specify auxiliary variables by algebraic functions of $\boldsymbol{x}_S$. In modelling, it is often practical, but not necessary, to perform various kinds of algebraic (non-dynamic) calculations by separate equations. Here, library functions from e.g. a Continuous System Simulation Language such as mathematical functions, logical functions, conditional statements, table look-up functions and time-related functions can be included.

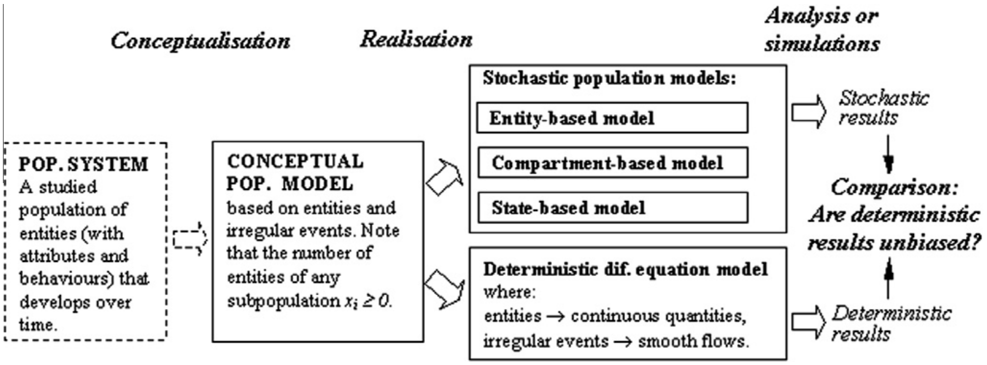


Fig. 1. Overview of the problem structure being studied. A conceptual model is defined from a population system at study. (This step is outside the scope of this paper.) The well-defined conceptual model is then translated into an executable stochastic model and an executable deterministic model with the same structures. Finally, the average results of the stochastic model are compared with those of the deterministic model.

In the flow equations[2] (1C), all flows into and out of each compartment are defined. It is here the demographic stochasticity is operating. For each time-step and each equation in (1C), the arguments of $Po[\Delta t \cdot G_i(\boldsymbol{x}_S(t+\Delta t), t+\Delta t)]$ or $Po[\Delta t \cdot H_i(\boldsymbol{x}_S(t+\Delta t), t+\Delta t)]$ are evaluated, and $Po[argument]$ is assigned a value by a Poisson-distributed random number generator [24]. Several inflows to a compartment i may be included in the function G_i and several outflows may be included in H_i. The state variables will remain integer-valued since they are initialised as (non-negative) integers and the $Po[.]$ terms in (1C) generate integer outcomes.[3]

The conceptual model should include the explicit formulation of the objective of the study. Eq. (1D) defines the output variable(s) that directly relate to the objective(s). An output is an algebraic function of one or several variables. The values of an output variable may be of interest during the whole trajectory or only at one particular instant, for example at stopping time. Note that the output variables $y_{s,k}$ may be functions also of auxiliary variables specified by the algebraic equations (1B).

The initialisation of the state variables is defined by (1E). In some studies, *stochastic initial values* may be used. This can e.g. be necessary when modelling an experiment where a number of individuals of a population are sampled to a study group. Initial values are then drawn at each simulation run from statistical distributions representing known conditions of the entire population.

Furthermore, parameters, e.g. temperature, precipitation or food supply, may be partly unknown or they may vary over time in not precisely known patterns. If the statistical effects of such irregularities on the model outcomes are of interest, then this can be incorporated into the model by letting the parameter assignments at (1F) be stochastic, with prescribed distributions and statistics of their time variability. This introduces *environmental randomness* into the simulation model by affecting the parameters of the functions G_i and H_i in (1C).

The stopping criterion (1G) needs to be defined, specifying when the solution/simulation (starting at time zero) is to be completed. The stopping criterion can be of two kinds:

1. A fixed stopping time criterion, or
2. Integration until some criterion, other than 'a fixed stopping time', is fulfilled. For this second type, stochasticity usually means that the executions will end at different points in time in different simulation runs.

Finally, in (1H), the integration rule and step-size Δt are specified. In this presentation, the Euler integration algorithm with a fixed step-size is used for simplicity. This means that the calculations are performed step-wise by holding the input intensities $G_i(\boldsymbol{x}_S(t),t)$ and the output intensities $H_i(\boldsymbol{x}_S(t),t)$ fixed during the time-step $(t,t+\Delta t)$. To avoid unnecessary complexity, throughout this paper it is assumed that the time-step is sufficiently small to cause negligible distortion of the time evolution of a model.

After the model equations are defined they are sorted, starting with all difference equations (1A), all algebraic equations (1B), and all flow equations (1C). At execution, the initial values (1E) and constant parameters (1F) are defined once and for all. Thereafter, in a loop over time, the Eqs. (1A), (1B) and (1C) (and (1F) in the case of time-varying parameters) are solved, followed by a redefinition of current time according to $t{:=}t+\Delta t$. This step-wise calculation over time continues until the stopping criterion is satisfied. For a stochastic model a large number of simulation runs are necessary to build the probability distribution/density functions (pdfs) of the state variables or output variables.

Now, we define a corresponding deterministic model in which all stochasticity is removed:

Definition 1. A '*corresponding deterministic model*' represents a model structure (1) where in (1C), $Po[\Delta t \cdot G_i(\boldsymbol{x}(t),t)]$ and $Po[\Delta t \cdot H_i(\boldsymbol{x}(t),t)]$ are replaced by the arguments $\Delta t \cdot G_i(\boldsymbol{x}(t),t)$ and $\Delta t \cdot H_i(\boldsymbol{x}(t),t)$ and where possible initial value stochasticity in (1E) and environmental stochasticity in (1F) are replaced by fixed deterministic values.

When we have only demographic stochasticity in (1), the corresponding deterministic model is said to be *embedded* in the stochastic model.

The deterministic model corresponding to (1) is given by:

$$x_{d,i}(t+\Delta t) = x_{d,i}(t) + In_i(t) - Out_i(t) \text{ and } x_{d,i} \geqslant 0; \quad i = 1 \ldots n \tag{2A}$$

$$x_{d,i}(t+\Delta t) = algebraic\_fcn_i(\boldsymbol{x}_d(t+\Delta t), t+\Delta t); \quad i = n+1 \ldots r \tag{2B}$$

$$In_i(t+\Delta t) = \Delta t \cdot G_i(\boldsymbol{x}_D(t+\Delta t), t+\Delta t); \quad i = 1 \ldots n \tag{2C-1}$$

$$Out_i(t+\Delta t) = \Delta t \cdot H_i(\boldsymbol{x}_D(t+\Delta t), t+\Delta t); \quad i = 1 \ldots n \tag{2C-2}$$

$$y_{d,k}(t) = C_k(\boldsymbol{x}_D(t), t); \quad k = 1 \ldots m \tag{2D}$$

$$x_{d,i}(0) = xo_{d,i}; \quad i = 1 \ldots n \tag{2E}$$

$$p_{d,j}(t); \quad j = 1 \ldots q \tag{2F}$$

$$\textit{Stopping criterion} \tag{2G}$$

$$\textit{Integration rule, size of } \Delta t. \tag{2H}$$

Here, $\boldsymbol{x}_D = (x_{d,1}, x_{d,2}, \ldots x_{d,r})^T = (\boldsymbol{x}_d^T, x_{d,n+1}, \ldots x_{d,r})^T$ denotes the extended state vector of the deterministic model. The vector of state

---

[2] We use the technical term 'flow equations' to denote Eq. (1C) although $\Delta t{*}flow\_rate$ is a somewhat unconventional construction since it depends on the user-defined quantity Δt.

[3] We could also include combined stochastic/deterministic models by letting some of the equations in (1C) be deterministic. However, this would here only complicate the description without adding further insight.

variables is real-valued here and is denoted by $\boldsymbol{x}_d = (x_{d,1}, x_{d,2}, \ldots, x_{d,n})^T$. If $x_{s,i}(0)$ and $p_{s,j}(t)$ in (1) are stochastic, then $x_{d,i}(0)$ and $p_{d,j}$ in (2) represent some kind of 'average' of $x_{s,i}(0)$ and $p_{s,j}(t)$. The execution takes place in the same way as for the stochastic model described above.[4] With no stochasticity, every simulation run will give the same result, so only one simulation run is needed.

Comparing (1) and (2), the fundamental concepts of indivisible entities and irregularly occurring instantaneous events in a stochastic population model are replaced by infinitely divisible amounts that change continuously over time in a deterministic model.

2.3. Simplifications

In the following we often use simplified forms of the models (1) and (2), in which the algebraic equations are eliminated. These forms are obtained by first including the algebraic equations (B) into the flow equations (C), and into the output equations (D). The flow equations (C) are then substituted into the difference equations (A). As a result of the substitutions, the functions $G_i(.)$ and $H_i(.)$ in the flow equations (C) and $C_k(.)$ in the output equations (D) are replaced by functions $g_i(.)$, $h_i(.)$ and $c_k(.)$ of the state variables only, since the auxiliary variables $x_{n+1}, x_2, \ldots, x_r$ have been eliminated.

In the stochastic case, model (1) then takes the simplified form:

$$\begin{aligned} &x_{s,i}(t+\Delta t) = x_{s,i}(t) + Po[\Delta t \cdot g_i(\boldsymbol{x}_s(t),t)] - Po[\Delta t \cdot h_i(\boldsymbol{x}_s(t),t)] \\ &\text{and } x_{s,i} \geqslant 0; \quad i = 1\ldots n, \\ &y_{s,k}(t) = c_k(\boldsymbol{x}_s(t),t); \quad k = 1\ldots m. \end{aligned} \tag{3}$$

Initial conditions, parameter values, the stopping criterion, the integration rule and the step-size are still part of the specification, but not explicitly repeated here.

The corresponding deterministic model (2) then takes the form

$$\begin{aligned} &x_{d,i}(t+\Delta t) = x_{d,i}(t) + \Delta t \cdot g_i(\boldsymbol{x}_d(t),t) - \Delta t \cdot h_i(\boldsymbol{x}_d(t),t) \\ &\text{and } x_{d,i} \geqslant 0; \quad i = 1\ldots n, \\ &y_{d,k}(t) = c_k(\boldsymbol{x}_d(t),t); \quad k = 1\ldots m. \end{aligned} \tag{4}$$

However, simplification by substituting flow equations into difference equations can't always be done without consequences. For example, a conservative system where a flow, F, transports entities from compartment x_1 to compartment x_2 with a flow rate proportional to x_1 is correctly described by:

$$\begin{aligned} &x_1(t+\Delta t) = x_1(t) - F \\ &x_2(t+\Delta t) = x_2(t) + F \\ &F = Po[\Delta t \cdot c \cdot x_1] \end{aligned}$$

Substituting the flow equation into the difference equations gives:

$$\begin{aligned} &x_1(t+\Delta t) = x_1(t) - Po[\Delta t \cdot c \cdot x_1] \\ &x_2(t+\Delta t) = x_2(t) + Po[\Delta t \cdot c \cdot x_1] \end{aligned}$$

This creates a problem, because now there are now two independent calls, with different return values, to the Po-generator (instead of one common call for the value of F) at each time-step. Then the total number of entities, $x_1 + x_2$, is no longer preserved.

However, for our purposes, inserting the flow equation into the difference equation causes no problem when analysing the average properties of the scalar state variable equations of (3) separately.

The main question for demographic stochasticity can now be rephrased as: When can dropping the $Po[\,]$ operators of (3) be justified in the sense that $y_{d,k}(t) = E[y_{s,k}(t)]$, $k = 1\ldots m$?

When comparing models (3) and (4), note that $E[Po[\Delta t \cdot g_i]] = \Delta t \cdot g_i$ and $E[Po[\Delta t \cdot h_i]] = \Delta t \cdot h_i$ (when the arguments are non-negative), so there is some hope that under certain conditions, the state variables of the stochastic model 'on average' could behave like those of the deterministic model. Additional conditions on the output equations $c_k(.)$ will also be required to ensure unbiased deterministic results.

3. Fundamental differences

This section presents a qualitative comparison of stochastic and deterministic models of a population in order to provide an understanding of why a deterministic model may produce biased results.

There are a number of (omitted) factors that might make the state variables and the results from a deterministic model of a population biased. Nine factors of particular importance are introduced below and discussed and exemplified in subsequent sections.

(a) *Non-linearity*: The models (3) and (4) may have flow rates $g_i(\boldsymbol{x},t)$ or $h_i(\boldsymbol{x},t)$; $i = 1\ldots n$, that are *non-linear functions* of the state variables. The state variables will then be more or less biased. For simplicity, consider the case where $n = 1$. Assume that a deterministic solution/trajectory for x at time t is $x(t)$ and that the stochastic solutions from two simulation runs pass through $x + \delta x$ and $x - \delta x$ at time t. At $t + \Delta t$ the *deterministic* model is updated by $\Delta t \cdot f(x,t)$; ($f = g$ or h), while the *stochastic* model is updated by $Po[\Delta t \cdot f(x+\delta x,t)]$ and $Po[\Delta t \cdot f(x-\delta x,t)]$. The average of $Po[\Delta t \cdot f(x+\delta x,t)]$ and $Po[\Delta t \cdot f(x-\delta x,t)]$ will then differ from that of $Po[\Delta t \cdot f(x,t)]$ if $f(\cdot)$ is a non-linear function of x.

(b) *Protection against a negative state variable*: A population or subpopulation can never have a negative number of entities. This property is preserved in the models in the state equations of (3) and (4) by the statements $x_i \geqslant 0$, $i = 1\ldots n$. This sometimes requires the inclusion of a conditional clause in a proper equation. For example, if $x_i(t+\Delta t) = x_i(t) - \Delta t \cdot a \cdot x_j(t)$, where $a \cdot x_j(t)$ is e.g. the number of trees felled, animals shot, enemies killed per time unit by x_j lumberjacks, hunters, soldiers, respectively, and where $i \neq j$, then a conditional statement is required to prevent x_i from becoming negative. One practical way is to limit the reduction of x_i to no more than x_i itself, for example with a construction such as: $x_i(t+\Delta t) = x_i(t) - MIN[\Delta t \cdot a \cdot x_j(t), x_i(t)]$. The same clause can be used for the stochastic case. Note, however, that when such a limiting clause becomes active, an otherwise linear model will proceed non-linearly, so the state variables of the deterministic model become biased. (This would be the case in Example 2, in the next section, if the simulation were to continue beyond the defeat of Force 2.)

(c) *Neglecting stochastic jumps*: For a deterministic model, the phase space can be divided into separate regions wherein a trajectory is restricted by the choice of initial values. For a stochastic model, '*stochastic jumps*' can cause transitions between such regions. For example, Poisson-distributed flows can take a population x_i to zero (*extinction*) in a finite time in a stochastic population model, even when this is impossible for the embedded deterministic model starting with $x_i > 0$. (Example 3, below.)

(d) *Neglecting oscillations generated by demographic stochasticity*: A deterministic model may reach a steady state while the corresponding stochastic model is repeatedly excited by

---

[4] In deterministic modelling, the total inflow rate G_i and the total outflow rate H_i per time-step to and from a compartment x_i are usually combined into a net term $(G_i - H_i)$. However, here they are kept separate to maintain the similarity to the stochastic case.

demographic variations. Even for a deterministic model starting from an equilibrium so that all x_i remain constant, the corresponding stochastic model, also starting from the same equilibrium, may generate oscillations. Absence of such oscillations may create bias for the deterministic model results. (Example 5 in Section 5.)

(e) *Neglecting that the stopping criterion should intervene at different points in time*: The question to be answered by analysis or model experiments is coupled to the stopping criterion. A fixed stopping time criterion (*If time* $\geqslant$ *T then Stop*) is a necessary condition for unbiased results. For other stopping criteria, a stochastic model will end at different points in time, which cannot be reflected by a corresponding deterministic model. (Example 2, in the next section.)

(f) *Averaging of stochastic initial values*: Stochastic initial values will create biased state variable values unless the model equations are linear.

(g) *Use of an average estimate as a fixed parameter value*: A treacherous but unnecessary error is to use an arithmetic average of a parameter that does not simply enter additively in an equation. (Example 8 in Section 8 illustrates this aspect.)

(h) *Neglecting environmental stochasticity*: With few exceptions, omitting environmental stochasticity will create bias in state variables.

(i) *Non-linear output function*: A question to be answered by analysis or model experiments is stated by an algebraic output function $y_k = c_k(x_1,\ldots,x_n,t)$. Even when the state variables $x_1,\ldots,x_n$ are unbiased, a non-linear output function will introduce bias into the results. This occurs for example when the quantity of interest is a ratio (x_1/x_2), a fraction $(x_1/(x_1 + x_2))$ or a non-linear cost function of $\boldsymbol{x}$. Also for a question such as: 'How often will x_i exceed a value slightly above the stationary level?', the deterministic and stochastic models may produce very different answers.

4. Illustrative examples of omitting demographic stochasticity

Replacing a demographically stochastic population model with a deterministic model can cause hugely biased results.

Example 1 (*A queuing system*). Consider an $M/M/1$ queuing system [26], where random arrivals to a single service station have an intensity of $\lambda = 1/10$ entities per minute. Let the station serve the entities with an exponentially distributed service time with an average capacity of $\mu = 1/9$ entities per minute. This model can be written in the form of Eq. (3) with the help of a conditional $MIN[a,b]$ clause that selects the smallest of the two arguments [27]:

$$x(t+\Delta t) = x(t) + Po[\Delta t \cdot \lambda] - MIN[Po[\Delta t \cdot \mu], x(t)],$$

where $x(t)$ denotes the actual number of queuing and served entities. The exponentially distributed time between arrivals is on average 10 min ($1/\lambda$), and the average of the exponentially distributed service time is 9 min ($1/\mu$). Theory or simulations will then produce the average number of queuing entities waiting for service as 8.1 persons, with an average waiting time of 1 h and 21 min.

Removing the stochastics, and instead using a constant arrival rate of $\lambda = 1/10$ per minute and a constant service rate of $\mu = 1/9$ per minute will give no waiting entities at all and a zero waiting time. Thus neglecting stochastics to describe irregular events can cause extremely misleading results.

More generally, queues and waiting for service will arise irregularly for all $\rho = \lambda/\mu < 1$, although less frequently for smaller values of ρ. For the corresponding deterministic model no queues or waiting are ever created for $\rho < 1$. □

The following example illustrates that linearity alone does not ensure unbiased results for any question at hand.

Example 2 (*Lanchester's model of warfare*). Lanchester's model of warfare [8,28], first published in 1916, is a simple model of a battle between two fighting forces. Force 1 is characterised by its number of entities (soldiers, aircraft, vessels, etc.), x_1, and hitting power, a, and Force 2 by its number, x_2, and hitting power, b. The hitting power is the (average) number of eliminated entities per time unit that each member can inflict on the enemy. The combat proceeds until the last entity of the losing force is eliminated.

In deterministic terms, the change in Force 1 depends on the number of entities in Force 2 and their hitting power, so it can be described by the differential equation: $dx_1/dt = -b \cdot x_2$, and the change in Force 2 has the similar equation $dx_2/dt = -a \cdot x_1$. This is a linear dynamic system, since it can be written as $d\boldsymbol{x}/dt = \boldsymbol{Ax}$, where:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{ and } \boldsymbol{A} = \begin{pmatrix} 0 & -b \\ -a & 0 \end{pmatrix}.$$

Here we are interested in the questions: Which side will win? What number of entities will remain after the battle? How long will the fight take? For simplicity, also assume that the hitting power of both forces is the same, i.e. $a = b = 0.1$.

Dividing the first equation by the second then gives $dx_1/dx_2 = x_2/x_1$ or $\int x_1 dx_1 = \int x_2 dx_2$. When integrated over the time of the fight 0 *to* τ this gives:

$$x_1^2(\tau) - x_1^2(0) = x_2^2(\tau) - x_2^2(0)$$

if the forces are initially $x_1(0)$ and $x_2(0)$. Thus when the weaker force (say Force 2) is eliminated at time τ (i.e. $x_2^2(\tau) = 0$), the relationship $x_1^2(\tau) = x_1^2(0) - x_2^2(0)$ holds. For example, with $x_1(0) = 5$ and $x_2(0) = 3$ we obtain $x_1^2(\tau) = 5^2 - 3^2 = 16$, so the model predicts precisely $x_1(\tau) = 4$ remaining entities after the combat. Simulation of the deterministic model gives $\tau \approx 6.94$ time units for $a = b = 0.1$.

Beautiful, but are these results correct? Does Force 1 always win? Will the average number of remaining entities in Force 1 be $x_1(\tau) = 4$? Will the battle on average take 6.94 time units? To answer these questions the assumptions of Lanchester's model are kept, but the number of entities eliminated for each time-step is now integer and stochastic. The stochastic model can be analytically solved (excluding time) using a probability tree. This is illustrated by Fig. 2 for the case $x_1(0) = 5$, $x_2(0) = 3$.[5]

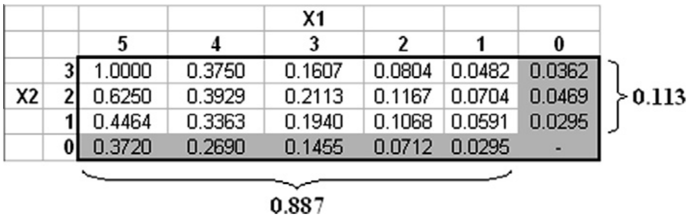
Summing up, we find that the probability of Force 1 winning is 0.887 and that of Force 2 winning is 0.113. Furthermore, the expected number of remaining entities after the combat for Force 1 is 3.545 ($\Sigma k \cdot p_{x1}(k) = 5 \cdot 0.3720 + 4 \cdot 0.2690 + \cdots + 1 \cdot 0.0295$), while for Force 2 it is 0.232.

For the stochastic model, the time of the fight will vary between simulation runs, with the average length estimated at 7.37 time units (7.27–7.47 for a 95% confidence level obtained from 10,000 simulation runs).

In conclusion: The elegant results from the deterministic model were all wrong. The stochastic results revealed the following major discrepancies with the deterministic model: (1) x_1 wins about 89% of the fights, not all of them. (2) The expected number of $x_1$entities after the fight is 3.545 entities, which deviates significantly from the deterministic model result of 4 entities. (3) The expected

---

[5] The probabilities are calculated by considering the sequence of successive elimination events. Conditional to being at state (5,3) the next event will take the model to state (5,2) or to state (4,3). The probability of reaching a border state such as (4,3) is $P(4,3) = P(4,3|5,3) = P(5,3) \cdot 3/(5+3) = 1 \cdot 3/8 = 0.375$, etc. Non-border states (i,j) can be reached from both $(i+1,j)$ and $(i,j+1)$, so such a state is calculated as the sum of two conditional probabilities, i.e. $P(i,j) = P(i,j|i+1,j) + P(i,j|\, i,j+1) = P(i+1,j) \cdot j/(i+1+j) + P(i,j+1) \cdot i/(i+j+1)$.



		X1					
		5	4	3	2	1	0
	3	1.0000	0.3750	0.1607	0.0804	0.0482	0.0362
X2	2	0.6250	0.3929	0.2113	0.1167	0.0704	0.0469
	1	0.4464	0.3363	0.1940	0.1068	0.0591	0.0295
	0	0.3720	0.2690	0.1455	0.0712	0.0295	-

Fig. 2. Probability tree presented as a spreadsheet model starting at $x_1 = 5$ and $x_2 = 3$. The cells show the probability of visiting state (x_1, x_2).

number of x_2 is 0.232 entities, not *zero*. (4) The length of the battle time is significantly longer (about 7.37 rather than 6.94 time units).

In Fig. 3a, stochastic trajectories from three simulation runs starting at $x_1(0) = 10$ and $x_2(0) = 8$ are shown in the $x_1 - x_2$ phase plane. As seen the trajectories may cross the demarcation line between the areas of the phase plane where x_1 and x_2 have the advantages. This makes it possible for x_2 to win. In Fig. 3b the deterministic trajectories can never cross each other.

When scaling up the initial number of entities on both sides in the same proportions, i.e. from (5,3) to (10,6) and to (20,12), the relative errors [(Deterministic_result – Stochastic_estimate)/stochastic_estimate] decrease from 14% to 8% and to 4% for $x_1(\tau)$, and decrease from 13% to 4% and to 1% for the probability that $x1$ will win the battle. However, the battle time, τ, remains about 7% too short even after scaling up the number of entities proportionally. □

A simple example of how elimination of stochasticity can block phenomena such as extinction and persistent excitations/oscillations is as follows.

Example 3 (*A prey-predator model with prey competition*). A *deterministic* Lotka–Volterra model [29], where competition between prey is added [8], is represented by:

$$dx_1/dt = a \cdot x_1 - b \cdot x_1x_2 - k \cdot x_1^2$$
$$dx_2/dt = c \cdot x_1x_2 - d \cdot x_2$$

where x_1 is the number of prey and x_2 the number of predators. The prey will breed as $f_1 = a \cdot x_1$, be reduced by predators as $f_2 = b \cdot x_1x_2$ and also by intra-species competition as $f_3 = k \cdot x_1^2$, where f_1 is an inflow to and f_2 and f_3 are outflows from x_1. The predators will breed as $f_4 = c \cdot x_1x_2$ because of food and will be reduced as $f_5 = d \cdot x_2$ because of natural deaths, where f_4 is an inflow to and f_5 is an outflow from x_2.

This system of equations has three stable equilibrium solutions, which are analytically obtained by setting the derivatives equal to zero and solving the resulting algebraic equations. These are $(x_1 = 0, x_2 = 0)$, $(x_1 = a/k,\ x_2 = 0)$ and $(x_1 = d/c, x_2 = a/b - kd/bc)$, corresponding to no prey or predators, only prey, and both prey and predators, respectively.

When simulating this model, the outcome will depend on the initial values of $x_1(0)$ and $x_2(0)$. If they are both set to zero they will remain so. If only $x_2(0)$ is set to zero a logistic solution is obtained, while if both $x_1(0)$ and $x_2(0)$ are non-zero, we obtain a solution where both species will coexist. In a phase plane the stability regions are strictly separated and a species can recover even from, say, 10^{-10} entities!

For a corresponding *stochastic model* with discrete entities [24,30], where $\Delta t \cdot f_i$ is replaced by $Po[\Delta t \cdot f_i]$; $i = 1 \ldots 5$, the system will never relax but will be persistently excited. Furthermore, the stochastic model can go from a region where $x_1 \neq 0$ and $x_2 \neq 0$ to a region where $x_1 \neq 0$ *and* $x_2 = 0$ or to a region where $x_1 = 0$ and $x_2 \neq 0$, leading to $x_1 = 0$ *and* $x_2 = 0$, but not in opposite directions, meaning that one or both species may become extinct. See Fig. 4.

For the deterministic model starting within $x_1 \neq 0$, $x_2 \neq 0$, the solution is trapped in the region where both x_1 *and* x_2 are strictly positive. The trajectories of the deterministic model will then converge towards $x_1 = d/c$, $x_2 = a/b - kd/bc$. The stochastic model, which has a non-zero probability of reaching $x_2 = 0$ or $x_1 = 0$ (leading to $x_2 = 0$), will do so sooner or later if the simulation is long enough, or will do so in a fraction of N simulations if N is large enough. The absorbing states $(x_1 \neq 0, x_2 = 0)$ and $(x_1 = 0, x_2 = 0)$ will then affect the expected values of the stochastic model results.

To make the discussion more general, the phase plane of x_1 and x_2 is rescaled to $x_1/(d/c)$ and $x_2/(a/b - kd/bc)$ in Fig. 5, so that the deterministic equilibrium point gets the coordinates (1, 1).

To examine what will happen when the stochastic and deterministic models are scaled up, we double and quadruple the deterministic equilibrium point from (60, 28) to (120, 56) and to

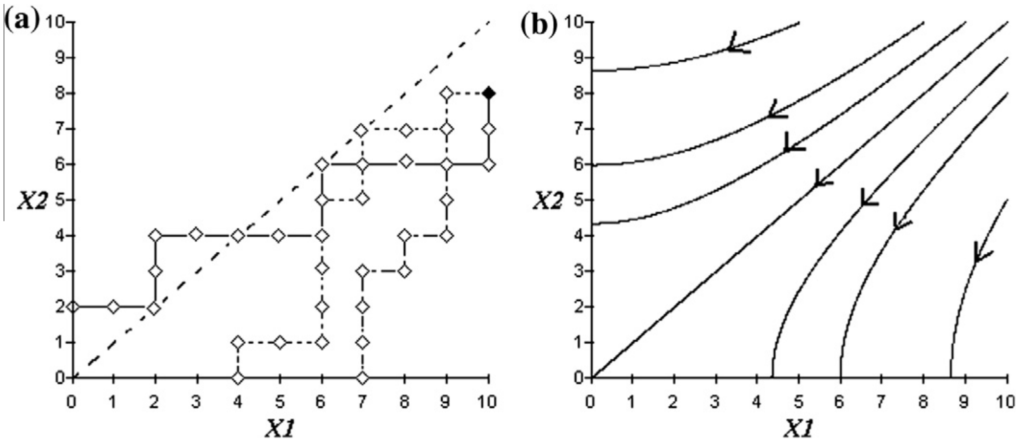


Fig. 3. (a) Three stochastic simulation runs starting at (10,8). The dotted diagonal demarcates the areas where x_1 and x_2 have the advantages. (b) Deterministic trajectories in form of hyperbolas. The straight line of equilibrium between the forces has the equation $x_2 = \sqrt{a/b} \cdot x_1$ (here $a = b$).

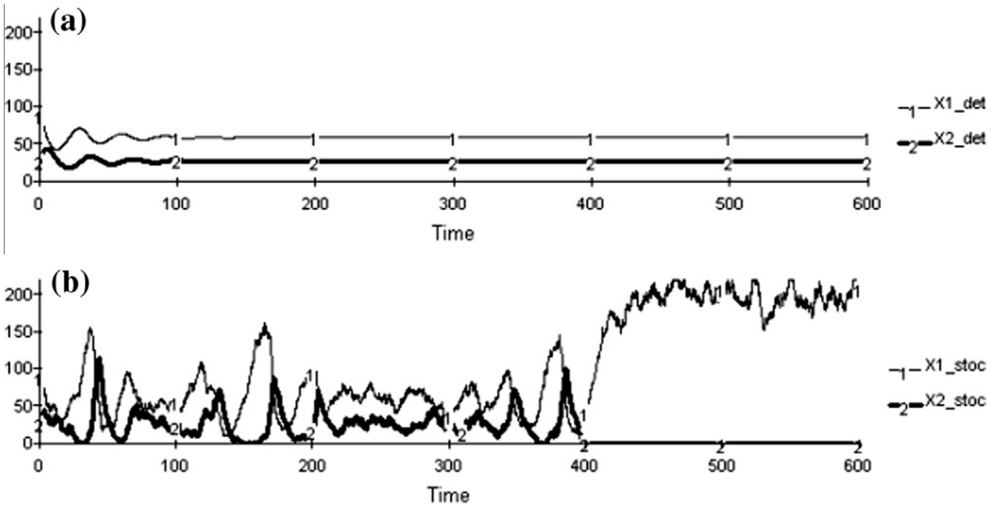


Fig. 4. Results from the prey-predator models of Example 3 with $a = 0.2$, $b = 0.005$, $c = 0.005$, $d = 0.3$, $k = 0.001$ and $x1(0) = 90$, $x2(0) = 28$. (a) Deterministic simulation. (b) A Poisson Simulation run. The stochastic oscillations are persistent, but at about 400 time units the predators x_2 become extinct, making the prey x_1 behave in a logistic manner.

(240, 112). With five parameters, this can be accomplished in different ways. We did so by doubling and quadrupling the values of a and d. The effects of the up-scaling are shown in Table 1.

As seen from Table 1, the deterministic equilibrium values were somewhat lower for x_1 and higher for x_2 than the stochastic averages (until extinction or 1000 time units). In relative terms this deviation of course decreases with scale.

The risk of extinction in the stochastic model decreases heavily for the larger scale. The deterministic model cannot pick up this risk, but for a sufficiently large scale this effect can be neglected and the deterministic model can be used.

The average distance from the equilibrium point for the stochastic model is calculated by: $Dist(t) = \sqrt{(x_1(t) - d/c)^2 + (x_2(t) - a/b - kd/bc)^2}$ and $Av\_dist = \int_0^{sim\_time} dist(t)dt/sim\_time$.

As seen the average distance will increase in absolute numbers when the model is scaled up, but it will decrease in relative terms. This gives another view of the deterministic model, which cannot reflect the variations, but at least gives a reasonable average for the variations when the scale 'grants' no extinction.

In the stochastic model variations are continuously generated by demographic stochasticity but these variations are damped by the competition coefficient k. In the deterministic model only the damping remains and takes the model to an equilibrium. The magnitude of k is crucial for possible extinction, and thus affects when a deterministic model might be useful.

5. Large flows where the local linearised dynamics is asymptotically stable

Now to one of our main questions: *Can a population model with demographic stochasticity be simplified into a system of deterministic difference equations without producing biased state variables?* The answer is affirmative for two different cases, discussed in this and the following section.

The first case where deterministic modelling generates unbiased state variables is when the flows stay so large that they can always be considered continuous, and where the dynamics around its trajectory is also locally asymptotically stable. A large flow of entities expressed by $Po[m]$ will then have small and symmetrical

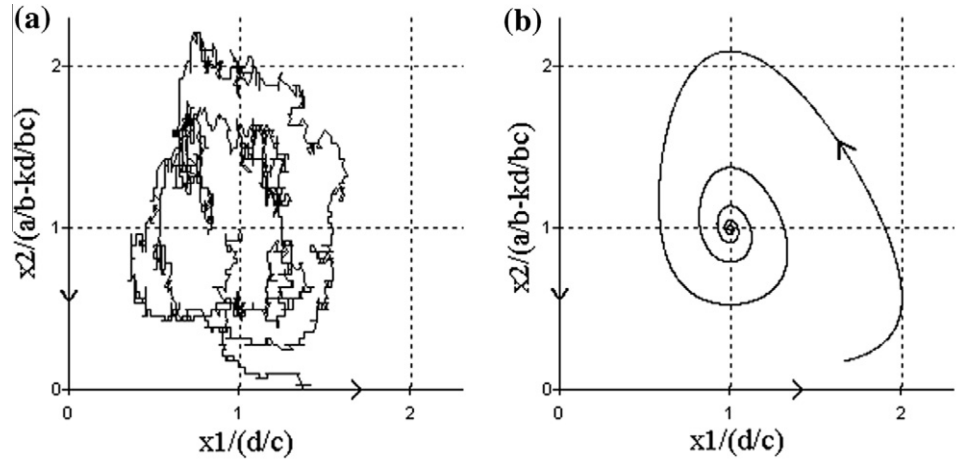


Fig. 5. Phase planes for the scaled variables $x_1/(d/c)$ and $x_2/(a/b - kd/bc)$ with $a = 0.2$, $b = 0.005$, $c = 0.005$, $d = 0.3$, $k = 0.001$. (a) A stochastic trajectory from a simulation run starting at the deterministic equilibrium $x_1(0) = d/c$ and $x_2 = a/b - kd/bc$ hits the absorbing x_1-axes. (b) A deterministic trajectory seeks its equilibrium point: $x_1 = d/c$ and $x_2 = a/b - kd/bc$ from all points within the first quadrant of the phase plane. Starting on the x_1-axis it will continue logistically to $x_1/dc = ac/kd = 3.33$, and starting on the x_2-axis it will proceed to the origin.

Table 1
Results from 1000 simulation runs over 1000 time units of the stochastic model with parameters *a* and *d* doubled and quadrupled.

$(d/c, a/b - kd/bc)$	Stochastic average	Average distance from $(d/c, a/b - kd/bc)$	*P* (extinction before 1000 time units)	Average simulation time (until extinction or 1000 time units)
$(\boldsymbol{60}, \boldsymbol{28})$	(61.5, 26.4)	24.6	88.6%	439.8
$(\boldsymbol{120}, \boldsymbol{56})$	(122.8, 54.7)	35.7	9.8%	949.0
$(\boldsymbol{240}, \boldsymbol{112})$	(242.5, 110.8)	49.6	0.1%	999.8

variations around the mean value m. This is directly seen in that $Po[m]$ approaches $N[m,\sqrt{m}]$ (in $N[\mu,\sigma]$ notation) when m increases. The relative variations $\sqrt{m}/m$ will then approach zero and the normal distribution approaches a point distribution.

More specifically, for one class of models it can be shown that there exist non-linear deterministic models for which $x_d \approx E[x_s(t)]$ for large flows. These models have asymptotically stable, local linearised dynamics that govern perturbations around the trajectory produced by the deterministic model. For large flows and fixed stopping time criteria, the difference between trajectories of the stochastic and the deterministic models can then be approximated as being generated by a set of relatively small zero mean Gaussian disturbances, acting on a stable, linearised perturbation model. The resulting perturbations then have approximately zero means. See Appendix A.

In the literature, a (sufficiently) large population or large subpopulations are sometimes taken as a justification for using a deterministic model [6,11]. The concept '*large population*' is avoided in this paper in favour of '*large flows*', for two reasons. First, the size of a flow, $Po[m]$, decides whether the variations around m can be regarded as relatively small. Second, although 'a sufficiently large population' may also scale the flows to be large, a focus on population size is misleading. A population can be very large, but a flow to, from or between compartments may still be small because of a small transition probability or a small subpopulation. What might happen when one compartment holds a small subpopulation, although the total population is large, is demonstrated by the following example.

Example 4 (*An epidemic SIR model*). Models of infectious diseases are often based on a sequence of the three consecutive stages: Susceptible, Infectious and Recovered. Such a model is therefore denoted an SIR model. The first SIR model was published by Kermack and McKendrick in 1927 [31].

The model has the following setting: The population, consisting of N individuals, is exposed to an infectious disease. Every individual of the population meets every other under equal conditions in each time unit. When a susceptible individual meets an infectious, the risk per time unit that the former becomes infected is b and the expected sojourn time in the infectious stage has an exponential distribution with the average of T time units. Using the Euler algorithm, the stochastic model takes the form:

$$
\begin{aligned}
&S(t+\Delta t) = S(t) - \Delta t \cdot f1(t)\\
&I(t+\Delta t) = I(t) + \Delta t \cdot f1(t) - \Delta t \cdot f2(t)\\
&R(t+\Delta t) = R(t) + \Delta t \cdot f2(t)\\
&\Delta t \cdot f1(t) = Po[\Delta t \cdot b \cdot S(t) \cdot I(t)]\\
&\Delta t \cdot f2(t) = Po[\Delta t \cdot I(t)/T].
\end{aligned}
$$

The embedded, deterministic model is easily obtained by eliminating the $Po[\,]$ operators of the last two equations.

The initial population $N = S + I + R$ here consists of 1000 susceptible individuals, a single infectious individual and no recovered individuals. The infection rate parameter is set to $b = 0.0003$ per individual and time unit and the sojourn time in the infectious compartment is $T = 4$ time units.

Simulations of the stochastic SIR model to study the number of susceptible individuals being infected, S(0)−S(∞), give an ensemble of results with the average value of 53.1 [*95% C.I. 50.8–55.5*] individuals falling ill. However, the corresponding deterministic model produces 318.5 individuals. Thus the large relative error [(Deterministic result–Stochastic estimate)/Stochastic estimate] is 500 percent.

The reason why the deterministic model distorts the results is because the initial infectious subpopulation $I(0)$ is small, although the total population N is large. There is then a large chance that the single infectious individual will recover before infecting another susceptible, so that no epidemic will occur. The probability of such an 'extinction of an outbreak' is $P[extinction] \approx R_0^{-I(0),}$ (and gives $P[outbreak] \approx 1 - R_0^{-I(0)}$); where $R_0 = b \cdot S(0) \cdot T$ is 'the basic reproduction number'. For $R_0 = 1.2$ we get $P[\text{outbreak}] \approx 0.1667$. Considering this we can adjust the deterministic results by $P[outbreak] \cdot Deterministic\_result \approx 0.1667 \cdot 318.5 = 53.08$. This formula gives a good adjustment for large N and small $I(0)$ and when $R_0 > 1$.

The parameter values and the initial conditions chosen in this example created a very large relative error. Now we will investigate what happens for other parameter values of $R_0 = b \cdot S(0) \cdot T$ and initial values $I(0)$ of the initial number of infectious.

To explain what happens for different $I(0)$ and R_0, we regard the state space of S versus I in Fig. 6. (The number of recovered, R, is not of interest here since it does not intervene in the equations of S and I. R can afterwards be calculated algebraically by $R = N - S - I$.)

To study the scale effect of different sizes of the population, N, we can rewrite $b = \beta/N$ (e.g. $\beta = 3$ and $N \approx S$). Then when starting with $I(0)/N$ the fraction of the population acquiring the disease is the same for e.g. $N \approx S = 1000$, 10000, 100000 and 1000000 as well as the trajectories in the phase space of S/N versus I/N.

The relative errors of the number of infected individuals $(S(0) - S(End) \equiv R - I(0))$ caused by using a deterministic model are shown in Table 2 for $R_0 = 1$, 1.2, 2, 5 and 10 (obtained by varying b) and for $I(0) = 1$, 2, 5 and 10.

As seen from Table 2, the error from using a deterministic model is significant if one of R_0 and $I(0)$ is not large. Remember that an epidemic in a population usually starts with one or a few infectious individuals. □

The stochastic model (3) presented in Section 2 updates $x_{s,i}(t)$ over the time interval $(t,t+\Delta t)$ by the terms $Po[\Delta t \cdot g_i(\boldsymbol{x}_s(t),t)]$ and $Po[\Delta t \cdot h_i(\boldsymbol{x}_s(t),t)]$, where Δt for dynamic reasons has to be sufficiently smaller than the smallest time constant of the model. However, Δt can be arbitrarily small to increase precision and $E[Po[\Delta t \cdot f]] = E[Po[\frac{1}{2}\Delta t \cdot f] + Po[\frac{1}{2}\Delta t \cdot f]]$ so $\Delta t \cdot f$ is not a good measure of the size of m when evaluating how $Po[m]$ approaches $N[m, \sqrt{m}]$. The size of $f(\boldsymbol{x}_s(t),t)$ is also not a suitable option because its dimension is the number of entities per time unit, and we can use any period (century, month, second) as time unit.

Therefore, the size of m should be the number transferred during a time interval $(t,t+\Delta\tau)$, where $\Delta\tau$ is the time during which $f(\boldsymbol{x}_s(t),t)$ can be regarded as sufficiently constant. It is not a trivial task to estimate when the stochastic flows are large enough to be approximated by deterministic flows without producing too much bias in the state variables. This complexity is best avoided by *testing* whether the stochastic model and its embedded deterministic model will produce sufficiently similar results for the particular question at study. See Section 10.

For locally unstable non-linear deterministic models, it is possible to construct simple stochastic models of arbitrarily large populations and flows where the embedded deterministic model produces large biases in the state variables, as shown by the next example.

Example 5 (*Lotka–Volterra model*). The stochastic model of Example 3, but without the competition term, is given by:

$$x_1(t+\Delta t) = x_1(t) + Po[\Delta t \cdot a \cdot x_1(t)] - Po[\Delta t \cdot b \cdot x_1(t)x_2(t)]$$
$$x_2(t+\Delta t) = x_2(t) + Po[\Delta t \cdot c \cdot x_1(t)x_2(t)] - Po[\Delta t \cdot d \cdot x_2(t)]$$

and its embedded deterministic model is:

$$x_1(t+\Delta t) = x_1(t) + \Delta t \cdot a \cdot x_1(t) - \Delta t \cdot b \cdot x_1(t)x_2(t)$$
$$x_2(t+\Delta t) = x_2(t) + \Delta t \cdot c \cdot x_1(t)x_2(t) - \Delta t \cdot d \cdot x_2(t)$$

Both models orbit in $x_1 - x_2$ space around the equilibrium point $x_{1eq} = d/c$, $x_{2eq} = a/b$. For the deterministic model, the local linearised dynamics governing perturbations will not be asymptotically stable over the whole orbit cycle.

Now, we can choose the parameters a, b, c and d so that x_{1eq} and x_{2eq} take any large values. Then by choosing $x_1(0)$ and $x_2(0)$, the trajectory is an orbit that has e.g. $x_{1min} = x_{1eq}/2$ and $x_{2min} = x_{2eq}/2$ so that both $x_1(t)$ and $x_2(t)$ and all flows remain very large for all t. However, this will *not* grant unbiased state variables. The stochastic model will sometimes jump outwards and sometimes inwards in a random-walk like fashion, but larger orbits require somewhat longer orbit times than smaller orbits. Therefore, the phase of the stochastic orbits will drift somewhat relative to that of the deterministic orbit. This drift will accumulate over time.

Now, consider the state variable $x_1(T)$ after a long time T. Then $E[x_{1s}(T)]$ will be close to x_{1eq}, while the deterministic state variable $x_{1d}(T)$ can take any value between $0.5x_{1eq}$ and (about) $1.5x_{1eq}$. In particular, if $x_{1d}(T)$ is a peak value, the bias can be as large as $1/2x_{1eq}$. □

6. Positive linear systems studied over a fixed time interval

We now investigate whether any class of embedded deterministic models will produce unbiased state variables (relative to a demographically stochastic population model) when the flows are *not* large. For a subclass of linear systems the answer is yes, provided that the question under study can be answered by simulation over a fixed time interval.

6.1. Heuristic discussion of conditions for unbiased state variables

Consider one equation of the demographically stochastic model given by the system of Eq. (3) where the index i is dropped:

$$x_s(t+\Delta t) = x_s(t) + Po[\Delta t \cdot g(\boldsymbol{x}_s(t),t)] - Po[\Delta t \cdot h(\boldsymbol{x}_s(t),t)].$$

Consider the corresponding equation of the embedded deterministic model (4):

$$x_d(t+\Delta t) = x_d(t) + \Delta t \cdot g(\boldsymbol{x}_d(t),t) - \Delta t \cdot h(\boldsymbol{x}_d(t),t).$$

Now assume that the deterministic model has produced unbiased state variables up to time t. Then to remain unbiased, it is required that:

$$E[x_s(t+\Delta t)] = x_d(t+\Delta t), E[x_s(t+2\Delta t)] = x_d(t+2\Delta t), \ldots$$

Since x is only updated by inflows (i.e. by $Po[\Delta t \cdot g(\boldsymbol{x}_s(t),t)]$ and $\Delta t \cdot g(\boldsymbol{x}_d(t),t)$) and/or by outflows (i.e. by $Po[\Delta t \cdot h(\boldsymbol{x}_s(t),t)]$ and $\Delta t \cdot h(\boldsymbol{x}_d(t),t)$), unbiasedness requires that:

$$E[Po[\Delta t \cdot g(\boldsymbol{x}(t),t)]] - E[Po[\Delta t \cdot h(\boldsymbol{x}(t),t)]]$$
$$= \Delta t \cdot g(\boldsymbol{x}(t),t) - \Delta t \cdot h(\boldsymbol{x}(t),t) \text{ is true for all } t.$$

A sufficient condition for this to hold is that:

$$E[Po[\Delta t \cdot g(\boldsymbol{x}(t),t)]] = \Delta t \cdot g(\boldsymbol{x}(t),t) \text{ and } E[Po[\Delta t \cdot h(\boldsymbol{x}(t),t)]]$$
$$= \Delta t \cdot h(\boldsymbol{x}(t),t), \text{ for all } t.$$

6.1.1. Reason for requiring linearity

When $Po[\Delta t \cdot f(\boldsymbol{x}_s(t),t)]$ (where f denotes g or h) is executed, the argument is sent to a random number generator returning a Poisson-distributed outcome that (provided the argument is nonnegative) on average equals the argument; i.e. $E[Po[\Delta t \cdot f(\boldsymbol{x}_s(t),t)]] = \Delta t \cdot f(\boldsymbol{x}_s(t),t)$. This means that if $x_s(t) = x_d(t)$ then $E[x_s(t+\Delta t)] = x_d(t+\Delta t)$. However, the result of the Poisson sampling from many simulation runs is an ensemble of dispersed outcomes of $x_s(t+\Delta t)$. This ensemble dispersion will create bias for the next time-step $t+2\Delta t$ unless the argument is a linear function of the state variables. Thus, linearity is a *necessary* condition to keep exact unbiasedness of the state variables from time $t+2\Delta t$ onwards.

6.1.2. Reason for requiring positivity[6]

A real population or subpopulation can, of course, not be smaller than zero. However, for a model (not guarded by $x_i \geqslant 0$) the size of a sub-population x_i may become negative. (This can happen in the linear *Lanchester's model of warfare*, see Example 2 in Section 4). One then has the choice of either including a guarding mechanism that after extinction prevents further reduction of the subpopulation or of ending the analysis/simulation run when extinction happens. If one wants to continue the solution/execution of the model beyond the time of extinction of the subpopulation, the guarding mechanism will introduce a non-linearity. On the other hand, if the analysis/simulation run is ended when the extinction occurs, one cannot use a fixed stopping time criterion. In both cases, the unbiasedness is lost. Thus, positivity is a necessary condition for unbiasedness of the state variables.

6.1.3. Reason for requiring a fixed stopping time criterion

That linearity and positivity together are not sufficient to secure unbiased results is shown by the following example.

Example 6 (*Time independent failure rate*). Consider a simple first order, positive, linear model describing the failure of N entities, e.g. lamps in a room. Let $x(t)$ represent the number of functioning (or surviving) entities at time t, with the failure rate constant a per time unit. The stochastic population model: $x(t+\Delta t) = x(t) - Po[\Delta t \cdot a \cdot x(t)]$ and the embedded deterministic model: $x(t+\Delta t) = x(t) - \Delta t \cdot a \cdot x(t)$ of $x(0) = N$ functioning entities are studied. For all fixed time intervals, 0 to T, the expected number of functioning entities $E[x(T)]$ in a stochastic model will then be equal to the state variable value of the deterministic model.

However, the answer to questions such as 'When does there remain M functioning entities?' ($M < N$) will produce different answers for the stochastic and the deterministic models. For $a = 0.1$ time units^{-1} and $N = 10$, the deterministic model will have $M = 1$ remaining functioning entities at 23.0 time units, while the stochastic model gives an estimated 19.3 time units until $M = 1$ (95% *C.I. 19.1 to 19.4*). The deterministic result is thus here 19 percent too large. The time until all entities fail is 29.3 time units (95% *C.I. 29.0 to 29.5*) according to the stochastic model, rather than infinity (obtained from the deterministic model).

---

[6] The concept 'positive system' includes both the case $x_i > 0$ and the case $x_i \geqslant 0$. In this paper we are mainly interested in the non-negative case $x_i \geqslant 0$.

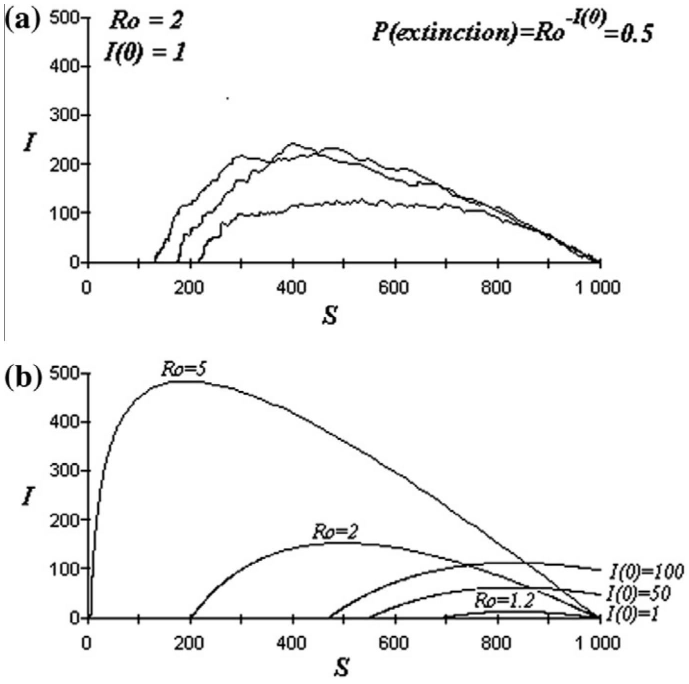


Fig. 6. A phase plane over S and I of the SIR model. (a) Six replications of the stochastic model (whereof three with extinction). (b) Trajectories for the deterministic model with $S(0) = 1000$ and $I(0) = 1$, 50 and 100 infectious individuals with $Ro = 1.2$ and also for $I(0) = 1$ with $Ro = 1.2$, 2 and 5 are shown.

Table 2
Relative errors caused by the use of a deterministic model for some R_0 and $I(0)$. The table area where the relative errors from using a deterministic model is below 5 percent is given in bold.

	$I(0)$			
	1	2	5	10
R_0				
1	2.42	1.44	0.66	0.32
1.2	5.00	2.26	0.72	0.28
2	0.95	0.33	**0.03**	**0.003**
5	0.24	**0.04**	**0.001**	**0.000**
10	0.11	**0.014**	**0.000**	**0.000**

We see that the size of the error from the deterministic model is related to the final number of functioning entities when using a time-indefinite stopping criterion. The error is only large when this final number is small or zero. □

6.2. Positive linear deterministic and stochastic models studied over a fixed time interval

So far, we have found that non-linearity, populations that must be protected from being negative, or criteria other than those based on a fixed study time will produce biased state variables in a deterministic model, in cases where flows cannot be ensured to be large at all times. In this subsection we therefore restrict the study to the class of so-called positive (non-negative) linear systems studied under a fixed stopping time criterion. The following theorem relates the solutions of a stochastic model and its embedded deterministic model.

Theorem 1. *For a stochastic system of the linear form*

$$\begin{aligned} x_{s,i}(t+\Delta t) &= x_{s,i}(t) + \Sigma_j Po[\Delta t \cdot a_{ij} \cdot x_{s,j}(t)] + Po[\Delta t \cdot b_i(t)] \\ &\quad - \Sigma_j Po[\Delta t \cdot c_{ij} \cdot x_{s,j}(t)] - Po[\Delta t \cdot d_i(t)]; \quad i = 1 \ldots n \end{aligned} \tag{5}$$

where all a_{ij}, b_i, c_{ij} and $d_i \geqslant 0$, the expected value of the solution over time equals the solution of the embedded deterministic system

$$\begin{aligned} x_{d,i}(t+\Delta t) &= x_{d,i}(t) + \Sigma_j \Delta t \cdot a_{ij} \cdot x_{d,j}(t) + \Delta t \cdot b_i(t) \\ &\quad - \Sigma_j \Delta t \cdot c_{ij} x_{d,j}(t) - \Delta t \cdot d_i(t); \quad i = 1 \ldots n \end{aligned} \tag{6}$$

as long as all state variables $x_{s,i}$ and $x_{d,i}$ remain non-negative.

For a proof, see Appendix B.

The task now is to characterise when the system has the property that its state variables, x_i, will remain non-negative. This issue is first examined for the deterministic model, for which there exist necessary and sufficient conditions. For the stochastic model, sufficient conditions exist but are stricter than those for the corresponding deterministic model.

6.2.1. Positivity for deterministic linear models

A model (4) that is linear implies that: $\boldsymbol{g} = \boldsymbol{Ax} + \boldsymbol{b}$ and $\boldsymbol{h} = \boldsymbol{Cx} + \boldsymbol{d}$, where $\boldsymbol{g} = (g_1(\boldsymbol{x}(t),t),\ldots,g_n(\boldsymbol{x}(t),t))^T$ and $\boldsymbol{h} = (h_1(\boldsymbol{x}(t),t),\ldots,h_n(\boldsymbol{x}(t),t))^T$. Then the deterministic model in vector form (without the guard to keep $x_i \geqslant 0$) is:

$$\begin{aligned} \boldsymbol{x}_d(t+\Delta t) &= \boldsymbol{x}_d(t) + \Delta t \cdot [(\boldsymbol{A} - \boldsymbol{C})\boldsymbol{x}_d(t) + (\boldsymbol{b} - \boldsymbol{d})] \\ &= \boldsymbol{x}_d(t) + \Delta t \cdot [\boldsymbol{Q}\boldsymbol{x}_d(t) + \boldsymbol{r}]. \end{aligned} \tag{7}$$

Linear systems $d\boldsymbol{x}/dt = \boldsymbol{Qx} + \boldsymbol{r}$ that remain non-negative, i.e. where all $x_i \geqslant 0$ for all time, are denoted *positive (non-negative) linear systems.*

For positive linear systems there exists a number of powerful theorems for analysis. For example, dominant eigenvalues simplify the analysis of long-term behaviour. Furthermore, there is a direct correspondence between the existence of a positive equilibrium point and stability. For a stable system at rest at its equilibrium point, there are powerful statements about what happens when some parameter of the system is slightly modified, see [8].

A *homogeneous nth* order linear differential equation has the form: $d\boldsymbol{x}(t)/dt = \boldsymbol{Q}\boldsymbol{x}(t)$; where $\boldsymbol{Q}$ is a real $n \times n$ system matrix with the elements q_{ij}. If $q_{ij} \geqslant 0$ for all $i \neq j$, in other words if all nondiagonal elements are non-negative, then $\boldsymbol{Q}$ is called a *Metzler matrix*. This is *a necessary and sufficient condition* for the homogeneous linear system to be non-negative.

Example 6 continued: A deterministic model of entities with time independent failure rate is: $dx/dt = -ax$. Since only the diagonal element '$-a$' is negative, $\boldsymbol{Q} = [-a]$ is a Metzler matrix and the model is a positive (non-negative) linear system, so x will always remain non-negative. □

Example 2 continued: Lanchester's model of warfare is a linear system but the matrix $\boldsymbol{Q}$ is not a Metzler matrix since the elements $q_{12} = -b$ and $q_{21} = -a$ are negative, so this system is not a non-negative one. □

A *non-homogeneous* linear differential equation has an additional external driving force vector $\boldsymbol{r}$ of n components r_i; $i = 1,2,\ldots,n$. The system then has the form: $d\boldsymbol{x}(t)/dt = \boldsymbol{Qx}(t) + \boldsymbol{r}$. This system is positive if and only if $\boldsymbol{Q}$ is a Metzler matrix *and* the vector $\boldsymbol{r}$ is non-negative (all $r_i \geqslant 0$).

6.2.2. Positivity for stochastic linear models

For a stochastic linear model

$$\boldsymbol{x}_s(t+\Delta t) = \boldsymbol{x}_s(t) + Po[\Delta t \cdot \boldsymbol{Ax}_s(t)] + Po[\Delta t \cdot \boldsymbol{b}] - Po[\Delta t \cdot \boldsymbol{Cx}_s(t)] - Po[\Delta t \cdot \boldsymbol{d}], \quad (8)$$

chance may still produce a larger number of departures than of arrivals causing a component of $\boldsymbol{x}_s$ to be negative, even when the corresponding deterministic model remains positive.

To ensure that all sub-populations in (8) remain non-negative, inflows by $\boldsymbol{A}$ and $\boldsymbol{b}$ will cause no problem, but the $\boldsymbol{C}$-matrix must be limited to diagonal elements and the vector $\boldsymbol{d}$ must be removed.

The sub-class of linear stochastic models (8) that will grant positivity (non-negativity) in a general case is very restricted. However, a specific model where the $\boldsymbol{C}$-matrix has non-zero elements off the diagonal, or where the $\boldsymbol{d}$-vector has non-zero elements, may in practice stay positive during a specified time interval of study. However, this must then be tested for that specific model under given initial and parameter values and the actual stopping criterion, as outlined in Section 10.

7. Bias from neglecting initial value stochasticity

So far, only the consequences of removing demographic stochasticity have been examined. Stochastics can also enter into initial values (1E). Consideration of initial value stochasticity may be important in a number of situations. For example:

- The initial situation is unclear because of lacking information, bad measurements, etc.
- In a randomised cohort trial, one cohort of n cases to be exposed and another of m reference cases to remain unexposed are randomly sampled from a population to see whether they will develop a certain disease during a follow-up period. However, a certain fraction p of the population already has the disease, although not diagnosed. If p is not negligibly small, the exposed and unexposed cohorts must be initialised by $Bi[n,p]$ and $Bi[m,p]$ distributions, respectively. Otherwise, the variations and thus the confidence intervals around the results will become too small.
- There are also cases where the distributions of initial values are theoretically known. A linear population system (without physical feedback) starting in a stationary state is being studied. The state variables are then Poisson-distributed around an average value [26,27,32,33]. To not disturb the stationary situation, the initial values must also be drawn from these distributions.

In this section the issue is what might happen when stochastic initial values are replaced by deterministic averages in terms of bias in state variables.

Theorem 1 states that for positive linear models, the deterministic model will follow the average of the demographic stochastic model, even though the solution of the latter model for each step is split up in a dispersed distribution. In the proof of this theorem (Appendix B), this split-up after one time-step can be taken as a distribution of initial values, although here it is based on the Poisson distribution. However, 'Step II' of this proof does not depend on a particular distribution. It requires only that the split-up distribution has the same average as the deterministic model, and that this distribution is normalised to unity. Thus, stochasticity in initial values will not create biased state variables in a linear model as long as the state variables remain nonnegative.[7]

However, for a non-linear model, the situation is different.

Example 7 (*SIR epidemics with random initial values*). Assume that a group of $n = 3$ persons are returning after visiting an infected area. For simplicity, we assume equal probabilities of 0, 1, 2 and 3 persons returning infected.

With four models we investigate a possible epidemic among the population of 1000 individuals: (1) A stochastic model where demographic and initial value stochasticity are included. (2) A model with only demographic stochasticity. (3) A model with only initial value stochasticity. (4) A deterministic model.

Assumptions from Example 4 are used: $b = 0.0003$ per individual and time unit, $T = 4$ time units. Further, in the cases of deterministic initial values, the average initial number of infectious persons, $Av\_I0 = 1.5$, is used. To keep the total population constant at 1000 persons, the number of susceptibles, S, is initiated to: $S0 = 1000 - Av\_I0$.

As can be seen from Table 3, removing initial value stochasticity increases the expected number of persons (including the originally infected) getting the disease from 71.5 to 98. Using the 'fully stochastic model '(1) as reference this gives a relative error of 37%. Keeping the initial values stochastic but instead removing the demographic stochasticity changes the number of infected persons to 239.5 (a relative error of 235%), and removing both demographic and initial value stochasticity changes it to 317 (a relative error of 343%). The deterministic model may thus obtain a considerable bias, and neglecting initial value stochasticity contributes to this bias. □

8. Bias from neglecting environmental stochasticity

The concept 'environmental stochasticity' [1,2] is somewhat vague. Here it is used as synonymous with parametric stochasticity. Environmental stochasticities are introduced in the parameter equations (1F) by making parameters stochastic.

Unlike the intrinsic demographic stochasticity, which is completely specified by the model structure and the actual number of entities, environmental stochasticity is used to describe lack of knowledge about various types of influences and irregularities from the environment. Such irregularities may be variations in temperature, wind, precipitation, food supply, pollution, noise, or disturbances of any kind. To include such external irregularities in a model, the value of a parameter is sampled from an appropriate statistical distribution to describe the variations in e.g. fertility, mortality, risk, reaction speed, etc. This statistical distribution may have several parameters. When and how often a new sample from

---

[7] Of course, the initial distribution must cover only positive values. For example a $Normal[\mu,\sigma]$ distribution always has a (perhaps theoretical) risk of returning negative values!

the distribution should be drawn may also be described by some statistical function. There may also be correlation between consecutive sample values.

Here we are interested in whether elimination of environmental stochastics might introduce bias in state variables. Using (3), elimination of environmental stochastics would produce a vanishing approximation error in the state variables if

$$\begin{aligned} x_{s,i}(t+\Delta t) &= x_{s,i}(t) + Po[\Delta t\cdot g_i(\boldsymbol{x}_s(t),\boldsymbol{p}_s(t),t)] \\ &\quad - Po[\Delta t\cdot h_i(\boldsymbol{x}_s(t),\boldsymbol{p}_s(t),t)] \text{ and } x_i \geqslant 0;\ i=1\ldots n, \\ &\approx x_{s,i}(t) + Po[\Delta t\cdot E[g_i(\boldsymbol{x}_s(t),\boldsymbol{p}_s(t),t)]] \\ &\quad - Po[\Delta t\cdot E[h_i(\boldsymbol{x}_s(t),\boldsymbol{p}_s(t),t)]] \text{ and } x_i \geqslant 0;\quad i=1\ldots n, \end{aligned} \tag{9}$$

where $E[.]$ represents averaging with respect to the parameter vector $\boldsymbol{p}_s(t)$ with the components $p_{s,j}$, $j = 1\ldots q$. (Removal of also the demographic stochasticity would then result in the deterministic model (4).)

Approximation of solutions of difference, or differential equations, by averaging over some of the variables, in particular averaging over high-frequency periodic functions, has a long tradition. Averaging over stochastic parameters has been used e.g. in the analysis of recursive parameter estimation algorithms [34–36]. While not directly applicable to the present problem, these methods and results indicate that one can only hope to attain the approximations (9) asymptotically, in the limit of fast time variability of the parameters $p_{s,j}(t)$ relative to the variability of all state variables. Furthermore, it would in general lead to erroneous results to average the parameters directly, by using $p_{av,j} = E[p_{s,j}(t)]$. These points are illustrated below.

8.1. Averaging over fast time-varying parameters

Example 8 (*An exponential growth model with fast time-varying environmental stochasticity*). Let a population $x(t)$ grow according to the first order equation:

$$x(t+\Delta t) = x(t) + Po[\Delta t\cdot x(t)/\tau(t)],\ x(0) = 1, \tag{10}$$

where the time constant $\tau(t)$ is a randomly varying parameter. This linear model is studied during the fixed time interval $0\ldots T$, where $T = 10$ time units. Assuming $\tau(t)$ to have a uniform distribution between 1 and 3 time units, $U[1,3]$ so that the pdf $f(\tau)$ has the value 1/2 for $1 \leqslant \tau \leqslant 3$ and zero otherwise. It is assumed here that $\tau(t)$ varies over time very much faster than the time constant of the process (1 to 3 time units); say a new sample after each interval of $\delta = 0.02$ time units.

Table 3
The number of susceptible individuals falling ill, $S(0)$–$S(\infty)$, with 95% C.I. based on 10,000 simulation runs.

Model type	Number of infected
(1) Demographic and initial value stochasticity	71.5 (68.8–74.2)
(2) Only demographic stochasticity	98.0 (95.0–100.9)
(3) Only initial value stochasticity	239.5 (236.7–242.2)
(4) Deterministic	317.0

Three models are compared. Model 1 has both demographic and environmental stochasticity, Model 2 only demographic stochasticity and model 3 is deterministic. Averaging of Model 1 by (9) gives the model with demographic stochasticity only:

$$x(t+\Delta t) = x(t) + Po[\Delta t\cdot x(t)E[1/\tau(t)]] = x(t) + Po[\Delta t\cdot x(t)[1/\tau]_{av}],$$

where

$$[1/\tau]_{av} = \int_1^3 \left(\frac{1}{\tau(t)}\right) f(\tau)d\tau = \frac{\ln(3)}{2} \approx \frac{1}{1.8205},$$

corresponding to an effective average: $\tau_{av} = 1.8205$.

As a comparison, a direct arithmetic average over the $U[1,3]$ distribution of $\tau(t)$ would give the fixed parameter value $\tau_{aa} = 2$. The expected values of $x(T)$ based on 10,000 simulation runs are shown in Table 4.

With appropriately performed averaging of the very fast time-varying parameter τ, very good agreement is obtained. Using the arithmetic average $\tau_{aa} = 2$ gives very biased state variables (about 39 percent too small). □

8.2. Slow parameter variations

We proceed with more realistic cases where the parameter $\tau(t)$ changes slowly or remains constant within a simulation run

.

Example 8 continued: The population $x(t)$ grows according to (10) where $\tau(t)$ is sampled from a $U[1,3]$ distribution. We investigate the omission of environmental stochasticity when the values of $\tau(t)$ are to be resampled with different sampling intervals, δ: The cases $\delta \ll$ *the (varying) time constant*, $\delta \approx$ *the time constant*, and $\delta \gg$ *the time constant* (the last case implying a new, fixed value of τ only for each simulation run) are considered. The average value of τ_{av} is set to $2/ln(3) \approx 1.8205$ when environmental stochasticity is omitted. The expected values of $x(T)$ of the three models with different sampling intervals δ are summarised in Table 5.

The expected value of $x(T)$ of the model with both environmental and demographic stochasticity varies strongly with the resampling interval δ. This is also the case for a model with only environmental stochasticity (not shown) giving values close to those of Model 1 in Table 5. However the deterministic model and the model with only demographic stochasticity produce very biased estimates of $x(T)$ when the sampling time is not very short. The relative error for the deterministic model is negligible for

Table 4
The value of $x(T)$ with 95% C.I. based on 10,000 simulation runs.

Fixed τ for Models 2 and 3	Model 1 (x_{ss}): environmental and demographic stochasticity	Model 2 (x_s): only demographic stochasticity	Model 3 (x_d): deterministic
$\tau_{av} = 1.8205$	235.7(231.1–240.2)	236.0 (231.4–240.6)	235.8
$\tau_{aa} = 2$	same as above	143.7 (140.9–146.5)	144.8

Table 5
The expected value of $x(T)$ with 95% C.I. based on 10,000 simulation runs.

Sampling interval	Model 1: environmental and demographic stochasticity	Model 2: only demographic stochasticity	Model 3: deterministic
$\delta = 0.02$ time units	235.7 (231.1–240.2)	236.0 (231.4–240.6)	235.8
$\delta = 1$ time unit	276.9 (269.4–284.3)	same as above	same as above
$\delta = 5$ time units	575.9 (540.6–611.2)	same as above	same as above
$\delta = 10 = T$ time units	1358.8 (1262.7–1454.9)	same as above	same as above

$\delta = 0.02$ time units but rises to -15% for $\delta = 1$, to -59% for $\delta = 5$, and to -83% for $\delta = 10$ time units.

Thus, neglecting environmental stochasticity, especially stochastics that produce slow time variations of parameters of the simulation models, may cause large bias. A linear model simulated over a fixed time interval was selected to emphasise that bias due to neglecting environmental stochasticity is not a phenomenon restricted to non-linear dynamics. An unstable model was selected in these examples to amplify the studied effects, but similar phenomena can be observed for locally stable models. □

This is just one example of the hazard of relying on deterministic models. The hidden danger is that the modeller will not see the implicit and often erroneous simplifications of using average values, because of lacking a reference model with environmental stochasticity.

9. Biased results caused by non-linear output functions

Now, we discuss the coupling between state variables and study results defined in terms of the output functions. The output function may create or increase bias from unbiased or little biased state variables. The question is here technically formulated as: When does a deterministic model with unbiased *state variables* also produce unbiased *results*?

The algebraic and deterministic output equations in the models (3) and (4) are:

$$y_{s,k}(t) = c_k(\boldsymbol{x}_s(t), t), \quad k = 1 \dots m \tag{11}$$

and

$$y_{d,k}(t) = c_k(\boldsymbol{x}_d(t), t), \quad k = 1 \dots m \tag{12}$$

The output can e.g. be a state variable of interest, $y = x_2$, or the biomass of three species, $y = w_1x_1 + w_2x_2 + w_3x_3$, where w_i are average weights for the species. It can be a ratio ($y = x_1/x_2$), a fraction ($y = x_1/(x_1 + x_2)$) or a non-linear cost function $y = f(\boldsymbol{x},t)$. The output y can be a counter of specific events. For example, $dx/dt = ax - bx^2$ will approach a/b logistically. If then y is a counter for the number of times $x(t)$ will exceed $a/b + 1$, very different results may be obtained by the deterministic and the corresponding stochastic model.

Assume linear output functions, $y_{s,k}$, of the state variables:

$$y_{s,k} = c_k(\boldsymbol{x}_s(t), t) = \Sigma_t(\Sigma_i \alpha_{ik}(t) x_{s,i}(t) + \beta_k(t)), \quad i = 1 \dots n, \ t = 0, \Delta t, 2\Delta t \dots T \tag{13}$$

for deterministic parameters $\alpha_{ik}(t)$ and $\beta_k(t) \in R$, and similarly for $y_{d,k}$. If the deterministic state variables are unbiased ($x_{d,i}(t) = E[x_{s,i}(t)]$), then this will also secure unbiased results:

$$E[y_{s,k}(t)] = E[c_k(\boldsymbol{x}_s(t), t)] = \Sigma_t(\Sigma_i \alpha_{ik}(t) E[x_{s,i}(t)] + \beta_k(t)) = \Sigma_t(\Sigma_i \alpha_{ik}(t) x_{d,i}(t) + \beta_k(t)) = y_{d,k}(t), \quad i = 1 \dots n, \ t = 0, \Delta t, 2\Delta t \dots T. \tag{14}$$

On the other hand, if an output function $c_k(\boldsymbol{x}(t),t)$ is non-linear in $\boldsymbol{x}$, then the output y_d will deviate from $E[y_s]$. Say that two simulation runs of the stochastic model give the results $\xi + \Delta\xi$ and $\xi - \Delta\xi$, respectively, and that the deterministic model gives ξ. If the output function is $y = x^2$, then the stochastic result will be $E[y_s] = 1/2[(\xi + \Delta\xi)^2 + (\xi - \Delta\xi)^2] = \xi^2 + \Delta\xi^2$, while $y_d = \xi^2$. For e.g. $\xi = 4$ and $\Delta\xi = 2$ this gives $y_d = 16$ instead of $E[y_s] = 20$ (a relative error of -20%).

Thus unbiased state variables will produce an unbiased result if and only if the output function utilised is linear in the state variables.

10. A pragmatic approach requires testing of the specific model

This paper highlights the phenomena that cause results from a deterministic model of a population to be biased.

In Sections 5 and 6 two mathematical classes of embedded deterministic models securing unbiased state variables, obtained by removing demographic stochasticity, were identified. However, the applicability of these results to a specific simulation study is not completely straightforward. For example, what a ‘sufficiently large’ flow is has to be answered in a pragmatic setting. Furthermore, the class of linear positive systems is very restricted and models outside this class may remain non-negative given the specific initial conditions and time interval of the study. In particular, Theorem 1 states more general conditions for how long a linear deterministic model will remain unbiased. When also considering environmental stochasticity to be omitted in the corresponding deterministic model, some bias seems almost inevitable. Therefore, a complete theoretical approach to the question of unbiasedness remains out of reach.

From a pragmatic viewpoint, exact unbiasedness may not be required. A deterministic model with a sufficiently small bias may still be useful. A pragmatic approach, therefore, should consider a defined model under given conditions and issues to be answered rather than a mathematical class of models. A practical way to test whether a deterministic model will produce sufficiently unbiased results is the following:

Run the stochastic model for, say, 10,000 simulation runs. Then run the deterministic model once and compare whether the outcome is within the confidence interval of the stochastic results. This must be done for every specified purpose of the study.

11. Discussion and conclusions

There is an evident need for a comprehensive theory of population modelling and simulation. The importance of such a theory and of pragmatic guidelines is also magnified because of the very many fields where population modelling and simulation is used. The evolving theory described in [3] explains the relationships between stochastic entity-based (e.g. Discrete Event Simulation), compartment-based (e.g. Poisson Simulation) and state-based (e.g. Markov) population models and formulates the conditions for these types of models to be mutually consistent. These types of stochastic population models all incorporate demographic stochasticity as an intrinsic property.

The current paper first expands this theory by exploring the conditions for when demographic stochasticity can be omitted without creating biased state variables. Then the consequences of omitting initial value and environmental stochasticity are studied, but in less detail. Finally it is shown that the output function, relating the state variables to study results, may also create bias.

A main conclusion is that no deterministic model of a population will produce fully unbiased results for all possible questions under study. Whether a deterministic model under the actual conditions will produce insignificantly biased or unbiased results should therefore usually be tested, by comparing its results with those of a corresponding stochastic model.

The restricted classes of population models that have been found where various types of stochasticity can be removed without biasing the state variables or the results can be summarised as follows:

- If the flows in a model with demographic stochastics only stay large and if the local linearised dynamics is asymptotically stable, then unbiased state variables are approached in the limit of large flows, over a fixed time interval.

- For the class of linear dynamic systems, Theorem 1 states that a deterministic model will produce unbiased state variables as long as all state variables remain non-negative in the deterministic and in the demographically stochastic model. This also requires that the question under study is compatible with a solution over a fixed time interval. Unbiasedness here holds for demographic stochasticity and initial value stochasticity.
- Unbiased state variables will produce unbiased results if, and only if, the output function utilised is linear in the state variables.

A deterministic model remains useful for special tasks, provided that (sufficiently) unbiased results are produced. Such tasks can be to obtain insights from mathematical analysis of average behaviour or to simplify parameter estimation, optimisation or sensitivity analysis.

Broadening the scope beyond this study to a pragmatic situation where the type of population model is to be selected (the whole of Fig. 1), we advocate the use of a stochastic model. The main reason (except for possible bias) is that a deterministic model, even when it produces unbiased results, will not say anything about the variations. Even when an average of many simulation runs of a stochastic model lies close to that from a deterministic model, a single simulation run may well differ quite markedly from it. If there is knowledge about the structure of the system under study (and for population models there usually is),[8] at least demographic stochasticity should usually be included (model (1) instead of model (2)) to give some hint of the size of the stochastic variations.

Regarding this from another angle, it may be argued that a single set of observations is not sufficient to build and adjust a stochastic model. However, the use of a single set of observations to determine the properties and parameters of a deterministic model is also problematic. The observed system behaviour may represent an extreme outcome given the circumstances. A deterministic model has to take these (possible extreme) observations as hard facts, while a stochastic model will at least span a distribution of outcomes. A similar conclusion is drawn by Renshaw in [13]: “In practice inference is often drawn from a single set of observations, and blind obedience to deterministic results is clearly courting trouble. A deterministic model should always be conducted in parallel with a simulation study of a corresponding stochastic model.”

This investigation shows the treacherous nature of deterministic modelling of populations. One fundamental problem is that the deterministic concepts force us to make a profound distortion of reality by replacing entities by continuous quantities and irregularly occurring events by smooth flows. Although this strongly simplifies the modelling, it often excludes a number of real aspects and phenomena and creates more or less biased results. In Table C.1 in Appendix C the examples in this paper, demonstrating the various reasons for bias, are listed.

From within the conceptual framework of deterministic modelling, the distortion of the entity concept and the omission of stochasticity with subsequent effects on model behaviour is invisible, so the effects of omitting stochasticity may fall outside the scope of a deterministic modelling approach. Therefore, the conclusion is that the modeller must be aware of the quagmire of potential problems and that for all but the most trivial models, the behaviour and results of a deterministic model of populations must be tested against a corresponding stochastic model.

An important practical conclusion of this paper is that modellers should be very restrictive in their use of deterministic modelling of populations. Disciplines dealing with population models (ecology, epidemiology, etc.) should *not* base theories and studies only on deterministic models.

Appendix A. Sufficient conditions for existence of deterministic models with large flows creating asymptotically unbiased results

Consider the stochastic population model (3) and the corresponding embedded deterministic model (4) in vector form, assuming $x_i(t) > 0$ for all $i = 1,\ldots,n$ for a fixed simulation time interval $t \in [0,T]$. Also let $\Delta\tau$ represent the largest time-step acceptable from a dynamic perspective:[9]

$$\boldsymbol{x}_d(t+\Delta\tau) = \boldsymbol{x}_d(t) + \Delta\tau\cdot\boldsymbol{g}(\boldsymbol{x}_d(t),t) - \Delta\tau\cdot\boldsymbol{h}(\boldsymbol{x}_d(t),t), \tag{A.1}$$

$$\boldsymbol{x}_s(t+\Delta\tau) = \boldsymbol{x}_s(t) + Po[\Delta\tau\cdot\boldsymbol{g}(\boldsymbol{x}_s(t),t)] - Po[\Delta\tau\cdot\boldsymbol{h}(\boldsymbol{x}_s(t),t)], \tag{A.2}$$

where $\boldsymbol{x}_d(t) = (x_{d,1},\ldots,x_{d,n})^T$ and $\boldsymbol{x}_s(t) = (x_{s,1},\ldots,x_{s,n})^T$ while $\boldsymbol{g} = (g_1(\cdot),\ldots,g_n(\cdot))^T$, $Po[\Delta\tau\cdot\boldsymbol{g}] = (Po[\Delta\tau\cdot g_1],\ldots,Po[\Delta\tau\cdot g_n])^T$ and likewise for $\boldsymbol{h}(\cdot)$. The two models (A.1) and (A.2) are assumed to be initialised at the same state vector $\boldsymbol{x}(0)$, with $x_i(0) \geqslant 0$, $i = 1,\ldots,n$.

The dynamics of small perturbations $\boldsymbol{\Delta x}(t)$ of the state variables in the deterministic model (A.1), so that $\boldsymbol{x}(t) = \boldsymbol{x}_d(t) + \boldsymbol{\Delta x}(t)$, are given by the local linearised dynamic system:

$$\boldsymbol{\Delta x}(t+\Delta\tau) = \boldsymbol{\Delta x}(t) + \Delta\tau\cdot\boldsymbol{A}(t)\boldsymbol{\Delta x}(t) - \Delta\tau\cdot\boldsymbol{C}(t)\boldsymbol{\Delta x}(t) + \boldsymbol{w}(t), \tag{A.3}$$

where the matrices have components $[\boldsymbol{A}(t)]_{ij} = \partial g_i(\boldsymbol{x}(t),t)/\partial x_j(t)$ and $[\boldsymbol{C}(t)]_{ij} = \partial h_i(\boldsymbol{x}(t),t)/\partial x_j(t)$, all evaluated at $\boldsymbol{x}(t) = \boldsymbol{x}_d(t)$, where all derivatives are assumed to exist for all $t \in [0,T]$. In Eq. (A.3) we have included an additive external perturbation vector $\boldsymbol{w}(t)$ acting at time t. This perturbation must be small relative to $\boldsymbol{x}(t)$ for the linearisation approximation (A.3) to remain valid.

Now assume the following conditions:

1. All components of the flows $\Delta\tau\cdot\boldsymbol{g}(\cdot)$ and $\Delta\tau\cdot\boldsymbol{h}(\cdot)$ in (A.1) have very large numerical values for all $t \in [0,T]$.
2. The local linearised dynamics for a perturbation $\boldsymbol{\Delta x}(t)$, described by (A.3), around $\boldsymbol{x}_d(t)$, described by (A.1), exists and is asymptotically stable for all $t \in [0,T]$.

Then we show that $E[\boldsymbol{x}_s(t)] \approx \boldsymbol{x}_d(t)$, for all $t \in [0,T]$. This is demonstrated as follows. We focus on component i of the state variable vectors in (A.1), (A.2) and (A.3), below denoted $x_{d,i}$, $x_{s,i}$ and Δx_i, respectively. Under Condition 1, all flows stay large and are then approximately normally distributed since:

$$Po[m] \rightarrow N[m,\sqrt{m}] \quad \text{as } m \rightarrow \infty.$$

Therefore, in the following we assume that:

$$\begin{aligned} Po[\Delta\tau\cdot g_i(\boldsymbol{x}_s(t),t)] &\approx \Delta\tau\cdot g_i(\boldsymbol{x}_s(t),t) + (\Delta\tau\cdot g_i(\boldsymbol{x}_s(t),t))^{1/2}e_1(t) \\ &= m_1(t) + m_1^{1/2}e_1(t) \end{aligned} \tag{A.4}$$

$$\begin{aligned} Po[\Delta\tau\cdot h_i(\boldsymbol{x}_s(t),t)] &\approx \Delta\tau\cdot h_i(\boldsymbol{x}_s(t),t) + (\Delta\tau\cdot h_i(\boldsymbol{x}_s(t),t))^{1/2}e_2(t) \\ &= m_2(t) + m_2^{1/2}e_2(t), \end{aligned} \tag{A.5}$$

where $e_1(t)$ and $e_2(t)$ are mutually uncorrelated zero mean Gaussian real-valued variables with unit variance, and where $m_1(t)$ and $m_2(t)$ are assumed to be large for all $t \in [0,T]$.

Both (A.1) and (A.2) are now assumed to be initialised at the same state, so $x_{d,i}(0) = x_{s,i}(0)$ for all i. At the first time-step $(0,\Delta\tau)$, (A.4) and (A.5) give:

$$\begin{aligned} x_{s,i}(\Delta\tau) &= x_{s,i}(0) + Po[m_1(0)] - Po[m_2(0)] \approx x_{d,i}(0) + m_1(0) \\ &\quad - m_2(0) + m_1(0)^{1/2}e_1(0) - m_2(0)^{1/2}e_2(0). \end{aligned} \tag{A.6}$$

---

[8] Structural information such as: who eats whom, the successive sequence of stages, what stages are infectious, etc. is usually known.

[9] In simulation there is no restriction on using any shorter time-step Δt, see also ‘Comments’ at the end of this appendix.

Since $x_{d,i}(\Delta\tau) = x_{d,i}(0) + m_1(0) - m_2(0)$ by (A.1), equation (A.6) describes the state of the stochastic model as a perturbation of the state of the deterministic model:

$$x_{s,i}(\Delta\tau) \approx x_{d,i}(\Delta\tau) + \Delta x_i(\Delta\tau), \tag{A.7}$$

where

$$\Delta x_i(\Delta\tau) = m_1(0)^{1/2} e_1(0) - m_2(0)^{1/2} e_2(0).$$

Note that $E[\Delta x_i(\Delta\tau)] = 0$, when the average is taken w.r.t. $e_1(0)$ and $e_2(0)$. While the perturbation $\Delta x_i(\Delta\tau)$ may be numerically large, it represents a small *relative* perturbation, since $m_j^{1/2}$ are small relative to m_j for large m_j, where $j = 1$ or 2. For sufficiently small relative perturbations $\Delta x_i(\Delta\tau)/x_{d,i}(\Delta\tau)$, the state variable perturbation $\Delta x_i(\Delta\tau)$ will remain within the range of validity of the perturbation model (A.3). Use of this model at time $\Delta\tau$ gives:

$$\begin{aligned} g_i(\boldsymbol{x}_s(\Delta\tau),\Delta\tau) &= g_i(\boldsymbol{x}_d(\Delta\tau) + \boldsymbol{\Delta x}(\Delta\tau),\Delta\tau) \\ &\approx g_i(\boldsymbol{x}_d(\Delta\tau),\Delta\tau) + \boldsymbol{A}_i(\Delta\tau)\boldsymbol{\Delta x}(\Delta\tau) \\ h_i(\boldsymbol{x}_s(\Delta\tau),\Delta\tau) &= h_i(\boldsymbol{x}_d(\Delta\tau) + \boldsymbol{\Delta x}(\Delta\tau),\Delta\tau) \\ &\approx h_i(\boldsymbol{x}_d(\Delta\tau),\Delta\tau) + \boldsymbol{C}_i(\Delta\tau)\boldsymbol{\Delta x}(\Delta\tau), \end{aligned} \tag{A.8}$$

where the row vectors $\boldsymbol{A}_i(\Delta\tau)$ and $\boldsymbol{C}_i(\Delta\tau)$ represent the ith rows of $\boldsymbol{A}(\Delta\tau)$ and $\boldsymbol{C}(\Delta\tau)$ in (A.3), respectively, and where $\boldsymbol{\Delta x}(\Delta\tau) = (\Delta x_1(\Delta\tau),\ldots,\Delta x_n(\Delta\tau))^T$.

The state variable i of the stochastic model at the next time-step, $x_{s,i}(2\Delta\tau)$, can then again be expressed as a perturbation $\Delta x_i(2\Delta\tau)$ of the same state variable of the deterministic model:

$$\begin{aligned} x_{s,i}(2\Delta\tau) &= x_{s,i}(\Delta\tau) + Po[m_1(\Delta\tau)] - Po[m_2(\Delta\tau)] \\ &\approx x_{d,i}(\Delta\tau) + \Delta x_i(\Delta\tau) + m_1(\Delta\tau) - m_2(\Delta\tau) \\ &\quad + m_1(\Delta\tau)^{1/2} e_1(\Delta\tau) - m_2(\Delta\tau)^{1/2} e_2(\Delta\tau) \\ &\approx x_{d,i}(\Delta\tau) + \Delta x_i(\Delta\tau) + \Delta\tau[g_i(\boldsymbol{x}_d(\Delta\tau),\Delta\tau) + \boldsymbol{A}_i(\Delta\tau)\boldsymbol{\Delta x}(\Delta\tau)] \\ &\quad - \Delta\tau[h_i(\boldsymbol{x}_d(\Delta\tau),\Delta\tau) + \boldsymbol{C}_i(\Delta\tau)\boldsymbol{\Delta x}(\Delta\tau)] + w_i(\Delta\tau) \\ &= x_{d,i}(\Delta\tau) + \Delta t \cdot g_i(\boldsymbol{x}_d(\Delta\tau),\Delta\tau) - \Delta\tau \cdot h_i(\boldsymbol{x}_d(\Delta\tau),\Delta\tau) \\ &\quad + \Delta x_i(\Delta\tau) + \Delta\tau \cdot \boldsymbol{A}_i(\Delta\tau)\boldsymbol{\Delta x}(\Delta\tau) - \Delta\tau \cdot \boldsymbol{C}_i(\Delta\tau)\boldsymbol{\Delta x}(\Delta\tau) + w_i(\Delta\tau) \end{aligned} \tag{A.9}$$

$$= x_{d,i}(2\Delta\tau) + \Delta x_i(2\Delta\tau). \tag{A.10}$$

Here, (A.7) and the approximations (A.4) and (A.5) were used in the second line. The linearisation approximations (A.8) were then used, followed by use of (A.1) and (A.3) in the last line. The perturbation term $w_i(\Delta\tau)$ represents some of the terms of the second line in (A.9):

$$w_i(\Delta\tau) = m_1(\Delta\tau)^{1/2} e_1(\Delta\tau) + m_2(\Delta\tau)^{1/2} e_2(\Delta\tau). \tag{A.11}$$

We note by (A.11) that $E[w_i(\Delta\tau)] = 0$, with the average taken w.r.t. $e_1(\Delta\tau)$ and $e_2(\Delta\tau)$.

As in (A.7), the term $\Delta x_i(2\Delta\tau)$ in (A.10) then has zero mean, since it is generated by a zero-mean Gaussian input to a linear dynamic system (A.3). Therefore:

$$\begin{aligned} E[x_{s,i}(2\Delta\tau)] \approx x_{d,i}(2\Delta\tau) + E[\Delta x_i(2\Delta\tau)] = x_{d,i}(2\Delta\tau), \text{ for all } i \\ = 1,2,\ldots,n. \end{aligned} \tag{A.12}$$

This reasoning can be repeated for the whole solution.

Essentially, we have a zero mean (and approximately symmetrically distributed) perturbation vector $\boldsymbol{w}(t) = (w_1(t),\ldots,w_n(t))^T$ that affects the linear perturbation model (A.3) which describes the deviation of the stochastic model trajectory from the deterministic model trajectory. If $E[\boldsymbol{\Delta x}(t)] = 0$ from the previous step, then by (A.3):

$$E[\boldsymbol{\Delta x}(t+\Delta\tau)] = (\boldsymbol{I} + \Delta\tau\boldsymbol{A}(t) - \Delta\tau\boldsymbol{C}(t))E[\boldsymbol{\Delta x}(t)] + E[\boldsymbol{w}(t)] = 0,$$

as long as the perturbation trajectories are in a region of the state space in which the local linearisation remains valid. By induction, $E[\boldsymbol{x}_s(t)] \approx \boldsymbol{x}_d(t)$ then holds for all $t \in [0,T]$.

A.1. Comments

The Condition 2 for asymptotic stability of the linearised equations is required to ensure bounded results of perturbations. Locally unstable dynamics may otherwise amplify small perturbations so they end up outside the range of validity of the linearisation.

Asymptotic stability implies bounded-input-bounded-output stability [37]. In a deterministic setting, this will ensure that the perturbed state trajectory will remain within a domain of (approximate) validity of the linearised model if the relative perturbations are sufficiently small. In our present stochastic setting, asymptotic stability cannot exclude the possibility that the accumulated effect of perturbations will push the state outside of this validity range. The probability for this can, however, be made arbitrarily small by reducing the variance of the perturbations relative to the other right-hand side terms of the difference equations.

However, the time-step must be limited so that the vectors $\boldsymbol{g}(\cdot)$ and $\boldsymbol{h}(\cdot)$ are approximately constant within $\Delta\tau$. In practice, a much smaller simulation time-step, $\Delta t = \Delta\tau/k$, might be used for numerical reasons. This would reduce the flows per time-step Δt by a factor of k, on average. However, in the reasoning above we may collect the accumulated flows from k such smaller time-steps to aggregated flows during a larger step $\Delta\tau = k \cdot \Delta t$ since, in distribution:

$$\begin{aligned} &Po[\Delta t \cdot g_i(\boldsymbol{x}(\Delta t),\Delta t)] + Po[\Delta t \cdot g_i(\boldsymbol{x}(2\Delta t),2\Delta t)] + \cdots + Po[\Delta t \\ &\quad \cdot g_i(\boldsymbol{x}(k\Delta t),k\Delta t)] \\ &\quad = k \cdot Po[\Delta t \cdot g_i(\boldsymbol{x}(k\Delta t),k\Delta t)] = Po[k \cdot \Delta t \cdot g_i(\boldsymbol{x}(k\Delta t),k\Delta t)] \\ &\quad = Po[\Delta\tau \cdot g_i(\boldsymbol{x}(\Delta\tau),\Delta\tau)], \end{aligned}$$

as long as the function $g_i(\cdot)$ remains constant over the time interval $\Delta\tau$. Therefore, $\Delta\tau$ can be interpreted here as the largest dynamically useful time-step, but not necessarily the time-step that would actually be used in a simulation.

Appendix B. Proof of Theorem 1

The proof is performed in two steps, after which the result follows by induction.

Because all the n stochastic difference equations of system (5) have the same form, we only study the ith equation and drop the index i. Furthermore, in Step I of the proof we use the simplified notation $\Sigma_j g_j$ for $\Sigma_j[a_j \cdot x_j + b]$ and $\Sigma_j h_j$ for $\Sigma_j[c_j \cdot x_j + d]$. Eq. (5) is then expressed as:

$$x_s(t+\Delta t) = x_s(t) + \Sigma_j Po[\Delta t \cdot g_j(\boldsymbol{x}_s,t)] - \Sigma_j Po[\Delta t \cdot h_j(\boldsymbol{x}_s,t)]. \tag{B.1}$$

Step I. First study how the expected value, $E[x_s(t+\Delta t)]$ by (B.1) will differ from $x_d(t+\Delta t)$, obtained from the embedded deterministic equation:

$$x_d(t+\Delta t) = x_d(t) + \Sigma_j \Delta t \cdot g_j(\boldsymbol{x}_d,t) - \Sigma_j \Delta t \cdot h_j(\boldsymbol{x}_d,t); \tag{B.2}$$

when both equations are initiated at time t, where $\boldsymbol{x}_s(t) = \boldsymbol{x}_d(t) \geqslant \boldsymbol{0}$ (i.e. all state variables are non-negative).

Since a population model must start with non-negative numbers of entities for all components of $\boldsymbol{x}_s$ and $\boldsymbol{x}_d$, and since a_j, b, c_j and $d \geqslant 0$, the linear functions g_j and h_j are at that point in time also non-negative.

Then, from (B.1),

$$\begin{aligned} E[x_s(t+\Delta t)] &= E[x_s(t)] + \Sigma_j E[Po[\Delta t \cdot g_j(\boldsymbol{x}_s,t)]] - \Sigma_j E[Po[\Delta t \\ &\quad \cdot h_j(\boldsymbol{x}_s,t)]] \\ &= x_s(t) + \Sigma_j \Delta t \cdot g_j(\boldsymbol{x}_s,t) - \Sigma_j \Delta t \cdot h_j(\boldsymbol{x}_s,t) \\ &= x_d(t) + \Sigma_j \Delta t \cdot g_j(\boldsymbol{x}_d,t) - \Sigma_j \Delta t \cdot h_j(\boldsymbol{x}_d,t) \\ &= x_d(t+\Delta t), \end{aligned} \tag{B.3}$$

Table C.1
List of the examples in this paper demonstrating the various reasons for bias.

Example	Comments	Deterministic problems
1. Queuing system M/M/1	Demographic stochasticity	False "synchronisation"
2. Lanchester's model of warfare	Demographic stochasticity and linear model	Cannot reflect stochastic battle times, which in the stochastic case requires a time-indeterminate stopping criterion
3. Prey-predator model (Lotka–Volterra model with interspecies competition)	Demographic stochasticity	Cannot follow stochastic switching to other areas of the phase space – e.g. extinction
4. SIR model	Demographic stochasticity and large total population	Can give huge errors for R_0 close to 1 and small $I(0)$ because of excluding possible early extinction of the disease in the population
5. Lotka–Volterra model	Demographic stochasticity and large numbers	The "phase" in the oscillation cycle may be e.g. a peak in the deterministic case at final simulation time
6. Time-independent failure rate	Demographic stochasticity and linear model	Indefinite stopping criteria may create problems even for a linear model
7. SIR model with random initial values	Initial value stochasticity	Averaging stochastic initial values might create errors
8. Exponential growth with a stochastic parameter	Environmental stochasticity	Errors created by removing environmental stochasticities
Chapter 9. Non-linear output functions	Non-linear output function	Also linear output functions are required to avoid bias.

where $E[Po[m]] = m$ was used in the second equality and the initialisation $\boldsymbol{x}_s(t) = \boldsymbol{x}_d(t)$ was used in the third equality.

Step II. The effect of the first time-step was that the stochastic solution, $x_s(t+\Delta t)$, was split up into a set of possible outcomes $k = 0, 1, 2, \ldots$ with a spectrum (pdf) of probabilities $p(0), p(1), p(2), \ldots$; where $\Sigma_k p(k) = 1$. This pdf represents the set of starting points for the next time-step of the stochastic solution.

For the second time-step it is first demonstrated that since g_j and h_j are linear and non-negative, the expected result of the stochastic solution $E[x_s(t+2\Delta t)]$ will be the same if the calculation is started from $E[x_s(t+\Delta t)]$ or if it is started from the dispersed spectrum (pdf) of solutions at $t+\Delta t$ (where the probability for each outcome is taken into account).

Thus, starting from $E[x_s(t+\Delta t)]$ and calculating one time-step forward for linear and non-negative flows $g_j = \Sigma_j \cdot a_j \cdot x_{s,j} + b$ and $h_j = \Sigma_j\ \cdot c_j \cdot x_{s,j} + d$ gives:

$$\begin{aligned} x_s(t+2\Delta t) = {} & E[x_s(t+\Delta t)] + \Sigma_j Po[\Delta t \cdot a_j \cdot E[x_{s,j}(t+\Delta t)]] + Po[\Delta t \cdot b] \\ & - \Sigma_j Po[\Delta t \cdot c_j \cdot E[x_{s,j}(t+\Delta t)]] - Po[\Delta\ t \cdot d], \end{aligned}$$

where a_j, c_j, and b and $d \geqslant 0$. Then we have:

$$\begin{aligned} E[x_s(t+2\Delta t)] = {} & E\{E[x_s(t+\Delta t)] + \Sigma_j Po[\Delta t \cdot a_j \cdot E[x_{s,j}(t+\Delta t)]] \\ & + Po[\Delta t \cdot b] - \Sigma_j Po[\Delta t \cdot c_j \cdot E[x_{s,j}(t+\Delta t)]] \\ & - Po[\Delta t \cdot d]\} \end{aligned} \tag{B.4}$$

Utilising that the expected value $E[x_s(t+\Delta t)]$ equals $\Sigma_k\ k \cdot p(k)$ and $\Sigma_k p(k) = 1$ gives:

$$\begin{aligned} E[x_s(t+2\Delta t)] = {} & E\{\Sigma_k k \cdot p(k) + \Sigma_j Po[\Delta t \cdot a_j \cdot \Sigma_k k \cdot p(k)] \\ & + Po[\Delta t \cdot \Sigma_k p(k) \cdot b] - \Sigma_j Po[\Delta t \cdot c_j \cdot \Sigma_k k \cdot p(k)] \\ & - Po[\Delta t \cdot \Sigma_k p(k) \cdot d]\} = E\{0 \cdot p(0) \\ & + \Sigma_j Po[\Delta t \cdot a_j \cdot 0 \cdot p(0)] + Po[\Delta t \cdot p(0) \cdot b] \\ & - \Sigma_j Po[\Delta t \cdot c_j \cdot 0 \cdot p(0)] - Po[\Delta t \cdot p(0) \cdot d] + 1 \cdot p(1) \\ & + \Sigma_j Po[\Delta t \cdot a_j \cdot 1 \cdot p(1)] + Po[\Delta t \cdot p(1) \cdot b] \\ & - \Sigma_j Po[\Delta t \cdot c_j \cdot 1 \cdot p(1)] - Po[\Delta t \cdot p(1) \cdot d] + 2 \cdot p(2) \\ & + \Sigma_j Po[\Delta t \cdot a_j \cdot 2 \cdot p(2)] + Po[\Delta t \cdot p(2) \cdot b] \\ & - \Sigma_j Po[\Delta t \cdot c_j \cdot 2 \cdot p(2)] - Po[\Delta t \cdot p(2) \cdot d] + \cdots\}. \end{aligned}$$

However, $E[Po[n \cdot x]] = E[Po[x] + \cdots + Po[x]] = E[Po[x]] + \cdots + E[Po[x]] = n \cdot E[Po[x]]$. Thus:

$$\begin{aligned} & E[x_s(t+2\Delta t)] \\ & = p(0) \cdot \{0 + \Sigma_j Po[\Delta t \cdot a_j \cdot 0] + Po[\Delta t \cdot b] - \Sigma_j Po[\Delta t \cdot c_j \cdot 0] - Po[\Delta t \cdot d]\} \\ & \quad + p(1) \cdot \{1 + \Sigma_j Po[\Delta t \cdot a_j \cdot 1] + Po[\Delta t \cdot b] - \Sigma_j Po[\Delta t \cdot c_j \cdot 1] - Po[\Delta t \cdot d]\} \\ & \quad + p(2) \cdot \{2 + \Sigma_j Po[\Delta t \cdot a_j \cdot 2] + Po[\Delta t \cdot b] - \Sigma_j Po[\Delta t \cdot c_j \cdot 2] - Po[\Delta t \cdot d]\} \\ & \quad + \cdots \\ & = p(0) \cdot x_s(t+2\Delta t) \text{ starting from } x_s(t+\Delta t) = 0 \\ & \quad + p(1) \cdot x_s(t+2\Delta t) \text{ starting from } x_s(t+\Delta t) = 1 \\ & \quad + p(2) \cdot x_s(t+2\Delta t) \text{ starting from } x_s(t+\Delta t) = 2 \\ & \quad + \cdots \end{aligned} \tag{B.5}$$

Thus, it is shown that for the stochastic equation there is no difference with respect to the expected value, $E[x_s(t+2\Delta t)]$, if the second time-step starts from the expected value, $E[x_s(t+\Delta t)]$ (by Eq. (B.4)), or if it starts from the split-up spectrum of values obtained from the first time-step and weighted by the pdf (by Eq. (B.5)). Thus, the splitting up will not affect the unbiasedness if g_j and h_j are linear in $\boldsymbol{x}$ and non-negative.

Furthermore, if the stochastic and deterministic equations start from the same value at $t+\Delta t$ then the results at $t+2\Delta t$, $E[x_s(t+2\Delta t)]$ and $x_d(t+2\Delta t)$, will be equal (by (B.3)).

Thus, $E[x_s(t+2\Delta t)]$ starting from the pdf generated in the previous time-step equals $E[x_s(t+2\Delta t)]$ starting from $E[x_s(t+\Delta t)]$ which also equals $x_d(t+2\Delta t)$ starting from $x_d(t+\Delta t)$.

These arguments can be continued by induction over the whole time interval T.

Appendix C. The examples in this paper

In Table C.1 the examples in this paper, demonstrating the various reasons for bias, are listed.

References

[1] J.W. Haefner, Modelling Biological Systems: Principles and Applications, Chapman & Hall, International Thomson Publishing, New York, 1996.
[2] R. Lande, S. Engen, B-E. Sæther, Stochastic Population Dynamics in Ecology and Conservation, Oxford University Press, U.K., 2003.
[3] L. Gustafsson, M. Sternad, Consistent micro, macro and state-based population modelling, Math. Biosci. 225 (2010) 94.
[4] R.M. Anderson (Ed.), Population Dynamics of Infectious Diseases, Theory and Applications, Chapman and Hall, London, 1982.
[5] B. Hannon, M. Ruth, Modeling Dynamic Biological Systems, Springer Verlag, New York, 1997.
[6] F. Brauer, C. Castillo-Chávez, Mathematical Models in Population Biology and Epidemiology, Springer-Verlag, New York, 2001.
[7] M. Braun, Differential Equations and Their Applications, Springer-Verlag, New York, 1993.

[8] D.G. Luenberger, Introduction to Dynamic Systems, Theory, Models and Applications, John Wiley & Sons Inc., New York, 1979.
[9] A.G. McKendrick, Applications of Mathematics to Medical Problems, Proc. Edin. Math. Soc. 14 (1926) 98.
[10] M.S. Bartlet, Stochastic Population Models in Ecology and Epidemiology, Wiley, New York, 1960.
[11] N.T.J. Bailey, The Mathematical Theory of Infectious Diseases and its Applications, Griffin, London, 1975.
[12] R.M. Nisbet, W.C.S. Gurney, Modelling Fluctuating Populations, The Blackburn Press, NJ, 1982.
[13] E. Renshaw, Modelling Biological Populations in Space and Time, Cambridge University Press, Cambridge, 1991.
[14] M.J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton University Press, NJ, 2008.
[15] E. Vynnycky, R.G. White, An Introduction to Infectious Disease Modelling, University Press, Oxford, 2010.
[16] D. Claessen, J. Andersson, L. Persson, A.M. de Roos, Delayed evolution branching in small populations, Evol. Ecol. Res. 9 (2007) 51.
[17] L. Fahse, C. Wissel, V. Grimm, Reconciling classical and individual-based approaches in theoretical population ecology: A protocol for extracting population parameters from individual-based models, Am. Nat. 152 (1988) 838.
[18] P. Gómez-Mourelo, From individual-based models to partial differential equations. An application to the upstream movement of elvers, Ecol. Model. 188 (2005) 93.
[19] V. Grimm, Ten years of individual-based modelling in ecology: What have we learned and what can we learn in the future?, Ecol Model. 115 (1999) 129.
[20] W.G. Wilson, Resolving discrepancies between deterministic population models and individual-based simulations, Am Nat. 151 (1998) 116.
[21] L.J.S. Allen, A.M. Burgin, Comparison of deterministic and stochastic SIS and SIR models in discrete time, Math. Biosci. 163 (2000) 1.
[22] M.J. Keeling, J.V. Ross, On methods for studying stochastic disease dynamics, J. R. Soc. Interface 5 (2008) 171.
[23] L. Gustafsson, M. Sternad, Bringing consistency to simulation of population models – Poisson Simulation as a bridge between micro and macro simulation, Math. Biosci. 209 (2007) 361.
[24] L. Gustafsson, Poisson Simulation – A Method for Generating Stochastic Variations in Continuous System Simulation, Simulation 74 (5) (2000) 264.
[25] D.T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, J. Chem. Phys. 115 (2001) 1716.
[26] L. Kleinrock, Queueing Systems, Theory, Vol. 1, John Wiley & Sons, New York, 1975.
[27] L. Gustafsson, Poisson Simulation as an Extension of Continuous System Simulation for the Modeling of Queuing Systems, Simulation 79 (9) (2003) 528.
[28] F.W. Lanchester, Aircraft in Warfare, the Dawn of the Fourth Arm, Tiptree, Constable and Co., Ltd., London, 1916.
[29] V. Volterra, Fluctuations in the abundance of a species considered mathematically, Nature 118 (1926) 558.
[30] L. Gustafsson, Studying Dynamic and Stochastic Systems Using Poisson Simulation, in: H. Liljenström, U. Svedin (Eds.), Micro–Meso–Macro: Addressing Complex Systems Couplings, World Scientific Publishing Company, Singapore, 2005, p. 131.
[31] W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics, Proc. Royal Soc. A 115 (1927) 700.
[32] P.J. Burke, The Output of a Queueing System, Oper. Res. 4 (1956) 699.
[33] J.R. Jackson, Networks of Waiting Lines, Oper. Res. 5 (1957) 518.
[34] L. Ljung, Analysis of recursive stochastic algorithms, IEEE Transactions on Automatic Control AC-22 (1977) 551.
[35] L. Ljung, T. Söderström, Theory and Practice of Recursive Identification, MIT Press, Cambridge, MA, 1983.
[36] H. Kushner, Approximation and Weak Convergence Methods of Random Processes, MIT Press, Cambridge, MA, 1984.
[37] M. Vidyasagar, Nonlinear Systems Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1978.