Dynamic noise, chaos and parameter estimation in population biology

N. Stollenwerk*, M. Aguiar, S. Ballesteros, J. Boto, B. Kooi and L. Mateus

Centro de Matemática e Aplicações Fundamentais, Universidade de Lisboa, Avenida Prof. Gama Pinto 2, 1649-003 Lisbon, Portugal

We revisit the parameter estimation framework for population biological dynamical systems, and apply it to calibrate various models in epidemiology with empirical time series, namely influenza and dengue fever. When it comes to more complex models such as multi-strain dynamics to describe the virus–host interaction in dengue fever, even the most recently developed parameter estimation techniques, such as maximum likelihood iterated filtering, reach their computational limits. However, the first results of parameter estimation with data on dengue fever from Thailand indicate a subtle interplay between stochasticity and the deterministic skeleton. The deterministic system on its own already displays complex dynamics up to deterministic chaos and coexistence of multiple attractors.

Keywords: epidemiology; influenza; dengue fever; deterministic chaos; master equation; likelihood

1. INTRODUCTION

A major contribution to stochasticity in empirical epidemiological data is population noise, which is modelled by time-continuous Markov processes or master equations [1–3]. In some cases, the master equation can be analytically solved and from the solution a likelihood function be given [4]. The likelihood function gives best estimates via maximization or can be used in the Bayesian framework to calculate the posterior distribution of parameters [3]. Here, we start with an example of a linear infection model that can be solved analytically in all aspects and then generalize to more complex epidemiological models that are relevant for the description of influenza or dengue fever [5], on the cost of having to perform more and more steps by simulation to obtain the likelihood function by complete enumeration [6] or even in extreme cases just to search for the maximum.

Recent applications to a multi-strain model applied to empirical datasets of dengue fever in Thailand, where the model displays such complex dynamics as deterministic chaos in wide parameter regions [5,7], including coexistence of multiple attractors, e.g. in isolas [8], stretch the presently available methods of parameter estimation well to its limits. Finally, the analysis of scaling of solely population noise indicates that very large world regions have to be considered in data analysis in order to be able to compare the fluctuations of the stochastic system with the much easier to analyse deterministic skeleton. Such a deterministic skeleton can already show deterministic chaos [9,10], here via torus bifurcations [8], which are also found in other ecological models, such as the seasonally forced Rosenzweig–McArthur model [11].

The main aim of this contribution is to give an overview and didactic introduction to parameter estimation in population biology from simple analytically solvable cases to recently developed numerical methods applicable to larger models with such complex dynamical behaviour as deterministic chaos in its mean field deterministic skeleton. Although this is a review some new results will be presented here as well, mainly to include dynamic noise in the likelihoods of the numerical simulations of complex models in parallel with the first given analytical results in the simple models.

2. AN ANALYTICALLY SOLVABLE CASE

We first review some of the simplest epidemiological models, and show explicitly the relation between the formulation as a stochastic process and the deterministic models formulated as ordinary differential equation (ODE) systems. Then we will solve explicitly one example where all steps towards a closed solution can be given analytically. In this example model, a likelihood function can be derived, including an analytical expression of, as well as the maximum likelihood best estimate as, a full Bayesian formulation.

*Author for correspondence (nico@ptmat.fc.ul.pt).

One contribution of 11 to a Theme Issue ‘Mathematical and theoretical ecology’.

Received 19 October 2011
Accepted 5 January 2012

Published online 1 February 2012
The model can be generalized to give a fast approximation scheme for more complicated stochastic epidemic models which later will be used in the applications with likelihoods based on simulations. With such simulation models, all steps, which we show in the analytically solvable case, can be performed in more complicated models and applied to time series of real disease cases.

### 2.1. Basic epidemiological models including population noise

The simplest closed epidemiological model is the susceptible–infected–susceptible (SIS) model for susceptible and infected hosts, specified by the following reaction scheme:

\[
S + I \xrightarrow{\beta} I + I \bigg\} \quad (2.1)
\]

and

\[
I \xrightarrow{\alpha} S
\]

for stochastic variables \( I \) and \( S = N - I \) with \( N \) the total population size of hosts. The transitions are infection rate \( \beta \) and recovery rate \( \alpha \). From the reaction scheme, we can give the dynamics of the probability \( p(I, t) \) as follows:

\[
\frac{d}{dt} p(I, t) = \frac{\beta}{N} (I(1-I)N-I) p(I-1, t) + \alpha (I+1) p(I+1, t) - \frac{\beta}{N} I(N-I) + \alpha I p(I, t). \quad (2.2)
\]

With the definition of mean values, which is given here by \( \langle I \rangle := \sum_{I=0}^{N} I \times p(I, t) \), we can derive the dynamics of the mean number of infected in time by inserting the dynamics of the probability given above and obtaining after some calculation (for a more extended derivation and further details, see Stollenwerk & Jansen [3] and references therein)

\[
\frac{d}{dt} \langle I \rangle = (\beta - \alpha \langle I \rangle - \frac{\beta}{N} \langle I^2 \rangle, \quad (2.3)
\]

where higher moment \( \langle I^2 \rangle \) now appears in the dynamics. For dynamics of higher moments and in spatial systems, see the studies by Stollenwerk & Jansen [3] and Stollenwerk et al. [12] for the equivalent higher clusters (pairs, triples, etc.). Since, in spatial systems, the epidemiological thresholds are, in general, shifted away from the mean field values, further methods valid at and around these thresholds can be used [13].

To obtain a closed ODE, we neglect the variance, restricting the dynamics to its simplest deterministic part, the so-called mean field approximation, \( \text{var} := \langle I^2 \rangle - \langle I \rangle^2 \approx 0 \), and find the famous logistic equation (in a slightly unusual form but mathematically equivalent to the well-known formulation of Verhulst)

\[
\frac{d}{dt} \langle I \rangle = \frac{\beta}{N} \langle I \rangle (N - \langle I \rangle) - \alpha \langle I \rangle. \quad (2.4)
\]

In complete analogy, many other epidemiological models can be formulated, often with somewhat more effort, such as the susceptible–infected–recovered (SIR) epidemic for susceptible, infected and recovered hosts

\[
S + I \xrightarrow{\beta} I + I
\]

\[
I \xrightarrow{\gamma} R
\]

\[
R \xrightarrow{\alpha} S
\]

with the respective transition rates, infection rate \( \beta \), recovery rate \( \gamma \) and waning immunity \( \alpha \). For variables \( S \) and \( I \) with \( R = N - S - I \) given from the previous two classes, we can again give the dynamics of the probability \( p(S, I, t) \) by

\[
\frac{d}{dt} p(S, I, t) = \frac{\beta}{N} (I(1-I)(S+1)) p(S+1, I-1, t) + \gamma (I+1) p(S+1, I+1, t) - \alpha (N-(S+1)-I) p(S+1, I, t) - \left( \frac{\beta}{N} S + \gamma + \alpha (N-S-I) \right) p(S, I, t). \quad (2.6)
\]

And, again, the mean field approximation of now higher cross moments such as \( \langle SI \rangle \) gives a well-known closed ODE system for this SIR system.

A seasonally forced SIR system can already be used to analyse diseases such as influenza with available long time series from various countries [14]. As we will show later, in realistic parameter regions for influenza, we find complex dynamic behaviours with bifurcations from limit cycles to double limit cycles and up to deterministically chaotic dynamics. Also, multi-strain epidemic models can be formulated in the frame work of stochastic processes given above by including host classes of infected with one or another strain.

One example is a dengue model, which can capture differences between primary and secondary infection, an important feature owing to the so-called antibody-dependent enhancement (ADE). Infected with one strain are labelled \( I \) and infected with a second different strain \( I_2 \); they can recover from that strain, called \( R_1 \) and \( R_2 \), respectively, and after a period of temporary cross immunity become susceptible to infection with another strain. Hence, a susceptible that has been infected with strain 1 previously is labelled \( S_1 \), and can now only be infected with strain 2, therefore becomes \( I_{12} \), etc. For details, see Aguiar et al. [5] and references therein.

The reaction scheme then looks for an infection with one strain and a subsequent infection with a second strain, such as

\[
S + I_1 \xrightarrow{\beta} I_1 + I_1
\]

\[
S + I_{21} \xrightarrow{\beta} I_1 + I_{21}
\]

\[
I_1 \xrightarrow{\gamma} R_1
\]

\[
R_1 \xrightarrow{\alpha} S_1
\]

\[
S_1 + I_2 \xrightarrow{\beta} I_{12} + I_2
\]

\[
S_1 + I_{12} \xrightarrow{\beta} I_{12} + I_{12}
\]

\[
I_{12} \xrightarrow{\gamma} R,
\]

\[
\text{and}
\]

\[
S + I_{12} \xrightarrow{\beta} I_{12} + I_{12}
\]

\[
I_{12} \xrightarrow{\gamma} R,
\]

\[
S + I_{21} \xrightarrow{\beta} I_{21} + I_{21}
\]

\[
S + I_{21} \xrightarrow{\beta} I_{21} + I_{21}
\]

\[
I_{21} \xrightarrow{\gamma} R_{21}
\]

\[
R_{21} \xrightarrow{\alpha} S_{21}
\]

\[
S_1 + I_{21} \xrightarrow{\beta} I_{12} + I_{21}
\]

\[
S_1 + I_{21} \xrightarrow{\beta} I_{12} + I_{21}
\]

\[
I_{12} \xrightarrow{\gamma} R_{12}
\]

\[
I_{12} \xrightarrow{\gamma} R_{12}
\]
leading to a 10-dimensional ODE system which can be reduced to 9 d.f. for constant population size. Already in its simplest form, the model shows complex dynamics in wide parameter regions, up to deterministic chaos characterized by positive Lyapunov exponents [5]. A version of this model including seasonality and import can easily explain the fluctuations observed in time series of severe disease (dengue haemorrhagic fever; DHF) in Thailand [7]. We will come back later to these models. First, we will analyse a simpler model.

2.2. Analytical solution of the linear infection model

To simplify further for analytical tractability, we assume only infection and only from an outside population. This simplified epidemiological model, an SI system, where infection is acquired only from the outside, leads to a master equation which is linear not only in probability but also in the state variables. A linear mean field approximation for the dynamics of the expectation values [3] is obtained from this process. Although very simple in its set-up, it can be applied to real world data of influenza in certain stages of the underlying SIR model, when considering the cumulative number of infected cases during the outbreak [4]. The reaction scheme is given by

\[ S + I^* \xrightarrow{\beta} I + I^* \]  

for infected I and susceptibles \( S = N - I \) with population size \( N \), and infection rate \( \beta \) as the only possible transition. The underlying model hypothesis is that infection can be acquired from outside the considered population of size \( N \), hence meeting infected \( I^* \).

The master equation reads for the probability \( p(I, t) \)

\[
\frac{d}{dt} p(I, t) = \frac{\beta}{N} I^*(N - (I - 1)) p(I - 1, t) \\
- \frac{\beta}{N} I^*(N - I) p(I, t),
\]

which can be solved using the characteristic function

\[
\langle e^{i\kappa} \rangle := \sum_{I=0}^{N} e^{i\kappa I} \times p(I, t) =: g(\kappa, t)
\]

or approximated by Kramers–Moyal expansion to preserve eventual attractor switching in more complicated models, relevant for influenza [15].

2.3. Solving the master equation

For suitable initial conditions, the master equation can be solved to obtain the transition probabilities needed to construct the likelihood function. From the definition of the characteristic function, equation (2.10), we see immediately that its derivatives are related to the moments \( \langle I^n \rangle \)

\[
(-i)^n \frac{\partial^n}{\partial \kappa^n} g(\kappa, t) \bigg|_{\kappa=0} = \langle I^n \rangle
\]

and \( g \) itself is a discrete Fourier transform using the definition \( \kappa =: (2\pi/(N + 1)) \times k \), hence

\[
g(\kappa, t) = \sum_{I=0}^{N} e^{i(2\pi/(N+1))k \times I} \times p(I, t) = \hat{g}(k, t).
\]

Owing to this, we can, via Fourier back transform, obtain the probabilities once we know the characteristic function

\[
p(I, t) \equiv \frac{1}{N+1} \sum_{k=0}^{N} e^{-i(2\pi/(N+1))k \times I} \times \hat{g}(k, t).
\]

Hence, we can derive a dynamics for the characteristic function from the dynamics of the probabilities, eventually solve the partial differential equation (PDE) for the characteristic function and then obtain, via Fourier back transform, the solution to the probability.

The dynamics of the characteristic function, obtained from the master equation, is

\[
\frac{\partial}{\partial t} g(\kappa, t) = \sum_{I=0}^{N} e^{i\kappa I} \times \frac{d}{dt} p(I, t),
\]

giving the PDE

\[
\frac{\partial}{\partial t} g(\kappa, t) = \beta^* N (e^{i\kappa} - 1) \times g(\kappa, t)
\]

\[
+ i\beta^* (e^{i\kappa} - 1) \times \frac{\partial g}{\partial \kappa}
\]

using \( \beta^* := (\beta/N) I^* \), and the initial condition \( p(I, t_0) = \delta_{I, I_0} \), and hence \( g(\kappa, t_0) = e^{i\kappa I_0} \) as the initial condition for the characteristic function. With a classical separation ansatz and the given initial conditions, we can solve the PDE analytically, obtaining as the solution (figure 1a)

\[
g(\kappa, t) = e^{i\kappa N} (e^{-i\kappa} e^{-\beta^*(t-t_0)} + (1 - e^{-\beta^*(t-t_0)})^{N-I_0}).
\]

Then an inverse Fourier transformation gives the transition probability \( p(I, t|I_0, t_0) \), explicitly

\[
p(I, t) = \left( \frac{N-I_0}{I-I_0} \right) (e^{-\beta^*(t-t_0)})^{N-I} (1 - e^{-\beta^*(t-t_0)})^{I-I_0},
\]

as the solution to the master equation. Owing to the special initial conditions of exactly having \( I_0 \) infected at time \( t_0 \), this solution is at the same time the transition probability \( p(I, t|I_0, t_0) \), as needed to construct the likelihood function for the parameter estimation.

2.4. Likelihood function from the master equation

From the transition probabilities, we can construct the likelihood function, i.e. the joint probability of finding all data points \( I_0, I_1, \ldots, I_n \) from our empirical time series interpreted as a function of the model parameters

\[
p(I_n, t_n, \ldots, I_0, t_0) = \prod_{i=0}^{n-1} p(I_{v+1}, t_{v+1}|I_v, t_v) \times p(I_0, t_0)
\]

\[
=: L(\beta).
\]
With the analytical solution, we obtain here explicitly using $\Delta t := t_{i+1} - t_i$

$$L(\beta) = \prod_{v=0}^{n-1} \left( \frac{N - I_v}{I_{v+1} - I_v} \right) (e^{-\beta \Delta t})^{N - I_v - 1} (1 - e^{-\beta \Delta t})^{I_v - 1},$$

and can plot the likelihood as a function of the parameter, here $\beta$ as shown in figure 1b, and its maximum as the best estimator $\hat{\beta}$. Here, the best estimator can be calculated analytically as well [3] by evaluating the first derivative of the likelihood to be zero at $\hat{\beta}$. We obtain analytically

$$\hat{\beta} = \frac{1}{\Delta t} \times \ln \left( \frac{N - (1/n) \sum_{v=0}^{n-1} I_v}{N - (1/n) \sum_{v=0}^{n-1} I_{v+1}} \right)$$

(2.16)
as the best estimate for the parameter $\beta$.

### 2.5. Fisher information

To obtain an approximation of the confidence intervals in the frequentist’s approach, after maximizing the likelihood, one assumes that the likelihood essentially reflects a Gaussian distribution, an assumption justified by the central limit theorem of probability in the case of many data points available for the likelihood. Hence

$$p(\beta) := \frac{1}{\sigma \sqrt{2\pi}} e^{-(\beta - \hat{\beta})^2 / 2\sigma^2}$$

(2.17)
is a Gaussian distribution around the best estimate $\hat{\beta}$ with a standard deviation $\sigma$ to be obtained from the curvature of the likelihood around its maximum

$$\frac{\partial^2 p(\beta)}{\partial \beta^2} \bigg|_{\beta = \hat{\beta}} = -\frac{1}{\sigma} \quad \text{and} \quad \sigma = \frac{1}{\Delta t} L(\beta) / \partial \beta^2 \bigg|_{\beta = \hat{\beta}}.$$

We can test this assumption of Gaussianity and its performance by the following experiment: we simulate the stochastic process, i.e. the master equation via the Gillespie algorithms [16,17], many times and each time calculate the best estimate, obtaining a histogram of the estimates around the initial model parameter. Then we take one typical realization from the bulk of the distribution and calculate the Gaussian distribution around the best estimate with variance given by the negative inverse of the curvature of the likelihood, the Fisher information, and compare the Gaussian with the histogram of many realizations (figure 2). The comparison is relatively good, but the histogram and previously the likelihood function are slightly asymmetric around the maximum, requiring a refined approach to better capture the distribution of parameters given the experimental or empirical data. This can be achieved by using the Bayesian framework, as will be shown below.

### 2.6. Bayesian framework

The Bayesian framework starts by using the likelihood function interpreted as the probability of finding the data given a parameter (or parameter set in more general cases); hence $L(\beta) = p(\mathbf{I} | \beta)$ with $\mathbf{I} = (I_0, I_1, \ldots, I_n)$ being the data vector and parameter $\beta$ and constructs from it the probability of the parameters conditioned on the present data, here from the time series of the epidemic system, the Bayesian posterior $p(\beta | \mathbf{I})$. This can be achieved only by imposing a prior $p(\beta)$, a probability of plausible parameter sets, hence we have the Bayes formula

$$p(\beta | \mathbf{I}) = \frac{p(\mathbf{I} | \beta)}{p(\mathbf{I})} p(\beta)$$

(2.18)

with the likelihood function $p(\mathbf{I} | \beta)$ as before, and a normalization $p(\mathbf{I}) = p(\mathbf{I} | \beta) \cdot p(\beta) d\beta$ once a prior $p(\beta)$ is specified. For analytical calculations, a so-called conjugate prior is often used, i.e. a distribution for the parameter in the same functional form as the parameter appears in the likelihood function, here a beta distribution. Then the posterior also has functionally the same form, here again a beta distribution, but the parameters, called hyperparameters, now contain information from the data points. For the linear infection model, all steps can still be performed analytically.

The posterior is given explicitly by

$$p(\beta | \mathbf{I}) = \frac{\Gamma(\tilde{a} + \tilde{b})}{\Gamma(\tilde{a}) \Gamma(\tilde{b})} \times e^{-\beta \Delta t \cdot \frac{1}{\Delta t} (1 - e^{-\beta \Delta t})^{\tilde{a} - 1} (e^{-\beta \Delta t})^{\tilde{b} - 1}},$$

$$\tilde{a} = \frac{1}{\Delta t} \sum_{v=0}^{n-1} I_v \Delta t$$

$$\tilde{b} = \frac{1}{\Delta t} \sum_{v=0}^{n-1} (N - I_v) \Delta t$$

$$\tilde{a} = \frac{1}{\Delta t} \sum_{v=0}^{n-1} I_v \Delta t$$

$$\tilde{b} = \frac{1}{\Delta t} \sum_{v=0}^{n-1} (N - I_v) \Delta t$$

Figure 1. (a) An example of the characteristic function in real and imaginary parts with fixed $\Delta t$ given by the sampling rate of the time series. (b) From the characteristic function we can obtain the solution to the stochastic model and with it the likelihood function, as shown here with its best estimate of the model parameter as maximum.
with hyperparameters \( \tilde{a} = a + \sum_{n=0}^{a-1} (I_{n+1} - I_n) \) and \( \tilde{b} = b + \sum_{n=0}^{a-1} (N - I_{n+1}) \) from the prior parameters \( a \) and \( b \). For the detailed calculation, see Stollenwerk & Jansen [3]. One can observe that, for large datasets, the sums over the data points are giving much larger values in the hyperparameters than the original prior parameters, hence a soft prior is achieved. If the prior parameters are of the same order of magnitude as the terms originating from the data points, the prior determines much more the form of the posterior than the observations. Such hard priors are not always avoidable when only small datasets are available. Then, the posterior gives some indication of the underlying insecurities of the stochastic process but should also not be overinterpreted.

Figure 3 shows the comparison of a histogram of best estimates from many realizations of the stochastic process (red), information that is not usually given in most empirical systems; often, we only have a single realization, and the results from the parameter estimation, a Gaussian approximation from the best estimate and the Fisher information (green), and finally the Bayesian posterior (black) obtained from a conjugate prior (blue), which is here nicely broad, not imposing much restriction on the parameter values considered.

**2.7. Some further remarks on empirical situations**

Before continuing with more complex population models, we make some remarks on what can happen in empirical situations, which we illustrate here using analytical examples.

A parameter can be close to one of its boundaries and not well measured by the given data points. In figure 4, we show an example from a time series of independent exponentially distributed waiting times \( p(t) = a \times e^{-at} \) (a typical null model for spiking neurons). This model can also be treated analytically in the same way as the linear infection model, which we have calculated above explicitly. The only parameter \( a \) of this model is always positive. However, the approximation of the likelihood by a Gaussian distribution can have a large tail in the negative region of the parameter space (green curve). In this case, clearly, the histogram of our Gedanken experiment (red curve) is highly asymmetric around the maximum, with a long tail shifted to the right, i.e. towards positive best estimates. Also the conjugate prior (blue curve) vanishes at \( a = 0 \), indicating that a negative value of \( a \) is meaningless in this model. The Bayesian posterior (black curve) captures the asymmetric distribution of estimates much better than the Gaussian approximation, on which the Fisher information is based, would suggest. Here, clearly the Bayesian approach is
performing better, while the Gaussian approximation has a large tail for negative parameter \(a\) values.

Again, we illustrate the second remark on the previously studied linear infection model. Previously, we compared the histogram of many stochastic realizations with the likelihood of a ‘typical’ realization, by which we mean a realization close to the bulk of the histogram. However, in any empirical situation our only available dataset, let us say of an epidemiological or ecological time series, is more ‘atypical’ in the sense that it is still quite likely to be obtained under the model assumptions, but not coming from the central bulk. Just by using another seed of our random generator than the one used in figure 3, we obtain the situation shown in figure 5a. And we have to remember that, in empirical situations, we do not even have the histogram of our Gedanken experiment, hence we have the information from figure 5b only. A reasonable confidence interval still covers the true parameter value, i.e. the one we originally used to generate the model realization, but the maximum likelihood best estimate as well as the maximum of the Bayesian prior can be quite far away from that empirically unknown true parameter value.

As a third remark, we already mention here that the binomial transition probability from the linear infection model can be generalized to a numerical approximation scheme for more complex models, the so-called Euler multi-nomial approximation. The assumption of linear transition rates holds for short time intervals and gives a binomial transition rate for staying in the present state or moving into one other or a multi-nomial for leaving to more than one other state. We will use this Euler multi-nomial approximation later for numerical parameter estimation in an SIR model.

3. NUMERICAL LIKELIHOOD VIA STOCHASTIC SIMULATIONS

In cases where not all steps or even no step can be performed analytically, a comparison of stochastic simulations, depending on the model parameters, with the empirical data is the only available information on the parameters. One can compare many simulations and see if they fall within a region around the data, called \(\eta\)-ball, depending on the parameter set used for the simulations [6] obtaining estimates of the likelihood function. In figure 6, an SIR model is compared with empirical data from influenza; here, owing to the relative simplicity of the stochastic model, a numerical enumeration of the whole relevant parameter space is still computationally possible (equivalent to a flat but cut-off prior). In more complicated models, dengue fever models and data from Thailand, a complete enumeration of the parameter space is not computationally possible, although only six parameters and nine initial conditions have to be estimated. Particle filtering, i.e. an often quite restricted distribution of parameters and initial conditions, a hard prior in Bayesian language, is stochastically integrated and compared with the empirical time series, selecting the best performers as the maximum of the likelihood function.

In figure 6, we can also study confidence intervals for multiple parameters. The easiest generalization from the one-parameter case is the likelihood slice, in which around the maximum all but one parameter are fixed and then the confidence interval is evaluated for the remaining parameter, as in the one-parameter case. In figure 6, the maximum of the likelihood is not easily determined and we observe near equally likely parameter sets along a long comb.

To quantify this insecurity along the comb better, one can fix one parameter for which the confidence interval is to be determined, and maximize all the other parameters, which gives essentially a projection of the comb in figure 6. A so-called likelihood profile is obtained in this way. This method gives large confidence intervals for the two parameters chosen here, the infection rate and the initial number of susceptibles before the influenza season, quantifying well the insecurity of parameters
rather than what the eventually much smaller likelihood sections would suggest.

3.1. Application to more complex models

In such cases, in which the model can display deterministic chaos, like the one in dengue fever that we investigated [5,7], even the short-term predictability and long-term unpredictability (as measured by the largest Lyapunov exponent) prohibit the comparison of stochastic simulations with the entire time series. Hence, iteratively short parts of the time series are compared with the stochastic particles, and, via simulated annealing, the variability of the particles is cooled down, i.e. the priors are narrowed in Bayesian language. The final method is called ‘maximum likelihood iterated filtering’ (MIF) [18–21]. We apply this method to dengue data from Thailand (a detailed manuscript is in preparation [22]), obtaining wide likelihood profiles for some of the parameters. Here, we display bifurcation diagrams obtained from the best estimates, around a difficult to estimate but important model parameter, the import rate \( \varrho \). Even in systems with population sizes well above 1 million, the noise level of the stochastic system is enormous.

To illustrate the ideas behind iterated particle filtering and including the explicit effects of dynamic noise as investigated in the previous example analytically, we...
will now look at a susceptible–infected–recovered–susceptible (SIRS) model with seasonality and import, which is motivated by recent studies on influenza [14] and which is reported to show deterministic chaos, and see how a suitable particle filter can perform in such a scenario.

4. A TEST CASE FOR ITERATED FILTERING UNDER CHAOTIC DYNAMICS

Looking back to the roots of filtering in time-series analysis, the earliest complete filters are the Stratonovich–Kalman–Bucy filters [24–29], also known as Kalman filters, for linear stochastic systems under additional observation noise.

4.1. A brief look at the history of filtering

A state-space model \( \frac{d}{dt}x = Ax + Be \) with state vector \( x \) and noise vector \( e \) and in the simplest case a linear scalar as observation \( y = Cx + \xi \) with noise \( \xi \) can be solved explicitly to find \( p(x|y) \), i.e. from the measurement \( y \) the underlying states \( x \) can be obtained, a task which often occurs in signal analysis in physics. But already including the estimation of parameters, here the matrix \( A \), leads to a nonlinear problem that cannot be solved analytically, but has to be approximated, via extended Kalman filters; hence, the calculation of \( p(x|A|y) \) or \( p(A|y) \) is already non-trivial.

Similarly for nonlinear systems \( (d/dt)x = f(x, \theta, e) \), where the parameter vector \( \theta \) now replaces the coefficients of the matrices \( A \) and \( B \), and observation \( y = g(x, \xi) \) has often been treated with simulation procedures to solve the classical filtering problem, i.e. to estimate the states \( x \) from the observations \( y \), but turns out to be much more difficult, in computational terms, again for parameter estimation, i.e. to estimate \( \theta \) from \( y \), although it is tempting to just consider parameters as additional states and leave the conceptual framework unchanged. But then, as in MIF, the likelihood function only considers the observation noise, not the dynamic noise. In mathematical terms, the likelihood function \( p(y|x) \) is taken from the observation process \( y = g(x, \xi) \) and includes only indirectly the dynamic noise \( \frac{d}{dt}x = f(x, \theta, e) \) giving zero weight to dynamic realizations below the observations. For extended reviews on filtering, see [30–33].

We include dynamic noise explicitly in the iterated filtering with resampling and obtain good results for difficult-to-estimate parameters such as the import rate under chaotic dynamics.

4.2. Seasonal SIR model with chaotic dynamics

The SIR model with import \( e \)

\[
\begin{align*}
S + I & \xrightarrow{\beta(t)} I + I \\
I & \xrightarrow{\gamma} R \\
R & \xrightarrow{\alpha} S \\
S & \xrightarrow{\mu} I,
\end{align*}
\]

and seasonal forcing given by

\[
\beta(t) = \beta(1 + \theta \times \cos(\omega t)),
\]

and parameters in the uniform phase and chaotic amplitude (UPCA) region, relevant for influenza [14], \( \alpha = 1/6y, \gamma = 1/3d = 365/3y, \beta = 1.5 \times \gamma, \theta = 0.12 \), shows around the import \( \ln(g) = -15 \) deterministic chaos. We take stochastic simulations from this model to test time series for the illustration of iterated filtering with dynamic noise in the likelihood function estimated via the Euler multi-nominal approximation, again a stochastic simulation method comparing the data by using an \( \eta \)-distance measure; see figure 9 for the comparison of the example dataset and the approximating simulation. They diverge because of the deterministically chaotic dynamics after about 2 years of the simulation, rather than because of the different simulation schemes. We also tested that two Gillespie simulations depart around this time horizon.

In this model situation, we illustrate the iterated filtering. The states are \( x = (S, I) \), with the stochastic dynamics \( f \) as specified by the reaction scheme equation

\[
\begin{align*}
\frac{d}{dt}S &= -\beta(t)S(I + e) \\
\frac{d}{dt}I &= \beta(t)S(I + e) - \gamma I, \quad \frac{d}{dt}R = \gamma I - \alpha R, \\
\frac{d}{dt}S &= \mu S - \beta(t)S(I + e), \quad \frac{d}{dt}I &= \beta(t)S(I + e) - \gamma I - \alpha R,
\end{align*}
\]
and the observations $y = I_{\text{cum}}$, the cumulative incidences in a given time interval, here one month, and an additional observation process could be modelled, in which not all new cases are observed, but only those with a probability equal to or less than 1, specifying the observation process $g$. Parameters in $\theta$ are given by the seasonality and by the import. We assume for simplicity now that the other parameters are known and keep them fixed, but they could be included in the parameter vector to be estimated as well. We also estimate all initial conditions, hence these are included in the generalized parameter vector $\theta$.

### 4.3 Constructing a particle filter: particle weights from dynamic noise

We compare the dataset $I_E = (I_1, I_2, \ldots, I_E)$, with dimension $E$ (here $E = 6$ months), with $K$ Euler
Figure 12. (a) Initial distribution versus and (b) final distribution of the seasonality parameter $\theta$ after filtering.

Figure 13. Results for time slice 39 of the last $m$-run through the data time series for $M = 5$ runs. (a) Cumulative distribution for $K = 100$ stochastic realizations for each of the now $J = 100$ particles and $\eta$-ball size of $\eta = 0.005$. (b) Fluctuations around the data points originated from the particles. Simultaneous variation of parameters and initial conditions: two model parameters, $\ln(\varphi)$, the logarithm of the import, and $\theta$, the seasonality, are drawn from Gaussian distributions with the mean of the original values and standard deviation of 10% of the mean values. Also the two initial conditions $I_0$ (by 10%) and $R_0$ are varied. (c) Local likelihood estimate, i.e. the weight of each particle, for the particle parameter values $\theta$. (d) Local likelihood estimate for $\ln(\varphi)$. 
multi-nomial simulations $I_k(\theta)$ performed with parameter set $\theta_i$ (‘particles’ now in the context of filtering adopted for parameter estimation)

$$\hat{p}(I_E|\theta) = \frac{1}{K} \sum_{k=1}^{K} H(\eta - \| I_E - I_k(\theta) \|_E).$$

(4.3)

simulations in the $\eta$-ball region around the data, with $H(\cdot)$ being the Heaviside step function. This gives an estimate of the time-local likelihood function $p(I_E|\theta)$; hence, for $K \to \infty$ and $\eta \to 0$,

$$w_j := \hat{p}(I_E|\theta_i) \to p(I_E|\theta_i),$$

(4.4)

giving the weights of particles $w_j$ for the particle filter (figure 10).

We use $J = 10$ particles, original parameter set $\theta_i$, with $K = 100$ simulations each, distances $\text{dist} := \| I_E - I_k(\theta) \|$. Still we have some fluctuations without changing the parameters yet (figure 10).

As the next step, we impose a variation in the parameters given by a Gaussian distribution. We vary seasonality $\theta$ by 10 per cent with a Gaussian distribution

$$p(\theta) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(\theta-\mu)^2/2\sigma^2}$$

(4.5)

with $\mu = \theta_{\text{org}} = 0.12$, the original value used for the data simulation, and $\sigma = \mu/10$, which acts like a Gaussian prior, for $J = 10$ particles, with $K = 100$ simulations each. Most distances are larger than in figure 10, but some are even smaller now (figure 11).

In the same way, we can vary several parameters and initial conditions, we vary seasonality $\theta$, import $\ln(\varrho)$ and initial conditions $l_0$ and $R_0$, all Gaussian, with the same order of magnitude, and still find distances in the same region as before. As an aside, varying the number of initial susceptibles by 10 per cent changes the simulation trajectories a lot, since owing to conservation of the host population, $N = S + I + R$, a small change in $S$ means a huge change in $R$. For construction of priors, one should always bear in mind such side effects of dependencies as the one just mentioned.

Now the weight $w_j$ of particle $\theta$ from estimating the time-local likelihood function for dynamic noise is given by

$$w_j := \hat{p}(I_E|\theta_i) = \frac{1}{K} \sum_{k=1}^{K} H(\eta - \| I_E - I_k(\theta) \|_E),$$

and we perform filtering in the form of sequential importance resampling [32,33] proportional to the
weights $w_j$ of particles $\theta_j$ after each six-month time slice, using an $\eta$-ball size of $\eta = 0.005$ (figure 12).

### 4.4. Particle filter in action

Now going $M = 5$ times through the time series with each $L = 40$ time slice of six months and starting the parameter values not at $\theta = 0.12$ but at $\theta = 0.14$, and not at $\ln(\varrho) = -15.0$ but at $\ln(\varrho) = -13.0$, we find parameters converging to the expected values as shown in the following.

One extra step has to be performed once we go several times through the data time series. The initial variance of the parameters and the initial conditions are decreased by $h$ times through the data time series. The initial variance of the following.

\[ \frac{\partial}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} \]

which turns out to perform best, whereas suggested gradient-like update rules seem only to perform well very close to the final estimates, and otherwise receive rather destructive kicks at the main convergence region towards the best estimates. For the results after five times, see figure 13.

This shows the qualitative behaviour of the algorithm well, especially indicating that the import parameter is rather more difficult to estimate than the seasonality (figure 13c,d).

Now going $M = 20$ times through the time series and more particles, better $\eta$ resolution gives the results shown in figures 14 and 15. The effect of the simulated annealing is clearly visible, with smaller variation in subsequent $m$-runs showing up in the estimates. Surprisingly, the import parameter is now also very well estimated after the additional effort of longer filtering.

When comparing the simulations of the best estimates with the final piece of time series, the predictive power is quite good in the sense that the simulations now are very close to the data (figure 16).

### 5. CONCLUSIONS AND DISCUSSION

We presented the framework of parameter estimation in simple and in complex dynamical models in population biology, with special emphasis on epidemiological applications. Dynamic population noise, as opposed to observation noise, can be used to construct analytical likelihood functions in simple models and also be estimated via simulations in more complex and more realistic models. Especially in models, which show in its mean field approximation already deterministic chaos it is demonstrated to be possible to estimate well so difficult parameters like the import rate. Future research will now be possible to reinvestigate realistic systems such as dengue fever with very good datasets from Thailand, where import already has been identified as one of the key parameters to determine the underlying dynamical structure of the process.

The presented method of filtering is also close to some elements of approximate Bayesian computations (ABCs), especially the aspect of a distance measure between data and simulations [34,35], and, very recently, ABC methods have been applied to some examples from physics and immune signalling displaying deterministically chaotic behaviour [36]. For the selection between different possible models, see Stollenwerk et al. [37] for an application in an ecological case study, and for a recent application to the paradigmatic Nicholson’s blowfly system, an ecological example with chaotic or near chaotic dynamics, see Wood [38]. In the present context, for example, one could see if the available time-series data can distinguish between a four-strain dengue fever model and a simpler model distinguishing only primary and secondary infection and eventually select one and reject the other (which is not always possible given a certain dataset; see Stollenwerk et al. [39]). For an initial study of qualitative similarities and differences in terms of dynamical complexity of such models, see Aguiar et al. [39]. The large number of initial conditions might prohibit the more complex four-strain model from being sufficiently better than the primary versus secondary infection model. Another future research question will be the comparison of spatial stochastic simulations with data, where shifts in epidemiological thresholds are to be expected [40]. Finally, such epidemiological models
might help to guide vaccination policies [41], once the dynamical parameters are sufficiently well known to obtain information on the relevant dynamic scenario behind the epidemiological fluctuations.

This work has been supported by the European Union under FP7 in the projects EPIWORK and DENFREE and by FCT, Portugal, in various ways, especially via the project PTDC/MAT/115168/2009. Further, we would like to thank an anonymous referee for valuable additional references.

REFERENCES

24 Stratonovich, R. L. 1950 Optimum nonlinear systems which bring about a separation of a signal with constant parameters from noise. Radiofizika 2, 892–901.
35 Tonelli, T. & Stumpf, M. P. H. 2010 Simulation-based model selection for dynamical systems in systems and


39 Aguiar, M., Stollenwerk, N. & Kooi, B. W. Submitted. Irregularity in dengue fever epidemics: difference between first and secondary infections drives the rich dynamics more than the detailed number of strains.
