

Prediction and Predictability in Population Biology: Noise and Chaos

N. Stollenwerk¹, L. Mateus¹, F. Rocha¹
U. Skwara^{1,2}, P. Ghaffari^{1,3}, M. Aguiar¹ *

¹ Centro de Matemática e Aplicações Fundamentais, Universidade de Lisboa, Portugal

² Department of Mathematics, Maria Curie-Skłodowska University, Lublin, Poland

³ Complexity and Networks Group, Imperial College London, United Kingdom

Abstract. To determine best predictors and quantify prediction insecurities we investigate an analytically solvable stochastic system from epidemiology in which the time dependent solution of the model, the likelihood function and Bayesian posterior among other quantities can be calculated explicitly as function of given data. We give analytical expressions for the prediction probability conditioned on best estimators of parameters versus conditioned on data only, and marginalized over the parameters, and observe that the prediction insecurity is wider in the second case of conditioning on the data only, as it should be done in empirical studies. Though the concept becomes clear in the analytical study the differences between prediction based on data directly and prediction based on best estimates of parameters is small due to the simplicity of the model which allowed the analytic treatment. In an only slightly more complex model which however already cannot be treated analytically we clearly observe the expected large differences between the two predictions. Finally, we discuss additional aspects in more extended stochastic population biological systems which do not only exhibit fixed point dynamics but also bifurcations into deterministic chaos with short term predictability and long term unpredictability as quantified by the characteristic exponent, the largest Lyapunov exponent. All is developed in a unified framework of time dependent and state discrete stochastic processes typical for many population biological systems, in ecology and in epidemiology, and can also incorporate non-exponential waiting times by simply including more intermediate classes between transitions.

Keywords and phrases: ecology and epidemiology, stochastic processes, Bayesian inference and model comparison, Bayesian prediction, deterministic chaos, Lyapunov exponents

Mathematics Subject Classification: 92B05, 60G07, 60G25, 62M20, 37C70, 37C75

1. Introduction

Prediction is a question of probability, especially in population biology where we encounter relatively small numbers of individuals, also called "small system sizes" e.g. in condensed matter physics and chemistry. In physics and other more traditional fields of science with typically "large system sizes" a

*Corresponding author. E-mail: nico@ptmat.fc.ul.pt, nico.biomath@gmail.com

model was always judged if it could "describe" an experiment, but this also means it should predict the outcome of the experiment, the next time we perform it. Though in most classical fields the outcome of experiments is quite narrow in distribution, such that a clear yes or no can be given in most theories soon, this in part being the art of experimenting. Not so in life sciences where most of the time not experiments are done, but empirical data are the basis. And gathering such empirical data is most of the time already a difficult task and cannot easily be extended to more accuracy. Here stochastic fluctuations have to be taken into account not only in analyzing models describing empirical data but also any prediction has to take stochastic fluctuations into account in the underlying predictive models validated on empirical data. In order to investigate this situation we will analyze an analytically treatable example from epidemiology, in which we can study all steps from modeling via data analysis, i.e. parameter estimation [15], and model comparison [10] to the final step of prediction and predictability, where the predictability is based on data and depends in accuracy on the data quality.

Though the example is motivated by applications in epidemiology, the model ansatz is easily transferable to ecological systems of predator-prey interactions and also to neuronal excitation systems. Most of the analytical steps of this simple case study can be transferred to more complex systems, sometimes in parts or totally performed via simulations in comparison with empirical data. We will demonstrate such an extension to a system which cannot be treated analytically in all details.

Additional aspects of predictability arise typically in population biology, since most of these systems are by nature auto-catalytic. All classical predator-prey models are designed in such a way of auto-catalysis that predators have to find preys to replicated and even preys feed on restricted resources to reproduce themselves. Exactly the same structure is found, somehow less stylized and more obviously observed, in epidemiology, an essentially fast predator-prey system, where infected individuals have to encounter susceptible individuals to create more infected. We will see in the subsequent text that here we encounter challenging mathematical structures which make it very difficult to predict the future of such systems.

After investigating simplest epidemiological models in their stochastic behavior leading to prediction insecurities we then describe more complex systems, which in mean field approximation already display prediction insecurities due to uncertainty in initial conditions, though being completely deterministic in their dynamics. Specifically, based on previously experienced complexity in epidemiological systems [1–4], we now look at the implications for predictability originating from such sensitivity to initial conditions. As an example we will investigate some epidemiological systems with dynamics determined by strongly non-linear effects, i.e. bifurcations of fixed points into limit cycles, limit cycle bifurcations up to torus bifurcations and deterministic chaos [1,3]. Especially, chaotic attractors present positive Lyapunov exponents which measure exponential divergence of neighboring trajectories, and hence measure the prediction horizon beyond which no prediction is possible even with relatively well known, but not exactly known initial conditions.

Finally, we will investigate an epidemiological system in which the deterministically given sensitivity to initial conditions meets additional stochastic insecurities due to noise induced transitions. In a model typical for childhood diseases like measles transitions between co-existing trajectories give additional aspects to predictability. Depending on the nature of the co-existing attractors, locally in time we might find good predictability for some time, but then very poor predictability due to the transition from one to another attractor. In the extreme case we find a deterministically chaotic attractor or semi-attractor co-existing with a periodic attractor [4], hence locally in time divergence and then attraction, depending on which of the attractors the system is at a time.

In section 2 we will introduce the problem of empirical data bases underlying modeling and prediction of population biological systems illustrated in an example from dengue fever epidemiology. In section 3 we will describe the stochastic aspects of prediction and predictability, as well in analytically solvable cases as well as in cases where only numerical treatments are possible, but then along the lines given by the concepts developed in the analytically treatable cases. The Bayesian framework of model parametrization and model comparison is of major importance in developing these concepts. In section 4 we will describe

the implications of sensitivity to initial conditions of deterministic systems to predictability, and in section 5 we will see how stochastic aspects further influence such systems with already deterministically complex dynamics.

2. Empirical data and its quality as the basis of prediction

Let us first give a brief look at some empirical data, here taken from an example in dengue fever epidemiology. The World Health Organization (WHO) recently issued a warning of a major dengue fever outbreak in Singapore for 2013, see Fig. 1 a). This warning includes the recent data and a sharp unprecedented increase of cases is shown. However, one can easily find the data for dengue fever being notified for a number of years in the past, see Fig. 1 b), from which it appears that larger outbreaks than the one we observe in its starting phase that year have been observed before.

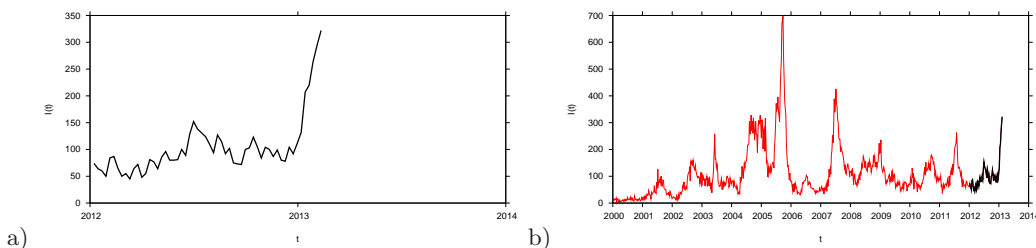


FIGURE 1. Prediction depends on the quality of previous experience in form of long data sets. a) Data of dengue fever cases from Singapore showing a huge increase in number of cases recently. b) In historic perspective the outbreak size at the moment is not unusual to happen over the years.

Hence in a historic perspective the present outbreak can be seen as one of more of this kind, and the challenging question is if it really will become a major outbreak or if it will be in the middle range of previously encountered outbreaks. A more detailed analysis will include extensive modeling of many aspects of dengue fever spreading, and yet we do not have good enough models and data to finally give predictions nor predictability bounds for such systems. The current best models include either multi-strain aspects [1] or at least a distinction between primary versus secondary infection as e.g. models with reinfection [16]. However, we see that a larger data base, here past outbreak data, can give insight into the probability for the presently increasing number of cases. Long data bases are urgently needed to develop descriptive and predictive models for population biological systems, as illustrated here in the case study of dengue fever.

3. Prediction from previous data in the Bayesian framework

In the following we will give a closer look to an example study of the linear infection model, in which from simulated data we can perform all necessary steps analytically towards a descriptive and predictive analysis, based on the underlying data. Previously we have analyzed parameter estimation, in the maximum likelihood framework as well as in the Bayesian framework [15] and recently also studied the important issue of model evaluation, in case several competing models exist for an empirical system and a given data set [10].

3.1. Prediction analytically one step ahead

We now investigate the question of prediction into the future, based on data of infected (I_0, I_1, \dots, I_n) at consecutive time steps with step size Δt . The joint probability of the data points gives the likelihood

e.g. for the linear infection model, which can be treated in all steps analytically, as is described in [15] in quite some detail. This formalizes the problem of data quality of previous times for prediction into the future, which we have just illustrated in the previous section on a management relevant problem like the Singapore dengue data. Here we will give first a brief summary of these results in order to then extend the framework to the new aspect of prediction.

3.1.1. *Stochastic epidemiological process analytically treatable: the linear infection model*

The linear infection model is given by the reaction scheme of susceptibles S becoming infected I by meeting infected from an external and much larger population I^* with infection rate β , hence



Any such reaction scheme can be translated easily into a dynamic process for the probabilities via a master equation, see [15] for the general set-up and many more examples, including spatially extended stochastic systems being describable in the same framework of master equations. For the present model of the linear infection model we have the master equation, i.e. the dynamics of the probabilities to have I infected at time t and population size N ,

$$\frac{d}{dt}p(I, t) = \frac{\beta}{N}I^* \cdot (N - (I - 1))p(I - 1, t) - \frac{\beta}{N}I^* \cdot (N - I)p(I, t) \quad (3.2)$$

with constant force of infection $\beta^* := \frac{\beta}{N}I^*$. A solution can be found via the characteristic function, see e.g. [12],

$$\langle e^{i\kappa I} \rangle := \sum_{I=0}^N e^{i\kappa I} \cdot p(I, t) =: g(\kappa, t) \quad (3.3)$$

with the property of

$$(-i)^n \left. \frac{\partial^n}{\partial \kappa^n} g(\kappa, t) \right|_{\kappa=0} = \langle I^n \rangle \quad (3.4)$$

generating moments. It can be interpreted as Fourier transform with $\kappa =: \frac{2\pi}{N+1} \cdot k$

$$g(\kappa, t) = \sum_{I=0}^N e^{i \frac{2\pi}{N+1} k \cdot I} \cdot p(I, t) = \hat{g}(k, t) \quad (3.5)$$

and then the probability p from g can be obtained via Fourier back-transformation

$$p(I, t) = \frac{1}{N+1} \sum_{k=0}^N e^{-i \frac{2\pi}{N+1} k \cdot I} \cdot \hat{g}(k, t) \quad . \quad (3.6)$$

The dynamics of the characteristic function is given by

$$\frac{\partial}{\partial t}g(\kappa, t) = \sum_{I=0}^N e^{i\kappa I} \cdot \frac{d}{dt}p(I, t) \quad (3.7)$$

with the master equation to be inserted into the expression $\frac{d}{dt}p(I, t)$. After some calculations this gives the partial differential equation (PDE) for the characteristic function

$$\frac{\partial}{\partial t}g(\kappa, t) = \beta^* N ((e^{i\kappa} - 1)) \cdot g(\kappa, t) + i\beta^*(e^{i\kappa} - 1) \cdot \frac{\partial g}{\partial \kappa} \quad . \quad (3.8)$$

The solution for the characteristic function can be obtained by the separation ansatz, and with initially exactly I_0 infected at time t_0 , hence with

$$p(I, t_0) = \delta_{I, I_0} \quad (3.9)$$

the initial condition for the characteristic function is

$$g(\kappa, t_0) = \sum_{I=0}^N e^{i\kappa I} \cdot p(I, t_0) = e^{i\kappa I_0} \quad . \quad (3.10)$$

The solution of the PDE, Eq. (3.8), is given by

$$g(\kappa, t) = e^{i\kappa N} \cdot \left(e^{-i\kappa} e^{-\beta^*(t-t_0)} + (1 - e^{-\beta^*(t-t_0)}) \right)^{N-I_0} \quad , \quad (3.11)$$

see [10], and inverted we obtain with $p(I, t) = \frac{1}{N+1} \sum_{k=0}^N e^{-i \frac{2\pi}{N+1} k \cdot I} \cdot g(\kappa(k), t)$ via the Fourier back-transformation

$$p(I, t) = \binom{N - I_0}{I - I_0} \left(e^{-\beta^*(t-t_0)} \right)^{N-I} \left(1 - e^{-\beta^*(t-t_0)} \right)^{I-I_0} = p(I, t|I_0, t_0) \quad (3.12)$$

the probability as the solution of linear infection model, giving also the transition probability $p(I, t|I_0, t_0)$, since we have the initial condition of exactly I_0 at t_0 .

3.1.2. Likelihood function and best estimate

From the transition probability we can, by investigating the joint probability of the data

$$p(I_n, t_n, I_{n-1}, t_{n-1}, \dots, I_1, t_1, I_0, t_0) = \prod_{\nu=0}^{n-1} p(I_{\nu+1}, t_{\nu+1}|I_{\nu}, t_{\nu}) \cdot p(I_0, t_0) \quad (3.13)$$

and inserting the solution of the stochastic process

$$p(I, t|I_0, t_0) = \binom{N - I_0}{I - I_0} \left(e^{-\beta(t-t_0)} \right)^{N-I} \left(1 - e^{-\beta(t-t_0)} \right)^{I-I_0} \quad , \quad (3.14)$$

obtain the likelihood function, using the abbreviation $\Delta t := (t - t_0)$,

$$L(\beta) = \prod_{\nu=0}^{n-1} \binom{N - I_{\nu}}{I_{\nu+1} - I_{\nu}} \left(e^{-\beta \cdot \Delta t} \right)^{N-I_{\nu+1}} \left(1 - e^{-\beta \cdot \Delta t} \right)^{I_{\nu+1}-I_{\nu}} \quad . \quad (3.15)$$

Then maximizing the likelihood $\frac{\partial L}{\partial \beta} = 0$ gives the best estimator for the parameter β

$$\hat{\beta} = \frac{1}{\Delta t} \cdot \ln \left(\frac{N - \frac{1}{n} \sum_{\nu=0}^{n-1} I_{\nu}}{N - \frac{1}{n} \sum_{\nu=0}^{n-1} I_{\nu+1}} \right) \quad (3.16)$$

see [10, 15] for more details.

3.1.3. Parameter estimation in the Bayesian framework

To give conceptually probabilities to all quantities of interest we can now extend to the Bayesian framework [11]. The data vector $\underline{I} = (I_0, I_1, \dots, I_n)$ is considered in a joint probability of data and parameters, here the parameter β , hence $p(\underline{I}|\beta)$ and then be decomposed into conditional probabilities

$$p(\beta|\underline{I}) = \frac{p(\underline{I}|\beta)}{p(\underline{I})} p(\beta) \quad . \quad (3.17)$$

We have $p(\underline{I}|\beta)$ to be identified with the likelihood function, Eq. (3.15),

$$p(\underline{I}|\beta) = \prod_{\nu=0}^{n-1} \binom{N - I_{\nu}}{I_{\nu+1} - I_{\nu}} \left(e^{-\beta \cdot \Delta t} \right)^{N-I_{\nu+1}} \left(1 - e^{-\beta \cdot \Delta t} \right)^{I_{\nu+1}-I_{\nu}} \quad (3.18)$$

or with abbreviation $\theta := 1 - e^{-\beta \cdot \Delta t}$ we have

$$p(\underline{I}|\theta) = \left(\prod_{\nu=0}^{n-1} \binom{N - I_\nu}{I_{\nu+1} - I_\nu} \right) (1 - \theta)^{\sum_{\nu=0}^{n-1} (N - I_{\nu+1})} \theta^{\sum_{\nu=0}^{n-1} (I_{\nu+1} - I_\nu)} \quad (3.19)$$

This has the functional form

$$p(\underline{I}|\theta) = k_1 \theta^{k_2} (1 - \theta)^{k_3} \quad (3.20)$$

and we have as final probability of the parameter given the data, which we were looking for, the so-called Bayesian posterior

$$p(\beta|\underline{I}) = \frac{p(\underline{I}|\beta)}{p(\underline{I})} p(\beta) \quad (3.21)$$

via $p(\beta|\underline{I}) = p(\theta|\underline{I}) \frac{d\theta}{d\beta}$. With the previously used likelihood function $p(\underline{I}|\beta) = L(\beta)$ and with the beta-function $B(a, b) := \int_0^1 x^{a-1} (1-x)^{b-1} dx$ and $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ the conjugate prior is

$$p(\theta) = \theta^{a-1} (1 - \theta)^{b-1} / B(a, b) \quad (3.22)$$

and corresponds to the functional form of the likelihood function, Eq. (3.20). To calculate the desired posterior $p(\beta|\underline{I})$, we obtain

$$\begin{aligned} p(\beta|\underline{I}) &= \frac{\Gamma(a + b + \sum_{\nu=0}^{n-1} (N - I_\nu))}{\Gamma(a + \sum_{\nu=0}^{n-1} (I_{\nu+1} - I_\nu)) \Gamma(b + \sum_{\nu=0}^{n-1} (N - I_{\nu+1}))} \\ &\cdot (1 - e^{-\beta \Delta t})^{a + \sum_{\nu=0}^{n-1} (I_{\nu+1} - I_\nu) - 1} (e^{-\beta \Delta t})^{b + \sum_{\nu=0}^{n-1} (N - I_{\nu+1}) - 1} \\ &\cdot e^{-\beta \Delta t} \cdot \Delta t \end{aligned} \quad (3.23)$$

as the probability of parameters given the underlying data.

3.1.4. One step ahead prediction in the linear infection model analytically

We have the likelihood function $L(\beta)$ given by the joint probability of finding all data points, interpreted as function of the model parameters, here the single parameter β ,

$$\begin{aligned} p(I_n, t_n, I_{n-1}, t_{n-1}, \dots, I_1, t_1, I_0, t_0 | \beta) &= \prod_{\nu=0}^{n-1} p(I_{\nu+1}, t_{\nu+1} | I_\nu, t_\nu, \beta) \cdot p(I_0, t_0) \\ &= L(\beta) \end{aligned} \quad (3.24)$$

with transition probabilities of a Markov process calculated from the underlying epidemiological model. In simulations we can use directly from the master equation [5, 8, 15, 17] the Gillespie algorithm [6, 7] to obtain stochastic realizations, see Fig. 2. These give us the data points, on which we can test all aspects of parameter estimation and model evaluation.

For prediction purposes, the transition probability can now also be used for the future times $t > t_n = t_{max}$ knowing I_n at t_n . Hence from the last data point we have the prediction distribution given by

$$p(I, t | I_n, t_n, \beta) = \binom{N - I_n}{I - I_n} \left(e^{-\beta(t-t_n)} \right)^{N-I} \left(1 - e^{-\beta(t-t_n)} \right)^{I-I_n} \quad (3.25)$$

and evidently it is a function of the model parameter β

$$p(I, t | I_n, t_n, \beta) = p(I, t | I_n, t_n, \hat{\beta}) \quad (3.26)$$

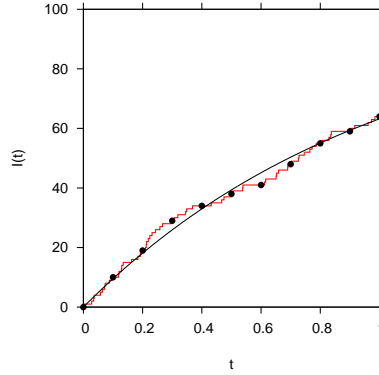


FIGURE 2. Linear infection model generates data, here 10 data points. The parameter is $\beta = 1$ for $N = 100$ individuals.

where in empirical studies β has to be estimated from the data set given. Then the prediction is given with maximum likelihood estimate $\hat{\beta}$ or any best value from the Bayesian posterior $p(\beta|\underline{I})$, maximum, median etc., inserted. Here the posterior distribution $p(\beta|\underline{I})$ can be analytically calculated ones the conjugate prior $p(\beta)$ of the likelihood function $p(\underline{I}|\beta)$ is fixed as shown above.

Then the best prediction \hat{I}_{n+1} for the next time step t_{n+1} is given by the maximum of the distribution $p(I_{n+1}, t_{n+1}|I_n, t_n, \beta)$ via

$$\left. \frac{\partial}{\partial I_{n+1}} \ln p(I_{n+1}, t_{n+1}|I_n, t_n, \hat{\beta}) \right|_{\hat{I}_{n+1}} = 0 \tag{3.27}$$

where the maximization can be performed by continuing the factorial by a gamma-function using $x! = \Gamma(x+1)$ or for large values Stirling's formula $x! \approx e^x \ln(x)$. For quantifying the insecurity of this prediction one can use

$$p(I_{n+1}, t_{n+1}|I_n, t_n, \hat{\beta}) \tag{3.28}$$

which is the distribution of the prediction with surely assumed underlying parameters, hence $\hat{\beta}$ as one fixed value. But then there arises the question: Where is the insecurity from the estimation of the parameter from the underlying previous data (I_0, I_1, \dots, I_n) ?

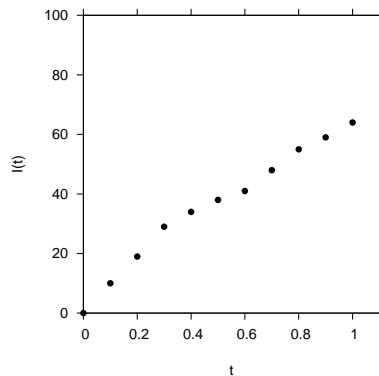


FIGURE 3. The task of prediction from presented data can be formulated in various ways: either model and parameters are known, or parameters have to be estimated. The estimation result depends highly on the data quality.

In order to answer this question we give a closer look to the Bayesian framework, and especially the Bayesian evidence of a model, as used in the model comparison via the Bayes factor [10]. From the prediction probability $p(I_{n+1}, t_{n+1} | I_n, t_n, \beta)$ and the Bayesian posterior $p(\beta | \underline{I})$

$$p(\beta | \underline{I}) = p(\beta | I_1, I_2, \dots, I_n) \tag{3.29}$$

with data vector \underline{I} for short notation, we can construct a joint probability as the product

$$p(I_{n+1}, t_{n+1} | I_n, t_n, \beta) \cdot p(\beta | \underline{I}) = p(I_{n+1}, t_{n+1}, \beta | \underline{I}) \tag{3.30}$$

and integrate over the model parameter β to obtain the prediction based on the underlying data only, and including the parameter insecurity naturally in this way,

$$p(I_{n+1}, t_{n+1} | \underline{I}) = \int_0^\infty p(I_{n+1}, t_{n+1} | I_n, t_n, \beta) \cdot p(\beta | \underline{I}) d\beta \tag{3.31}$$

Only in the limiting case of exactly known parameter in the Bayesian posterior, hence when the posterior is a Dirac delta-function $p(\beta | \underline{I}) := \delta(\beta - \hat{\beta})$, we obtain the previous result $p(I_{n+1}, t_{n+1} | \underline{I}, \hat{\beta})$ of a prediction probability with exactly known parameter also in the prediction distribution.

Hence the prediction probability $p(I_{n+1}, t_{n+1} | \underline{I})$ for the linear infection model, including parameter insecurity, is given by

$$\begin{aligned} p(I_{n+1}, | \underline{I}) &= \int_0^\infty p(I_{n+1}, t_{n+1} | I_n, t_n, \beta) \cdot p(\beta | \underline{I}) d\beta \\ &= \binom{N - I_n}{I_{n+1} - I_n} \frac{B(a + I_{n+1} - I_n + k_2, b + N - I_{n+1} + k_3)}{B(a + k_2, b + k_3)} \end{aligned} \tag{3.32}$$

in terms of the beta-function originating from the conjugate prior of the model, still depending on prior parameters but not explicitly on model parameter β , with $k_2 := \sum_{\nu=0}^{n-1} (I_{\nu+1} - I_\nu)$ and $k_3 := \sum_{\nu=0}^{n-1} (N - I_{\nu+1})$ only being data dependent terms. This is expected to have a wide distribution in case of few data, hence some insecurity on the underlying parameter.

As an example see the data points generated from the linear infection model, as given in Fig. 2, from which we intend to make a prediction to the next time step $t_n + \Delta t$, see Fig. 3. In Fig. 4 we give the actual future of the stochastic realization after gathering the 10 data points on which we will base our prediction. The full line is the mean value of the process, from which the actual stochastic realization deviates often to some extend.

As an additional feature of empirical systems we have to take into account that we might encounter atypical realizations of a stochastic process with quite some probability, see Fig. 5. Realizations can appear entirely above or below the mean curve, given for a specified parameter value. The prediction into the future will, of course, depend on how typical or atypical a realization is in practice.

When the underlying model parameter, here β is exactly known, as would be the case in a well prepared physical experiment, we obtain the binomial prediction distribution as given in Fig. 6.

We compare the prediction under known parameter value with the situation in which we have to estimate the parameter first from the already given data and integrating over the insecurity in the parameter originating from the parameter estimation, via the Bayesian posterior, see Fig. 7. The insecurity is especially pronounced in atypical runs, as can be observed in Fig. 8.

We now will investigate the more general case of prediction not into the near future, hence the next time step $t_n + \Delta t$ but into an arbitrary future time T , which cannot be solved analytically in closed form any more.

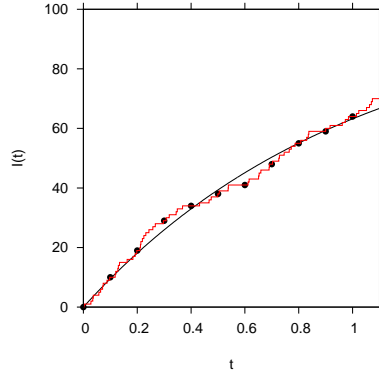


FIGURE 4. The actual history of the process continued into the future $t_{n+1} = 1.1$ after the 10 gathered data points between $t_0 = 0$ and $t_n = 1$.

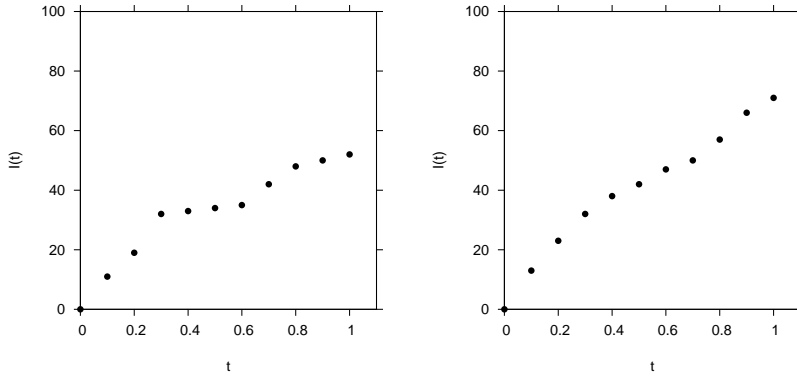


FIGURE 5. Atypical stochastic realizations of the same model with the same parameters as given in Fig 2.

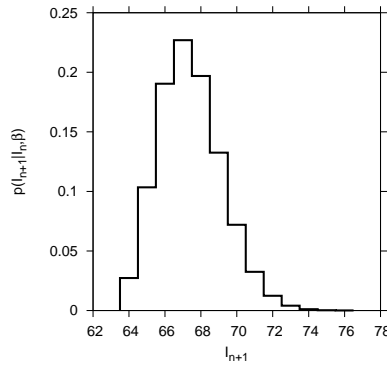


FIGURE 6. Prediction based on known model and known parameter, here $\beta = 1$.

3.2. Prediction analytically for arbitrary prediction period

We describe the prediction distribution with $\theta := 1 - e^{-\beta \Delta t}$ for the given sampling rate of the n data points and with the arbitrary prediction time T as

$$p(I_T, t_n + T | I_n, t_n, \theta) = \binom{N - I_n}{I_T - I_n} \left(1 - (1 - \theta)^{\frac{T}{\Delta t}}\right)^{I - I_n} (1 - \theta)^{\frac{T}{\Delta t} (N - I_T)} \quad (3.33)$$

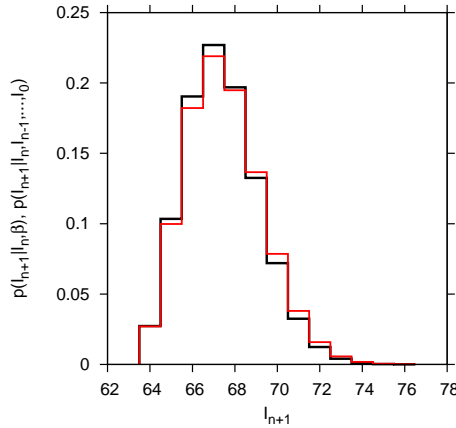


FIGURE 7. Prediction based on known model and known parameter, here β , in comparison with the prediction based on the previous data points only, with integration over model parameters. In the present figure we have a typical stochastic realization, the one from Fig. 2, so that the data points contain good information on the underlying model parameter.

and use the posterior from data estimation

$$p(\theta|\underline{I}) = \frac{\theta^{a+k_2-1}(1-\theta)^{b+k_3-1}}{B(a+k_2, b+k_3)} \quad . \quad (3.34)$$

Hence the prediction into future time T after the last data point, based on the data, is given by

$$p(I_T, t_n + T|\underline{I}) = \int_0^1 p(I_T, t_n + T|I_n, t_n, \theta) \cdot p(\theta|\underline{I}) d\theta \quad . \quad (3.35)$$

This is the situation as given in Fig. 9. It cannot easily be transformed into an integral of a beta-function, as it would be the case for $T = \Delta t$. Hence we only can evaluate the prediction integral numerically, see Fig. 10.

The prediction distribution is clearly wider for the prediction based on previous data only, than in the situation where we would know the model parameter. The present model is very simple in order to perform all steps analytically, hence the effect is here relatively mild, as compared to what we would expect in the case of real world empirical situations like given in Fig. 1 for the data on dengue cases in Singapore. In the analytically treatable linear infection model few data points give already quite good information on the underlying model parameter, whereas in more complex models many more data points would be needed in order to recover the information of eventually several parameters. All major steps which were performed here analytically, can also be performed in simulations for a given stochastic model in comparison with empirical data, see [13, 14] and especially in [12], where in dengue models we find large fluctuations near an absorbing boundary, even in large system size of e.g. a million population size.

We will now describe more complex systems than the analytically treatable given above. Especially, models which show in their mean field approximation already deterministic chaos, give a new twist to the question of prediction, as we will see in the next section. But first we will investigate now the still simple SIS epidemiological model which however cannot be solved analytically, hence no closed form for the likelihood function can be given.

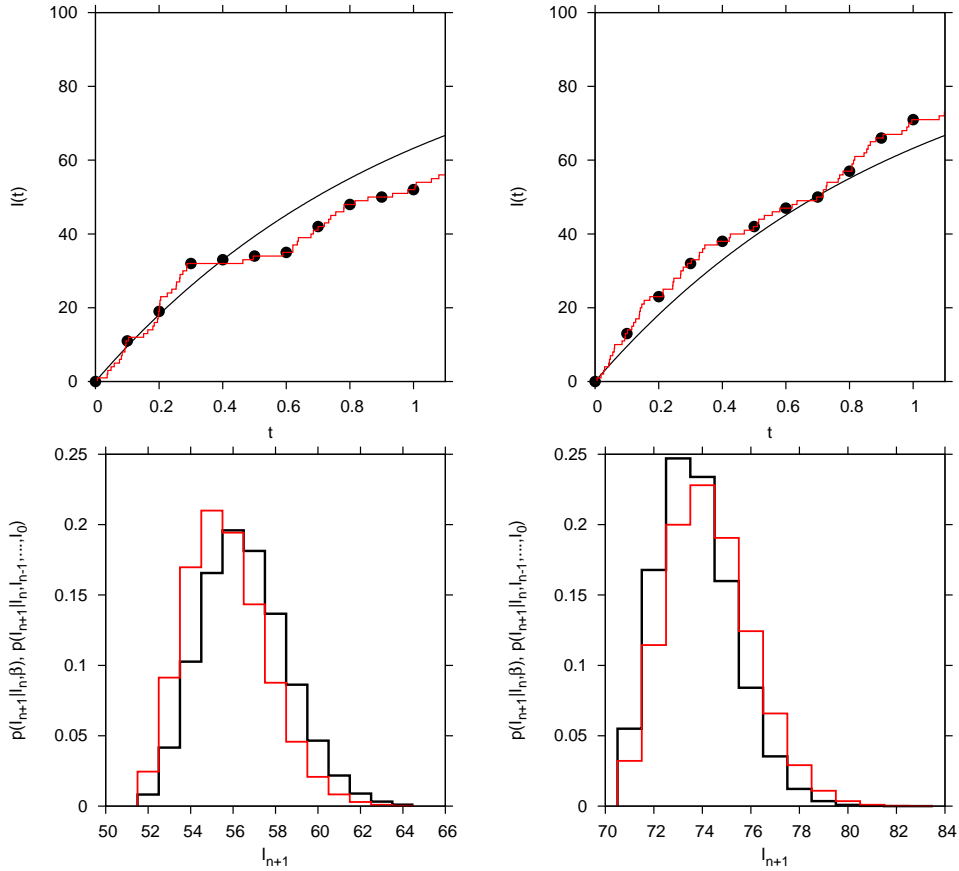


FIGURE 8. Prediction based on known model and known parameter, here β , in comparison with the prediction based on the previous data points only, this time for "atypical" stochastic realizations. Here the data based predictions deviate from the ones for the originally known underlying parameter, since the data give higher probabilities to slightly different parameter values than the originally underlying.

3.3. Prediction in an only numerically treatable case

The SIS epidemiological model describes the infection of susceptibles by meeting infected with rate β , and subsequently infected also can become susceptible again with rate α , hence the reaction scheme now is given by



with the transition probability for the infection process becoming nonlinear in I .

Explicitly the master equation is given by

$$\begin{aligned}
 \frac{d}{dt}p(I, t) &= \frac{\beta}{N}(I - 1) \cdot (N - (I - 1))p(I - 1, t) + \alpha(I + 1)p(I + 1, t) \\
 &\quad - \left(\frac{\beta}{N}I \cdot (N - I) + \alpha I \right) p(I, t)
 \end{aligned}
 \tag{3.37}$$

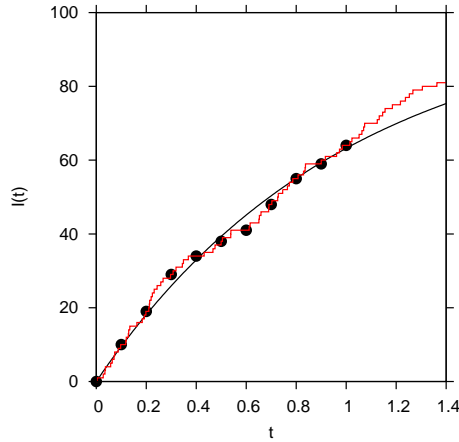


FIGURE 9. Prediction for a general prediction horizon T .

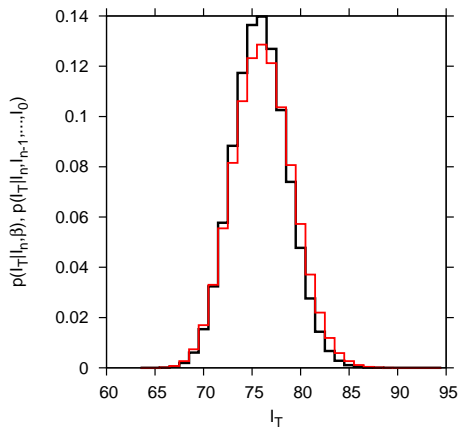


FIGURE 10. Prediction distribution based on known model and known parameter (black) or based on parameter estimation on data only (red), for a general prediction horizon T . We observe a wider distribution for the prediction, based on the data only, than for the prediction, based on the known initial parameter on both sides of the distribution.

and has no closed analytic solution. Hence after simulating again a test data set \underline{I} , we now can only simulate stochastic realizations of the model with varying parameter set, and compare the simulations with the data via a distance measure.

Explicitly, we consider the n dimensional space, where the data set \underline{I} is only one point, and the stochastic realizations of the model for a given parameter set are a cloud in this space. Changing parameters moves the cloud towards or away from the data set, and we count for each parameter set the number of simulation points in an η -ball vicinity of the data set. Fig. 12 gives a two dimensional projection of this n dimensional space, with the data set in the middle of its η -ball neighborhood. Changing a parameter, e.g. the infection rate β , we observe the cloud of simulations moving through the ball with different numbers of points inside the η -ball for each parameter set. The histogram of the number of points in the η -ball neighborhood by changing parameter gives an estimate of the probability of the data set by the given parameter set, here $p(\underline{I}|\beta)$ which is the likelihood function, (see Fig. 12) when the η -ball size around the data vector goes to zero and the number of cloud points goes to infinity. In practice a compromise of η -ball size and number of cloud points has to be made due to limited simulation time

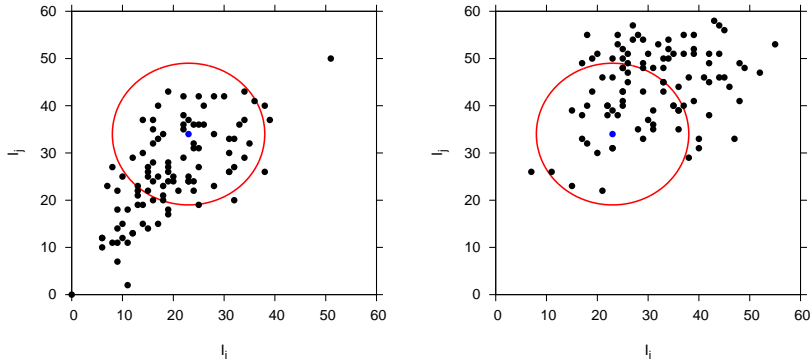


FIGURE 11. Varying parameter shifts the simulated data vectors through the η -ball vicinity of the data vector.

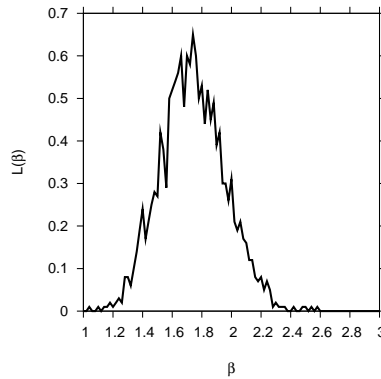


FIGURE 12. Counting the simulated data vectors in the η -ball vicinity of the data vector for changing parameter gives an estimate of the joint probability of the data vector for the given parameter values, i.e. the likelihood.

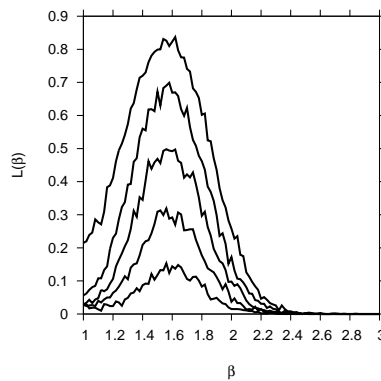


FIGURE 13. Changing the η -ball size keeps for reasonable ranges the maximum essentially at the same place of $\beta \approx 1.6$. The limit of η -ball size going to zero gives the joint probability of the data vector for the given parameter values, i.e. the likelihood, but on the cost of increasing numbers of simulations needed to avoid empty η -balls.

for the model giving the cloud points, but for various η -ball size the method gives stable results for best estimate and likelihood shape, see Fig. 13. For a more detailed description of the η -ball method, see [13] with a test case application, and [12] for more applications with complex dynamics, where the η -ball method is refined to also capture systems with positive Lyapunov exponents by using iterated filtering.

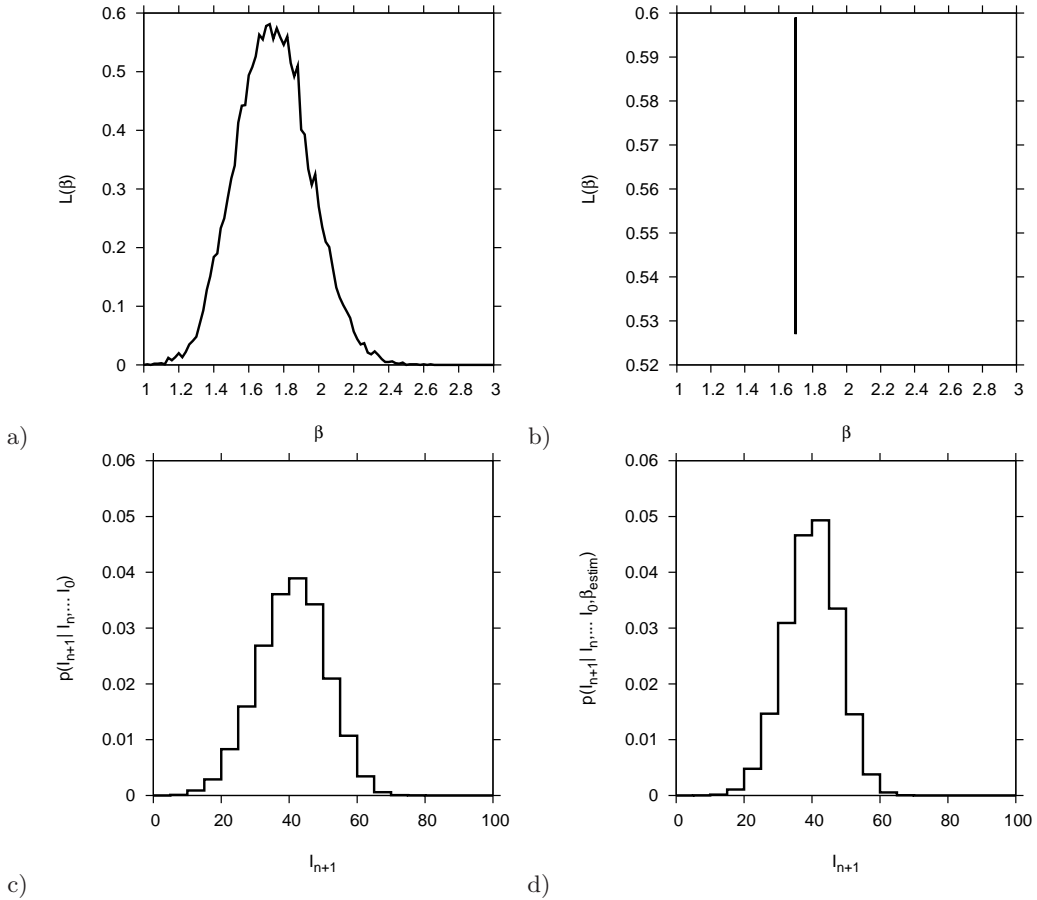


FIGURE 14. The prediction in c) is based on known model and varying parameter, here β varying according to the distribution from the η -ball method as shown in a), and d) is the prediction for know model and best estimate of the parameter β fixed as shown in b) to $\beta \approx 1.65$.

Now for the aspect of prediction and predictability on the basis of given data, we use the same η -ball method, but we continue the simulations giving the cloud per parameter set into the future beyond the time of the data set. Then all simulations with cloud points inside the η -ball vicinity of the data set are taken for the distribution of predictions, see Fig. 14.

We can now again, like in the linear infection model, distinguish between the prediction based on the best estimate from the data via the η -ball method, see Fig. 14 b) for the best parameter value and d) for the prediction distribution. Or we take all simulations for all parameter values, which are again in the η -ball vicinity of the data to construct the histogram of predictions, which corresponds to integrating over all parameter values (marginalizing the prediction probability over the parameters), see Fig. 14 a) for the estimated likelihood and c) for the resulting prediction distribution. This gives the quantified

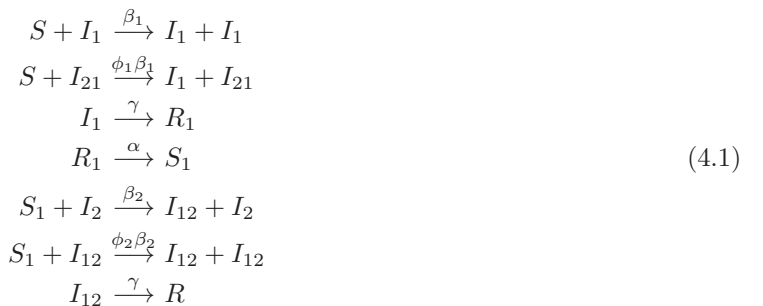
parameter insecurity of the given data set, while the first case of taking only the simulations of the best estimator parameter set neglects the parameter insecurity.

The prediction insecurity is now visibly much wider in the case of quantified parameter insecurity, Fig. 14 c), versus an estimated sharp parameter value d) for the SIS system than in the analytically treatable linear infection model in the previous subsection. The linear infection model is so simple that few data points already specify the parameter perfectly, and the marginalization over the parameter does not contribute much to the prediction insecurity which mainly originates from the population noise only. However in the case of the SIS model, already the parameter insecurity shows up by increasing the prediction insecurity well beyond the population noise insecurity.

An observation, which will only be subject to future detailed studies, but never the less an important observation is that by increasing the number of data points in the SIS system the difference between both prediction cases decreases, hence we are back to the situation of the linear infection model with parameter insecurity not contributing much to the prediction insecurity. There seems to be a certain number of data points or scales on which the data points are sampled, such that the information about the model is sufficient to neglect the parameter insecurity for prediction purposes, i.e. the difference in predictability of best estimator versus marginalization over all parameters is only small, like observed in the linear infection model. In this case the data give basically full information about the model. But as mentioned, a good quantification of this aspect can only be subject to future studies, and this threshold for sufficient information from the data is expected to be highly model specific.

4. Deterministic chaos and predictability horizon via Lyapunov exponents

We will now describe a case study from dengue fever, in which the multi-strain aspect leads to deterministically chaotic dynamics in wide parameter regions, see [1] for the mean field model and [2] for its stochastic counterpart. The simplest multi-strain models are of SIR-type, susceptibles, infected and recovered, with at least two strains, here a reaction scheme typical for dengue fever



from which mean field ODE system and the stochastic counter part follow immediately, see [15] for the framework and [1, 2] e.g for the detailed description of the present dengue fever specific model.

For such models we observe exponential divergence of trajectories in state space due to small insecurities in initial conditions, without any additional stochastic dynamics needed, see Fig. 15, where two time series with exactly the same model parameters are plotted which only differ slightly in the initial conditions. After about 8 years of closeness of the time series the exponential divergence leads to oscillations completely out of phase, which means that the predictability is lost entirely after on average these eight years.

This exponential divergence is measured by the largest Lyapunov exponent of a given dynamical system, see [1] with additional references therein. For the present dengue multi-strain model we calculated the Lyapunov spectrum over a parameter range of the parameter which models the contribution of secondary infection to the force of infection ϕ , see Fig. 16. The largest Lyapunov exponent is for small ϕ negative, indicating a fixed point as attractor, then becomes zero, which indicates a limit cycle as attractor, and for larger values a clearly positive largest Lyapunov exponent indicates deterministic chaos. Here we

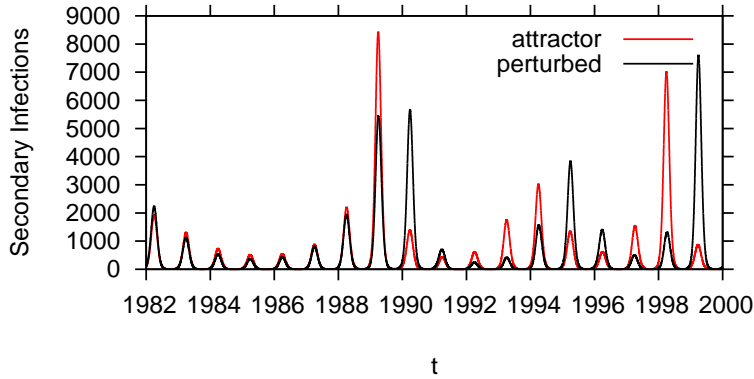


FIGURE 15. Comparison between two time series of the same model and similar initial conditions, only slightly perturbed in initial conditions one from the other.

have short term predictability versus long term unpredictability without any noise source other than the insecurity in the initial conditions.

In such cases noise adds more prediction insecurity to the system, but the largest contribution in such systems originates from the deterministic chaos, hence the deterministic skeleton of the system. For very large system size the stochastic trajectories follow closely the deterministic ones, unless we have stochastic transitions between co-existing attractors, the situation which we will analyze in the next section.

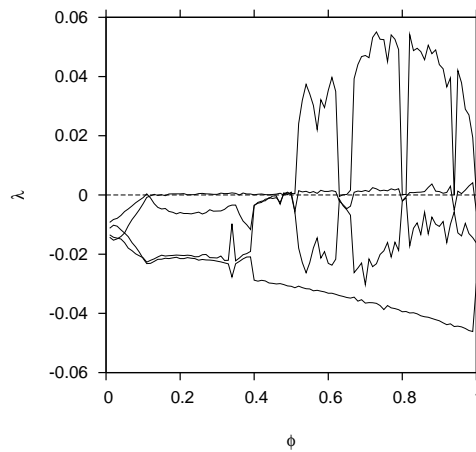


FIGURE 16. Lyapunov spectrum for the dengue multi-strain model, including a positive Lyapunov exponent in some parameter region.

For this we will investigate an epidemiological dynamical system which does not only present deterministic chaos but also transitions between co-existing attractors, a case study from classical childhood infection modeling, since here the state space is lower dimensional than in the multi-strain models and the effects hence easier to study. In this situation the deterministic skeleton of the system and the population noise have a subtle interaction and lead to even wider insecurities of prediction and predictability.

5. Coexisting attractors and noise transitions as challenges for prediction

As a final example we investigate a toy model for the maxima return maps of measles dynamics, similar to the ones investigated in [4]. It is a Hénon type map fitted to the logarithm of the measles maxima return map, as obtained from e.g. the monthly measles cases in new York City from the 1920th to the 1960th when a measles vaccine became available and the dynamics changed qualitatively. Such a two dimensional map can be considered as a Poincaré section of a three dimensional differential equation system, in the present case an SIR system with preserved population size and seasonal forcing of the infection rate given by the following reaction scheme, from which also directly the mean field ODE system can be derived [15],



with recovery rate γ , waning immunity α or in case of life long immunity the replacement of one generation by the next, newborns being susceptible, whereas the majority of death cases are elderly recovered and immune, and import ϱ . Finally the infection rate β is now in seasonally forced disease transmission time dependent with in the simplest case a cosine-forcing

$$\beta(t) = \beta_0 \cdot (1 + \theta \cdot \cos(\omega t)) \quad . \quad (5.2)$$

In such dynamical systems the Poincaré section then has the same topological, hence qualitative behavior as the original differential equation system. This means that we can identify fixed points, limit cycles and chaotic attractors of the original differential equation system by studying the two dimensional Poincaré map.

Several functional forms for the two dimensional map could be fitted to the data and gave the same qualitative dynamical picture. Here we give the map in the following final form

$$\begin{array}{l}
 x_{n+1} = ax_n e^{-h \cdot x_n} \cdot (y_n + c)^{-g} + b \\
 y_{n+1} = x_n
 \end{array} \quad (5.3)$$

with parameters $a = 410$, $b = 135$, $c = 600$, $g = 0.327$, $h = 0.000684$ for measles, evaluated from New York data. The same dynamical picture which we will describe below was also observed in an ordinary differential equation system which had been suggested before in order to model measles and other childhood diseases in a susceptible, exposed, infectious and recovered (SEIR) model. There, due to time scale separation, the exposed class follows closely the infected class and hence exhibits effectively on the epidemiologically relevant time scale the above mentioned SIR system with seasonal forcing.

We first give the dynamic scenario for a parameter value of $a = 400$, see Fig. 17 a). The parameters do not have any direct biological meaning, but determine the functional form of the logarithm of the next year's maximum, here represented by the state variable y_n as a function of the present year's logarithm of maximum of measles infected, represented by x_n . Then the parameter a essentially governs the strength of the non-linearity in the map. For more details see [4].

For the value of $a = 400$ we find a chaotic attractor of maxima return dots filling a fractal with many finer and finer lines when zooming into the attractor. Co-existing next to the chaotic attractor we observe a period three solution, given in Fig. 17 a) by the three full dots which are periodically visited in always the same order, hence the system is here perfectly predictable, whereas in the chaotic attractor the system has the short term predictability and long term unpredictability described already in the previous section. Also we find an unstable period three, given by the circles, in which we would have periodic behaviour only exactly on the sequence of circles, but any what so ever small deviation or any noise would drive the

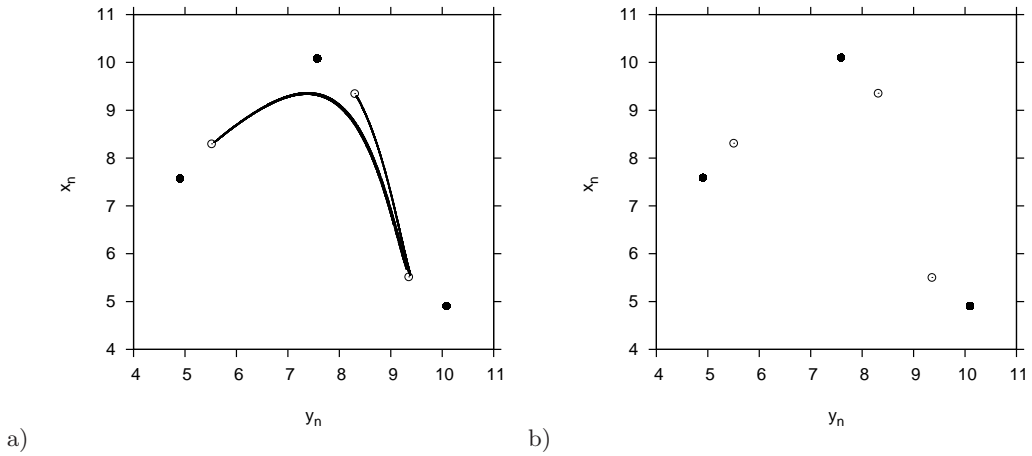


FIGURE 17. a) The attractors of the maxima return map at $a = 400$ and b) at $a = 410$.

system away from the unstable period three towards one of the stable attractors, the chaotic attractor or the stable period three.

Now we investigate the system for the actually best fitting parameter value of $a = 410$, see Fig. 17 b), where the chaotic attractor has disappeared and the stable period three is the only and globally attracting structure with perfect predictability, or good predictability in the noisy versions. Just for orientation we also show the unstable period three in this graph.

Only the further analysis of the unstable period three and its stable manifold as basin of attraction boundaries, as shown in Fig. 18, reveals the complete picture of the dynamics: The stable manifold is composed of all points which by iteration end up in the unstable period three, and any small perturbation will lead the initially on the stable manifold lying points to either the chaotic attractor or to the period three attractor. In this sense this manifold is a boundary of the basins of attraction of the co-existing attractors. Depending on which side of the manifold the state is initial located the trajectory will fall into one or the other attractor. This basin boundary is in itself again fractal.

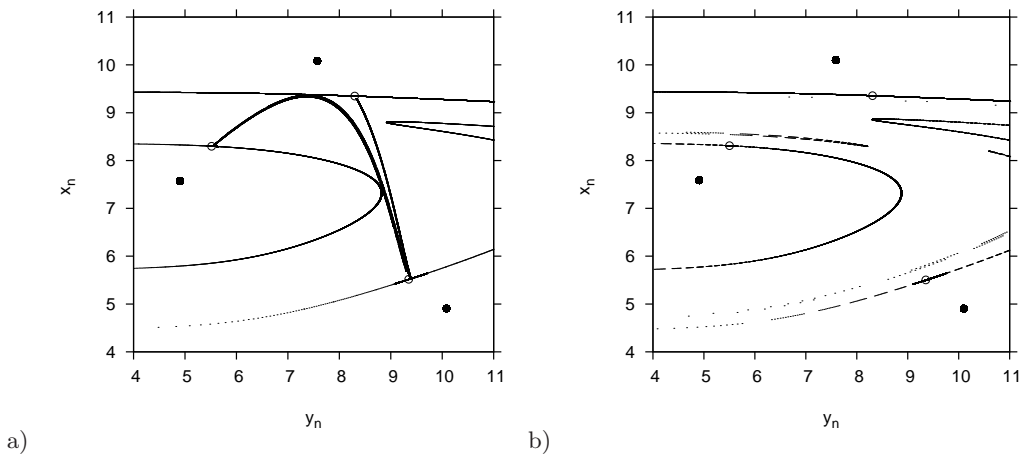


FIGURE 18. a) Maxima return map at $a = 400$ including basin boundaries as the stable manifold of the unstable period three. b) Same as in a) now for the map at $a = 410$. The basin boundary now cuts into the chaotic attractor, transforming it into a semi-attractor.

In Fig. 18 a) we show the situation for $a = 400$, where chaotic and periodic attractor co-exist. The basin boundary is given as filled lines around the unstable period three in addition to the already shown structures in Fig. 17 a).

When increasing the parameter a only slightly, hence the measles relevant value of $a = 410$, see Fig. 18 b), we observe that the stable manifold of the unstable period three cuts into the chaotic attractor, and does so infinitely often then. Hence when starting on the chaotic attractor after some iterations the system falls into the gaps which were cut into the chaotic attractor by the basin boundary, and ultimately ends up in the stable period three attractor. The chaotic attractor has disappeared in a so called boundary crisis. The chaotic attractor has become a transiently chaotic structure which will always end up as a repeller of the dynamics.

However, in the stochastic version of the same system, given by the stochastic map

$$\begin{aligned} x_{n+1} &= ax_n e^{-h \cdot x_n} \cdot (y_n + c)^{-g} + b + x_n \cdot \sigma \cdot \varepsilon_n \\ y_{n+1} &= x_n + y_n \cdot \sigma \cdot \eta_n \end{aligned} \quad (5.4)$$

with Gaussian white noise stochastic variables ε_n and η_n , standard deviation $\sigma = 0.3$, here in the simplest stochastic extension of the just described deterministic map, stabilizes the transiently chaotic repeller, as seen in Fig. 19. While the system is periodically oscillating around the deterministically stable period three with some simple stochastic fluctuations around, similar to the previously described situations e.g. in the example study of the linear infection model, it occasionally crosses the boundary into the semi-attractor, which is the transient remain of the chaotic attractor, and stays there for a while until only later it revisits the period three in a stochastic fashion.

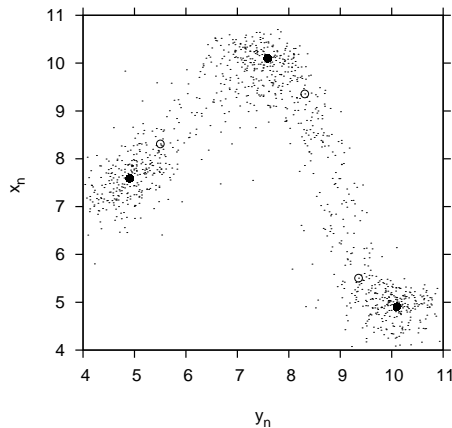


FIGURE 19. The map at $a = 410$ under noise. Transitions between fluctuations around period three and around the chaotic semi-attractor are visible.

The stochastic system is at times relatively well predictable, when visiting the period three solution, but becomes long term unpredictable and short term predictable (up to the noise fluctuations) when visiting the chaotic semi-attractor. Hence in such a system the predictability is a matter of this subtle interplay between coexisting structure in the deterministic skeleton and the noise induced transitions between these deterministic structures.

Similar scenarios of co-existing attractors were also found in higher dimensional multi-strain models for dengue fever [3]. Here the main bifurcation branch originating from a fixed point and finally going through a torus bifurcation into chaos co-exists with periodic limit cycles which appear as so called isolas which in their own again go through a period doubling bifurcation route into chaos. Since the system is in this case at least nine dimensional for only two interacting strains (and would be even higher for more

interacting dengue strains) we cannot easily investigate graphically the state space scenario as we could in the example of the measles maxima return map which gives in two dimensions already all information. A maxima return map for the simplest two-strain dengue model is eight dimensional. So one has to rely on bifurcation software to analyze the torus bifurcation and the isolas. But all qualitative statements on noise induced transitions and their consequences for prediction are holding in such more complex systems as well.

6. Conclusions and outlook

We have revisited several epidemiological systems which have been used as descriptive models for empirical data. The aspect of prediction, prediction insecurity and predictability have been newly studies in these systems. Especially the aspect of prediction based on previous data and hence carrying over the insecurity from the parameter estimation and model evaluation have not been previously studied extensively in the Bayesian framework (see e.g. standard references like [11]), and at least in epidemiology and ecology to our knowledge not been used as it should be.

Due to its inherently nonlinear structure population biological models often present additional features of unpredictability already visible in deterministic approximations of the underlying stochastic processes, namely bifurcations and deterministic chaos, hence sensitivity to initial conditions leading to short term predictability and long term unpredictability. We demonstrate the implications for prediction in a case study from dengue fever, where the multi-strain aspect of the pathogen, here the dengue viruses, gives complex behavior, i.e. bifurcations and deterministic chaos, in wide parameter regions.

Finally, we describe in a model of measles outbreak maxima return map the additional feature of noise induced transitions, hence illustrate in a very simple and naturally appearing example from epidemiology a very complex behavior with varying predictability along the time course, as measured of initially neighboring points deviating in the course of iteration. A given realization of the stochastic process alone would be quite difficult to be investigated in order to find the underlying dynamic structures. However, this is the level of complexity to be ultimately expected in real world applications. New tools of investigating data with possibly chaotic dynamics are now becoming available, still with a lot of technical difficulties, like iterated filtering of particles, i.e. parameter sets and initial conditions given by a dynamic stochastic model, see [12]. The challenge will be to see in how far we can now deal with such levels of complexity in descriptive and ultimately predictive models in epidemiology and other population biological systems, like in ecology and eventually neuro-biological systems of many neurons interacting.

All models described here are developed in a unified framework of time dependent and state discrete stochastic processes typical for many population biological systems, in ecology and in epidemiology, and can also incorporate non-exponential waiting times by simply including more intermediate classes between transitions. The waiting times between first and last transition follow Erlang distributions known from queuing theory, where the limiting case of many intermediate classes converges to a Gaussian distribution around a mean waiting time. So the apparent exponential transition time distributions between classes are no real restriction of the framework, but extensions can easily be incorporated.

Acknowledgements. This work has been supported by the European Union under FP7 in the projects EPIWORK and DENFREE and in various ways by FCT, Portugal, especially via the project PTDC/MAT/115168/2009. We thank Friedhelm Drepper, Jülich, Ralph Engbert, Potsdam, and Bob Kooi, Amsterdam, for many discussions on various aspects of the present review.

References

- [1] M. Aguiar, S. Ballesteros, B. Kooi, N. Stollenwerk. *The role of seasonality and import in a minimalistic multi-strain dengue model capturing differences between primary and secondary infections: complex dynamics and its implications for data analysis.* Journal of Theoretical Biology, 289 (2011), 181–196.
- [2] M. Aguiar, N. Stollenwerk, B. Kooi. *Scaling of stochasticity in dengue hemorrhagic fever epidemics.* Math. Model. Nat. Phenom., 7 (2012), 1–11.

-
- [3] M. Aguiar, N. Stollenwerk, B. Kooi. *Torus bifurcations, isolas and chaotic attractors in a simple dengue fever model with ADE and temporary cross immunity*. Intern. Journal of Computer Mathematics, 86 (2009), 1867–77.
- [4] F. Drepper, R. Engbert, N. Stollenwerk. *Nonlinear time series analysis of empirical population dynamics*. Ecological Modelling, 75/76 (1994), 171–181.
- [5] C.W. Gardiner. Handbook of stochastic methods. Springer, New York, 1985.
- [6] D.T. Gillespie. *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*. Journal of Computational Physics, 22 (1976), 403–434.
- [7] D.T. Gillespie. *Monte Carlo simulation of random walks with residence time dependent transition probability rates*. Journal of Computational Physics, 28 (1978), 395–407.
- [8] J. Honerkamp. Stochastic Dynamical Systems: Concepts, Numerical Methods and Data Analysis. VCH Publishers, Heidelberg, New York, 1993.
- [9] D. MacKay. (2005) Information theory, inference, and learning algorithms. Cambridge University Press, Cambridge, 2005.
- [10] L. Mateus, N. Stollenwerk, J.C. Zambrini. *Stochastic Models in Population Biology: From Dynamic Noise to Bayesian Description and Model Comparison for Given Data Sets*. Int. Journal. Computer Math., 90 (2013), 2161–2173.
- [11] D.S. Sivia. Data analysis: A Bayesian tutorial. Oxford University Press, Oxford, 1996.
- [12] N. Stollenwerk, M. Aguiar, S. Ballesteros, J. Boto, B. Kooi, L. Mateus. *Dynamic noise, chaos and parameter estimation in population biology*. Roy. Soc. Interface Focus, 2 (2012), 156–169.
- [13] N. Stollenwerk, K. Briggs. *Master equation solution of a plant disease model*. Physics Letters, A 274 (2000), 84–91.
- [14] N. Stollenwerk, F. Drepper, H. Siegel. *Testing nonlinear stochastic models on phytoplankton biomass time series*. Ecological Modelling, 144 (2001), 261–277.
- [15] N. Stollenwerk, V. Jansen. Population biology and criticality. Imperial College Press, London, 2011.
- [16] N. Stollenwerk, S. van Noort, J. Martins, M. Aguiar, F. Hilker, A. Pinto, G. Gomes. *A spatially stochastic epidemic model with partial immunization shows in mean field approximation the reinfection threshold*. Journal of Biological Dynamics, 4 (2010), 634–649.
- [17] N.G. van Kampen. Stochastic Processes in Physics and Chemistry. North-Holland, Amsterdam, 1992.