SHAR and effective SIR models: from dengue fever toy models to a COVID-19 fully parametrized SHARUCD framework

Maíra Aguiar[1,2,3*] **and Nico Stollenwerk**[1,4]

[1]Dipartimento di Matematica, Università degli Studi di Trento,
Via Sommarive, 14 – 38123 Povo (Trento), Italy
[2]Basque Center for Applied Mathematics (BCAM),
Alameda Mazarredo, 14, 48009 Bilbao, Spain
[3]Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[4]Center for Mathematics, Fundamental Applications and
Operations Research, Lisbon University, Portugal

[*]Email: m.aguiar@unitn.it

Abstract

We review basic models of severe/hospitalized and mild/asymptomatic infection spreading (with classes of susceptibles S, hopsitalized H, asymptomatic A and recovered R, hence SHAR-models) and develop the notion of comparing different models on the same data set as exemplified in the comparison of SHAR models with effective SIR models, where only the H-class of the SHAR model is taken into account in the SIR model. This is done via the so-called Bayes factor.

A simpler pair of models with analytical expressions up to the Bayes factor will be briefly mentioned as well. The notions developed with respect to dengue fever epidemiology will then be used to analyze recently becoming available data on coronavirus disease 2019, COVID-19, where models can be fully parametrized including hospital admission and more extensions like intensive care unit (ICU) admission and deceased, always with a close look on as simple as possible models but not simpler, as exercised in Ocham's razor and analyzed by e.g. the Bayes factor.

We present the resulting models of SHAR-type with additional classes of ICU admissions U, and deceased D, and for data analysis of cumulative disease data, also accounting the cumulative classes C, in the so-called SHARUCD framework. Besides a first basic version, SHARUCD model 1, we investigate also in detail a refined version, SHARUCD model 2, which could be achieved by a closer analysis of available data only obtained after the exponential growth phase of the epidemic, when lockdown control measures showed effects.

Namely, the ICU admissions turned out to be more in synchrony with the hospitalized than with e.g. the deceased cases, such that we could adjust the transitions so that ICU admissions are modeled like hospitalizations in model 2, and not like recovery or disease induced death as assumed in model 1, explaining much better the empirical data, specially after the effects of the lockdown became visible.

Special attention will be given here, for the first time, to the initial phase of the COVID-19 epidemics, before all variables entered into the exponential phase, and its interplay between asymptomatic and severe hospitalized cases, always in close check with the SIR-limiting case. Such improved understanding of the initial phase will help in the future analysis of re-emergent outbreaks of COVID-19, likely to happen in the next or a subsequent respiratory disease season in autumn or winter.

Keywords: COVID-19, Bayes factor, master equation, Fokker-Planck equations, epidemiological growth factors.
2010 MSC classification number: 92D30, 92B05, 60H35, 97Mxx, 93A30, 62C10.

1. INTRODUCTION

Good epidemiological data are often available for many infectious diseases outbreaks, from severe or hospitalized cases, well recorded and documented, however, the dynamic process of disease spreading can be as well driven by less severe infection, often reported as mild or even asymptomatic cases. Scenarios of epidemiological spreading of severe/hospitalized and mild/asymptomatic infection cases in susceptible, S, hospitalized, H,asymptomatic, A, and recovered, R, systems via SHAR-models have been evaluated for long

*Corresponding author
Received July 1^{st}, 2020, Revised July 9^{th}, 2020, Accepted for publication July 15^{th}, 2020. e-ISSN: 2549-2896, DOI:10.5614/cbms.2020.3.1.6

time [1], [2]. These models were mainly motivated by observations in dengue fever, where primary infections are often mild and the main risk of severe disease is epidemiologically linked to a secondary infection caused by heterotypic dengue virus [3], [4], [5]. The best documented dengue data are of severe and hospitalized notified cases [6], representing the "peak of the iceberg", which gives, nevertheless, good information about complex dynamic features due to recurrent outbreaks over many decades [3]. However, as the mild infections are mostly unrecorded, it is well known that those cases are of major importance for the epidemiological dengue spreading.

A wide spectrum of symptoms is observed for many infectious diseases, from severe and potentially lethal disease to milder forms such that the infected individual is often not aware of being infective due to a sub-clinical or even asymptomatic course of infection. In bacterial meningitis caused by Neisseria bacteria, for example, the pathogens are mostly commensal for the human host and only rarely, due to mutations, turn out to be pathogenic, the so called accidental pathogens [7], [8], [9]. Via critical fluctuations [10], [9], we observe that the more pathogenic the sooner they exhaust their basis, becoming extinct, whereas milder mutants would survive longer time periods in the host population, eventually becoming commensals and not harming their host in any noticeable way.

A similar pattern is noticed in the current coronavirus disease 2019 outbreaks, COVID-19, where we observe a relatively high disease induced mortality as compared to other well established influenza-like illnesses, ILI, or any other human coronaviruses circulating during in the seasonal respiratory diseases period, however milder forms are frequent. The COVID-19 case fatality ratio is decreasing over time as testing capacities are increasing allowing the detection of infectious cases with mild or no symptoms. Data on hospitalizations as well as on infected but not hospitalized cases, which we call mild or asymptomatic infections, are now available such that SHAR-type models can be parametrized on actual outbreak data. For COVID-19 cases, a variable spectrum of disease severity is observable, from very severe, prone to hospitalization up to intensive care unit admissions and disease induced death, to milder forms with some characteristic disease symptoms, but not needing hospitalization, to very mild forms with symptoms easily confused with other respiratory diseases to even asymptomatic. See for a very recent study on seroprevalence in Spain [11], where up to 30 % asymptomatic are reported to be detected among the seropositive for immunoglobin G, IgG, antibodies.

In this study we distinguish between severe and potentially hospitalized to be always detectable with a positive PCR test, in the H class, and mild up to asymptomatic, with a proportion ξ detected, which still contribute to the force of infection, in the A class. We use stochastic models to estimate parameters from the available data and models are parametrized as they can describe the exponential phase of the epidemic.

In March 2020, a multidisciplinary task force, the so-called Basque Modelling Task Force, BMTF, was created to assist the Basque Health managers and the Basque Government during the COVID-19 responses. Withing this BMTF we developed a modelling framework able to describe the COVID-19 outbreak in the Basque Country in so-called SHARUCD models, using data referring to all positive PCR tested cases classified as hospitalization, intensive care unit (ICU) admissions, deceased, and later also recovered. Besides the already described SHAR compartments and ICU admissions, U, compartments for cumulatively notified disease classes, C, namely C_H, C_A, C_U and C_R, and deceased, D, were introduced [12], [13]. The compartments H, A, U and R are dynamical compartments whereas the corresponding C compartments are counting only the entries cumulatively or in the case of C_R only notified parts of entries, to compare with the empirical cumulative cases in the Basque Country.

In the basic toy models for dengue fever, which are sketches of more extended multi-strain models [3], a comparison between SHAR models and simpler SIR models could be performed in the transition towards stationarity and fluctuations around stationarity, showing that data generated by the more complex SHAR model could still be described by the simpler SIR model when taking shifts in parameters into account [2]. In a Bayesian framework the original parameters of the SHAR model as well as effective parameters of the simplifying SIR model could be estimated and probabilities could be given for one model versus the other via the Bayes factor, a quotient of Bayesian evidences of the compared models. This demonstrates that often empirical data can give some information about dynamics of different models but not sufficient to distinguish between them. The key question is what parameters the different models obtain and how models can be compared [14].

In the present work we use the above notions to investigate for the SHARUCD models, describing and being fully parametrized on COVID-19 data from the Basque Country, limiting cases of SIR-type models and in how far extra information is obtained by adding more realistic features in the SHARUCD framework

without over-parameterizing. If in previous work we have demonstrated that on available empirical data some model refinements could be performed giving better results [12], here we describe how more information can eventually be extracted from the present and future revisited data, including the disease introduction phase. We face the present situation in COVID-19 research that many different models have been recently developed describing often similar data. So a more formal procedure for model comparison will help in future understanding of the recent pandemic and its future development in the soon upcoming respiratory disease seasons. From experiences in other respiratory diseases, ILI, it would be expected to observe some seasonal effect on the spreading of COVID-19. Although still difficult to disentangled disease control in the Norther hemisphere from the lockdown measures, we should consider differences in disease speeding via seasonality. Note that the first cases in China started in mid winter, with the disease spreading, via imported cases, into the European countries in different times, but still in winter. However, as the lockdowns started differently, most of these European countries were able to control disease transmission at similar times. That could indicate some seasonal effect, hence not completely excluding the possibility of a second wave to begin in autumn. Nevertheless, more information about SARS-CoV-2 virus has to be gained for further evaluations of risks, now being in the phase of isolated outbreaks after the first lockdown relaxation.

In Section 2 we describe the model comparison of different models applied to the same data set along the analytically treatable examples of the linear infection model and the Poisson model. In Section 3 the SHAR modelling framework is introduced and numerical comparison of the SHAR models with effective SIR models is tested. In Section 4 we describe the SHARUCD modelling framework to model COVID-19 based on the previous notions of the SHAR models, where here the SHARUCD models would be the minimalistic models describing quantitative features of the data, where many more detailed models have been suggested recently, see e.g. [15], where the role of asymptomatic infected is emphasized, and many compartments are used, just any difference in infectivity between severe and mild is lacking. Any difference between infectivity of mild versus severe cases, actually, gives us technically tools to distinguish between the SIR-limiting case and the SHAR specific properties of the disease dynamics. Section 4 gives two particular model versions of the SHARUCD-type, the basic model first analyzed on data of COVID-19 in the exponential growth phase, and an improved version, SHARUCD model 2, obtained in the phase of lockdown control measure effects, in order to describe new incoming data features. Here we include new results on this improved model to also capture more features of the disease introduction phase, i.e., before the COVID-19 exponential phase, where with "exponential phase" we mean the explosive growth phase of all observed variables with the same growth factor.

Since in the pre-exponential phase, some of the variables still lack behind they eventually give more information on some of the modelling parameters and the dynamics of any introduction or re-introduction of COVID-19 into a new area. Still more high quality data for this phase have to be sought to improve our understanding of this very important initiation phase including stochastic fluctuations from external index cases.

2. Model comparison in analytically treatable cases

2.1. Introduction

In this section we give a brief first example of how to compare different models on the same empirical data. We derive the basic notions of likelihood, Bayesian prior and posterior along an analytically treatable model, the linear infection model. Then we derive a simplifiction of this process, the Poisson model, in the same analytically treatable way, and finally compare both models on the same data set, obtained from stochastic simulations of the more complex linear infection model via probabilities of one model conditioned on the data and probabilities of the other model conditioned on the same model, in the so-called Bayes factor. For a more detailed description on this procedure see [14]. Then in the next section we will transfer all notions obtained here to model comparisons of SHAR and SIR type, where the analysis can only be performed numerically, but with the same notions as developed here in the analytically treatable case.

2.2. The linear infection model as a simple analytically tractable stochastic system

With susceptible individuals S and infected I giving a population of size N we have one of the simplest models in epidemiology given by the following reaction scheme

$$\begin{array}{rcl} S+I & \stackrel{\beta}{\longrightarrow} & I+I \\ I & \stackrel{\alpha}{\longrightarrow} & \end{array} \tag{1}$$

with infection rate β and recovery rate α. Hence the probability of the system is given by the stochastic variable I, and with constant population size following $S = N - I$ as master equation

$$\begin{aligned} \frac{d}{dt}\,p(I,t) &= \frac{\beta}{N}(I-1)(N-(I-1))\cdot p(I-1,t) + \alpha(I+1)\cdot p(I+1,t) \\ &\quad - \left(\frac{\beta}{N}I(N-I)+\alpha I\right)\cdot p(I,t) \end{aligned} \tag{2}$$

not easily solvable explicitly. However, for a population of susceptibles being infected from an outside and much larger pool of infected I^*, and no internal infection and no recovery (or counting only cumulative infected) we have the even simpler linear infection model with the following reaction scheme

$$S + I^* \stackrel{\beta}{\longrightarrow} I + I^* \tag{3}$$

and its master equation

$$\frac{d}{dt}\,p(I,t) = \frac{\beta}{N}I^*(N-(I-1))\cdot p(I-1,t) - \frac{\beta}{N}I^*(N-I)\cdot p(I,t) \quad . \tag{4}$$

For the linear infection model see e.g. [9] and references therein, and for general introduction to such stochastic processes e.g. [16], [17], [18]. With the abreviation $\beta^* := \frac{\beta}{N}I^*$ the linear infection model then is

$$\frac{d}{dt}\,p(I,t) = \beta^*(N-(I-1))\cdot p(I-1,t) - \beta^*(N-I)\cdot p(I,t) \tag{5}$$

and can now be analytically treated.

2.3. Analytic solution of the linear infection model with characteristic function

The characteristic function

$$\langle e^{i\kappa I}\rangle := \sum_{I=0}^{N} e^{i\kappa I}\cdot p(I,t) =: g(\kappa,t) \tag{6}$$

$$= \sum_{I=0}^{N} e^{i\frac{2\pi}{N+1}k\cdot I}\cdot p(I,t) = \hat{g}(k,t) \tag{7}$$

gives the tool for solving Eq. (5). With the wave number k defined via $\kappa =: \frac{2\pi}{N+1}\cdot k$ we have $\hat{g}(k,t)$ as Fourier transform with inverse Fourier transformation giving again the probability

$$p(I,t) = \frac{1}{N+1}\sum_{k=0}^{N} e^{-i\frac{2\pi}{N+1}k\cdot I}\cdot \hat{g}(k,t) \quad . \tag{8}$$

Hence we just have to solve a dynamical equation for g obtained from the dynamics of the probability and use Fourier back transform to obtain the solution for the probability p via

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

obtaining

$$\frac{\partial}{\partial t}\,g(\kappa,t) = \beta^* N\left((e^{i\kappa}-1)\right)\cdot g(\kappa,t) + i\beta^*(e^{i\kappa}-1)\cdot\frac{\partial g}{\partial \kappa} \tag{9}$$

as dynamic equation. A separation ansatz $g(\kappa,t) := h(\kappa)\cdot\ell(\kappa,t)$ solves this partial differential equation provided we use a second step of $\ell(\kappa,t) := m(\kappa)\cdot n(t)$. The first separation gives

$$\frac{\partial \ell}{\partial t} = i\beta^* \left(e^{i\kappa}-1\right)\right) \frac{\partial \ell}{\partial \kappa} \quad \text{and} \quad \frac{dh}{d\kappa} = iN\cdot h(\kappa)$$

with solution $h(\kappa) = e^{iN\kappa}$. A special solution of the second separation is

$$\frac{dn}{dt} = i\beta^*\cdot n(t) \qquad \Rightarrow \qquad n(t) = e^{i\beta^* t}$$

and

$$\frac{dm}{d\kappa} = \frac{1}{e^{i\kappa}-1}\cdot m(\kappa) \qquad \Rightarrow \qquad m(\kappa) = e^{-\kappa}\cdot\left(e^{i\kappa}-1\right)^{-i} \quad .$$

For the solution including initial conditions we have I_0 infected at time t_0

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

giving the initial condition of the characteristic function

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

With another function $\Phi(z)$ in the separation ansatz we include the intial conditions via $z(\kappa,t) = m(\kappa)\cdot n(t)$

$$g(\kappa,t) = h(\kappa)\cdot\Phi(z) = h(\kappa)\cdot\Phi(\ell(\kappa,t)) \quad .$$

giving

$$\Phi(z) = \left(1-z^i e^{\beta^* t_0}\right)^{N-I_0}$$

and the characteristic function has the solution

$$g(\kappa,t) = e^{i\kappa N}\cdot\left(e^{-i\kappa}e^{-\beta^*(t-t_0)} + (1-e^{-\beta^*(t-t_0)})\right)^{N-I_0} \tag{10}$$

with $p(I,t) = \frac{1}{N+1}\sum_{k=0}^{N} e^{-i\frac{2\pi}{N+1}k\cdot I}\cdot g(\kappa(k),t)$ giving the probability solution

$$p(I,t) = \begin{pmatrix} N-I_0 \\ I-I_0 \end{pmatrix} \left(e^{-\beta^*(t-t_0)}\right)^{N-I} \left(1-e^{-\beta^*(t-t_0)}\right)^{I-I_0} = p(I,t|I_0,t_0) \tag{11}$$

with the transition probability $p(I,t|I_0,t_0)$ as used below in the likelihood function.

2.4. Likelihood function and its maximization

For a data set $(I_0, I_1, ...I_n)$ we have the joint probability given by the product of the above calculated transition probabilities

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

Inserting this and omitting the star from β^* for notational ease we obtain the joint probability of all data points as a function of model parameters, here the parameter β, the likelihood function

$$L(\beta) = \prod_{\nu=0}^{n-1} \begin{pmatrix} N-I_\nu \\ I_{\nu+1}-I_\nu \end{pmatrix} \left(e^{-\beta\Delta t}\right)^{N-I_{\nu+1}} \left(1-e^{-\beta\Delta t}\right)^{I_{\nu+1}-I_\nu} \cdot p(I_0,t_0)$$

with $\Delta t := t_{\nu+1} - t_\nu$. The most likely parameter is the the maximum of $L(\beta)$, or hence also its logarithm $l(\beta) := ln(L(\beta))$, with the result

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

from the data I_ν.

2.5. Bayesian probabilities for parameters given the data

Of course, one wants to know the probability of the paramters given the data at hand, here $p(\beta|\underline{I})$, conditioned on the data set $\underline{I} = (I_0, I_1, ...I_n)$, in order to evaluate the parameter estimation and the information the data provide. The likelihood as constructed above is however the probability of the data given the parameters, hence $L(\beta) = p(\underline{I}|\beta)$, and not a normalizable probability in respect the the parameters. Considering the joint probability of data and parameters and conditioning on one or the other gives us the wanted probability of the parameters given the data $p(\beta|\underline{I})$ from the likelihood as obtained directly from the data $p(\underline{I}|\beta)$ via the Bayesian ansatz

$$p(\beta|\underline{I}) = \frac{p(\underline{I}|\beta)}{p(\underline{I})}\, p(\beta) \quad , \tag{12}$$

with the normalization constant $p(\underline{I}) = \int p(\underline{I}|\beta) \cdot p(\beta)\, d\beta$. See e.g. [19], [20] for a general introduction to the Bayesian framework.

The introduction of an a priori probability of the parameter $p(\beta)$ occasionally leads to some discussion between frequentists and Bayesians, since there is no real criterion how to chose this, and leaves some arbitrariness to the result $p(\beta|\underline{I})$ of the posterior, in case the prior is chosen narrowly. However, a braod prior is mostly overwritten by the data, which give much larger terms than the prior, hence a fair choice of relatively uninformed priors over the hopefully well informed and much narrower posterior is desirable [19]. For analytic convenience often conjugate priors are used, having the same functional form as the likelihood function, whereas in computaional methods often a flat prior, on the costs of a cut-off far outside the parameter region of interest is used, for normalization purposes.

The conjugate prior for the linear infection model is a beta-distribution with prior parameters a_1 and b_1, which should be specified beforehand and give uninformed priors. Then for the posterior we obtain the analytical expression

$$\begin{aligned} p(\beta|\underline{I}) \quad = \quad & \frac{\Gamma(a_1 + b_1 + \sum_{\nu=0}^{n-1}(N - I_\nu))}{\Gamma(a_1 + \sum_{\nu=0}^{n-1}(I_{\nu+1} - I_\nu))\, \Gamma(b_1 + \sum_{\nu=0}^{n-1}(N - I_{\nu+1}))} \\ & \cdot \left(1 - e^{-\beta\Delta t}\right)^{a_1 + \sum_{\nu=0}^{n-1}(I_{\nu+1} - I_\nu) - 1} \left(e^{-\beta\Delta t}\right)^{b_1 + \sum_{\nu=0}^{n-1}(N - I_{\nu+1}) - 1} \cdot e^{-\beta\Delta t} \cdot \Delta t \quad , \end{aligned} \tag{13}$$

with the detailed calculations given e.g. in [9].

The posterior is again a beta-distribution with parameters $\tilde{a} = a_1 + \sum_{\nu=0}^{n-1}(I_{\nu+1} - I_\nu)$ and $\tilde{b} = b_1 + \sum_{\nu=0}^{n-1}(N - I_{\nu+1})$, the hyperparameters with the large terms of the data overwriting the small prior parameters. For a further investigation of prior versus posterior performance see [21], and especially Fig. 3 therein.

Then the median or the mode of the posterior is taken as best estimate and the confidence intervals can be read off directly from the posterior distribution $p(\beta|\underline{I})$. Now we will investigate in how far a model is suitable at all to describe the data at hand, or more specifically if other models can describe the data better or worse than the presently considered model, hence we search now for probabilities of one model, given the data, in comparison to other models, given the same data.

2.6. Model comparison

Often in life science different research groups are working on the same data or similar data but with different models, often parametrized quite differently and emphasizing one or another feature of the dynamics of the system. Now in times of COVID-19 this happens more than ever, and even on a global scale. Hence to address this question of model comparison in the present Bayesian framework we look at another model with which we can compare the one we have been working with up to now. We consider now a further simplification of the linear infection model, i.e. for large N we have $\beta^*(N - (I - 1)) \approx \beta^*(N - I) \approx \beta^* N =: \lambda$ which is constant, and so the master equation reduces to the form

$$\frac{d}{dt}p(I,t) \quad = \quad \lambda \cdot p(I-1,t) - \lambda \cdot p(I,t) \tag{14}$$

now with constant transition probabilities λ. This describes the Poisson process and with initial conditions $p(I,t_0)=\delta_{I,I_0}$ has the solution, see [9] for the detailed calculations,

$$p(I,t)=\frac{(\lambda\cdot\Delta t)^{I-I_0}}{(I-I_0)!}\,e^{-\lambda\cdot\Delta t} \tag{15}$$

giving also the transition probability $p(I,t|I_0,t_0)$ needed for the likelihood function with the result

$$L(\lambda)=\prod_{\nu=0}^{n-1}\frac{(\lambda\Delta t)^{I_{\nu+1}-I_\nu}}{(I_{\nu+1}-I_\nu)!}e^{-\lambda\Delta t} \quad . \tag{16}$$

For the Poisson model the conjugate prior is the gamma-distribution with prior parameters a_2 and b_2 and we obtain

$$p(\lambda|\underline{I})=\frac{(b_2+n)^{a_2+k_2}}{\Gamma(a_2+k_2)}\,(\lambda\Delta t)^{(a_2+k_2-1)}e^{-\lambda\Delta t(b_2+n)}\Delta t \quad . \tag{17}$$

for the posterior.

Consider now the linear infection model and the Poisson model, for a given data set $\underline{I}$, which we obtain from Gillespie simulations from the more complex linear infection model. We will refered to the linear infection model M_1 and to the Poisson model as M_2. With same probability initially for both models $p(M_1)=p(M_2)=1/2$ the Bayes factor k gives the ratio of probability of model M_1 over model M_2, both conditioned on the same data set

$$k:=\frac{p(M_1|\underline{I})}{p(M_2|\underline{I})}=\frac{\frac{p(\underline{I}|M_1)}{p(\underline{I})}\cdot p(M_1)}{\frac{p(\underline{I}|M_2)}{p(\underline{I})}\cdot p(M_2)}=\frac{p(\underline{I}|M_1)}{p(\underline{I}|M_2)} \quad . \tag{18}$$

Computing the Bayesian evidence of models M_1 and M_2 by integrating over all parameters, which we already obtained via the normalization constants in each model individually, here $p(\underline{I}|M_1)=\int p(\underline{I}|\beta,M_1)\cdot p(\beta)\,d\beta$ for model M_1 and respectively $p(\underline{I})=\int p(\underline{I}|\lambda,M_2)\cdot p(\lambda)\,d\lambda$ for M_2, we get

$$p(\underline{I}|M_1)=k_1\cdot\frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)}\cdot\frac{\Gamma(a_1+k_2)\Gamma(b_1+k_3)}{\Gamma(a_1+k_2+b_1+k_3)} \tag{19}$$

and

$$p(\underline{I}|M_2)=k_4\cdot\frac{b_2^{a_2}}{\Gamma(a_2)}\cdot\frac{\Gamma(a_2+k_2)}{(b_2+n)^{a_2+k_2}} \tag{20}$$

and for Bayes factor

$$k=\frac{k_1\cdot\Gamma(a_1+b_1)\cdot\Gamma(a_1+k_2)\cdot\Gamma(b_1+k_3)\cdot\Gamma(a_2)\cdot(b_2+n)^{a_2+k_2}}{k_4\cdot\Gamma(a_1)\cdot\Gamma(b_1)\cdot\Gamma(a_1+k_2+b_1+k_3)\cdot\Gamma(a_2+k_2)\cdot b_2^{a_2}} \tag{21}$$

with constants which include the data information, i.e. $k_1:=\left(\prod_{\nu=0}^{n-1}\begin{pmatrix}N-I_\nu\\ I_{\nu+1}-I_\nu\end{pmatrix}\right)$ and further $k_2:=\sum_{\nu=0}^{n-1}(I_{\nu+1}-I_\nu)$, $k_3:=\sum_{\nu=0}^{n-1}(N-I_{\nu+1})$ and $k_4:=\prod_{\nu=0}^{n-1}\frac{1}{(I_{\nu+1}-I_\nu)!}$.

We evaluated the Bayes factor for a data set simulated with the linear infection model M_1. However, in some parameter regions often the simpler Poisson model appears to be more likely, i.e., has higher probability or evidence than the truly underlying linear infection model, see [14] for more details on various data sets generated from model M_1. A Bayes factor of $k>1$ indicates a higher probability of model M_1 than the probability for model M_2 while a Bayes factor smaller than 1, hence $k<1$, would indicate a higher probability for the simpler model M_2, though the data set was originally simulated via model M_1. Sometimes a value of k higher than 10 is considered as selection of model M_1, but there is no strict border. Still, if a Bayes factor is above 100, model M_1 is 100 times more likely than M_2 etc., hence the unfavoured model can be considered as inappropriate to describe the data set under investigation.

The Bayes factor was then computed for many realizations of the process and for different values of β. Fig. 1 shows the variation of the logarithm of k with β. We see that in some parameter regions the Bayes factor favours the model not underlying the simulation. This happens in parameter regions which do not easily give

sufficient information to the simulated data to reject the simpler model M_2. This is a common situation in empirical systems with limited available data. As expected, the underlying model is clearly selected if the parameter β is increased.

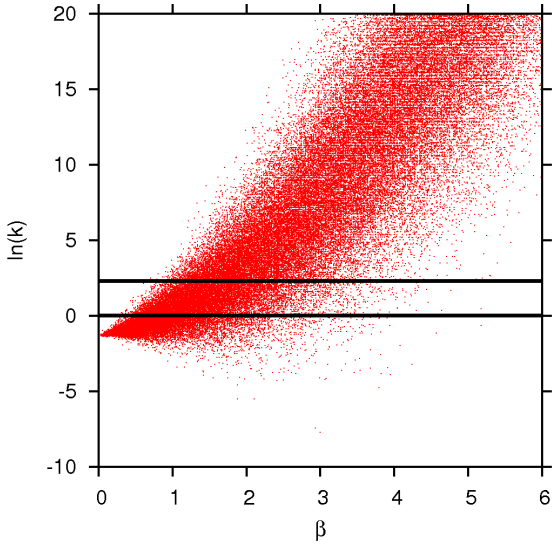


Figure 1: Logarithm of the Bayes factor for various values of β. Lines indicate $k = 1$, hence $ln(k) = 0$ and a value given often as theshold to distinguish one model from the other of $k = 10$, which however is not really a good choice. Clear cases are given for much larger Bayes factors only. For β-values up to $\beta \approx 1$ the simpler model has good chances to decribe the data, often even better than the linear infection model. From $\beta \approx 3$ on the linear infection model would be near to exclusively been chosen. In the intermediate region, often it depends on the realization of the data, which of the two models has better descriptive value.

In conclusion the present models, and especially the here presented aspect of model selection in the Bayesian framework, can be applied to influenza data, namely those from InfluenzaNet, an internet based surveillance system for influenza like illnesses which has been extended to various countries in Europe, see [22]. Another application is in dengue fever where more extended models have been suggested with potentially selective features which might be detectable with presently available empirical data or with future data, see [23], [24], [25], [3] and references therein. When using such more complex models, however, most of the steps which could be treated here analytically will have to be performed via simulations, see [21]. We will demonstrate the numerical aspects of model evaluation and model selection on SHAR and effective SIR models in the next section.

3. Dengue fever toy models of SHAR-type and effective SIR models

3.1. Introduction

In this section we derive stochastic versions of epidemiological models with severe and asymptomatic infection, which have previously been described as differential equation systems [1] for SHAR models. We fit data output of these models with simpler SIR type models to obtain numerically effective infection rates. The results can be compared with the previously stated simpler analytical arguments for the relation between effective infection rate in the SIR models and more complex parameter combinations in the SHAR models [1]. Namely, the relation between effective infection rate in the SIR system and the parameters of the SHAR model is confirmed by numerically evaluating the likelihoods of initial infection rate of the SHAR model and the effective infection rate of the SIR system, even outside stationarity of the respective processes, from which the analytical results are obtained. Similar approaches as we use here for the numerical calulation of likelihoods and Bayes factors have been used e.g. in [26], [27]. In our case study it turns out that the SHAR model, from which the toy data were generated, appears to be not much more likely than the effective SIR

model in terms of Bayes factors. For alternative approaches of model selection, based on the Kolmogorov-Smirnov test, see e.g. [28]. However, such approaches are not as easily generalizable as the Bayes factor, where simply probabilities are given to models based on data.

The present study of such initial models is well applicable to the situation of dengue fever infections in which severity of disease is associated with antibody dependent enhancement (ADE) since a long time [29], [30]. And recently field studies tried to clarify the question if asymptomatically infected are equally infectious as severe cases with their high viral load [31]. These authors come to the conclusion that asymptomatically infected can infect as much mosquitoes as severely infected, and there is the expectation that due to their mobility, asymptomatically infected might even contribute more to the force of infection than severe cases.

However, the best notification data of dengue infections from Thailand use only severe cases being hospitalized [6], which means that in relatively simple models which already can describe the empirical data well we would expect a "masked" effective infection rate as suitable rather than a "bare" infection rate, however this might be defined. Many present models of dengue fever have so complicated structures that any systematic analysis in terms of Bayes factors would most likely prefer the simpler models due to the limited number of empirical data. However, such effective models are the best tools for realistic predictions of future disease levels [32], and even more in analysis of control measures like vaccination and its impact [33], [34], [35]. In the present study we concentrate on rather simplified models like the SHAR and SIR model without any serotype interactions, which would be needed to describe the empirical data more accurately [3], [36], [37]. Since such multi-strain models show not only oscillations into stationary states, and eventually stabilized by noise as we observe here in our simple models, but chaotic attractors, we will leave the study of such models with the present techniques for future research. For chaotic systems up to now mostly random walks toward the likelihood maximum are in technical reach, see [21] including further references.

3.2. The SHAR model

The SHAR model with severe infection or hospitalized cases H and mild or asymptomatic cases A in an otherwise SIR-type epidemiological model, see [9] for more information on the notations and standart analysis techniques, is given by

$$\begin{aligned} \frac{d}{dt}S &= \alpha R - \frac{\beta}{N} S(A+\phi H) \\ \frac{d}{dt}H &= \eta \frac{\beta}{N} S(A+\phi H) - \gamma H \\ \frac{d}{dt}A &= (1-\eta)\frac{\beta}{N} S(A+\phi H) - \gamma A \\ \frac{d}{dt}R &= \gamma(A+H) - \alpha R \end{aligned} \tag{22}$$

and $R = N - S - A - H$. Here η is the ratio of infection leading to severe disease cases and ϕ is the change of infectivity of severe cases versus the "natural" infectivity via asymptomatic cases.

The role of asymptomatic infection becomes more recognized recently, since many pathogens appear as only accidentally pathogenic, see e.g. [9], [43] initially for bacterial meningitis [7], [10], [8], with fluctuations similar to crossing a vaccination threshold [44]. Similarly, in dengue fever the virus is completely asymptomatic in its animal reservoir, monkeys, and only turns occasionally pathogenic in humans, mostly in secondary infections. The SHAR model is an oversimplified model for dengue fever in the sense that higher viral load due to antibody dependent enhancement, ADE [29], [30], or lower infectivity due to hospitalization gives a ϕ different from unity, and the majority of primary infections is asymptomatic, mild or sub-clinical with ratio $(1-\eta)$, but here we do not consider primary versus secondary infection nor any further serotype interaction as done in [3], [37].

3.2.1. The stochastic SHAR model

We now want to generate "toy data" via a stochastic version of the SHAR model via the Gillespie algorithm [38], [39] from a master equation formulation [16], [17], [18], see in application to epidemiology explicitly

[9] including further references, with transitions given by the following reaction scheme

$$\begin{array}{ccc} S+A & \stackrel{\eta\beta}{\longrightarrow} & H+A \\ \text{„} & \stackrel{(1-\eta)\beta}{\longrightarrow} & A+A \\ S+H & \stackrel{\phi\eta\beta}{\longrightarrow} & H+H \\ \text{„} & \stackrel{\phi(1-\eta)\beta}{\longrightarrow} & A+H \\ H & \stackrel{\gamma}{\longrightarrow} & R \\ A & \stackrel{\gamma}{\longrightarrow} & R \\ R & \stackrel{\alpha}{\longrightarrow} & S \end{array} \tag{23}$$

giving the master equation for the dynamics of the probabilities

$$\begin{aligned} \frac{d}{dt}p(S,H,A,t) &= \eta\frac{\beta}{N}(S+1)A\; p(S+1,H-1,A,t) \\ &+(1-\eta)\frac{\beta}{N}(S+1)(A-1)\; p(S+1,H,A-1,t) \\ &+\eta\frac{b}{N}(S+1)\phi(H-1)\; p(S+1,H-1,A,t) \\ &+(1-\eta)\frac{\beta}{N}(S+1)\phi H\; p(S+1,H,A-1,t) \\ &+\gamma(A+1)\; p(S,H,A+1,t) \\ &+\gamma(H+1)p(S,H+1,A,t) \\ &+\alpha(N-(S-1)-H-A)\; p(S-1,H,A,t) \\ &-\left(\frac{\beta}{N}S(A+\phi H)+\gamma(A+H)+\alpha(N-S-H-A)\right)\; p(S,H,A,t) \quad . \end{aligned} \tag{24}$$

The master equation can be written in a generic form using densities $x_1 := S/N$, $x_2 := H/N$ and $x_3 := A/N$ instead of the total numbers of individuals in the population classes S, H and A, hence state vector $\underline{x} := (x_1, x_2, x_3)^{tr}$, as

$$\frac{d}{dt}\; p(\underline{x},t) \;=\; \sum_{j=1}^{n}\Big(\; Nw_j(\underline{x}+\Delta\underline{x}_j)\cdot p(\underline{x}+\Delta\underline{x}_j,t) \;\; - Nw_j(\underline{x})\cdot p(\underline{x},t)\Big) \tag{25}$$

with $n = 5$ different transitions and small deviation from state $\underline{x}$ as $\Delta\underline{x}_j := \frac{1}{N}\cdot\underline{r}_j$, see [40], [41], [42].

For the SHAR model we have explicitly the following transitions $w_j(\underline{x})$ and its vectors $\underline{r}_j$ given by

$$\begin{array}{lll} w_1(\underline{x}) = \eta\beta x_1(x_3+\phi x_2) & , & \underline{r}_1 = (1,-1,0)^{tr} \\ w_2(\underline{x}) = (1-\eta)\beta x_1(x_3+\phi x_2) & , & \underline{r}_2 = (1,0,-1)^{tr} \\ w_3(\underline{x}) = \gamma x_2 & , & \underline{r}_3 = (0,1,0)^{tr} \\ w_4(\underline{x}) = \gamma x_3 & , & \underline{r}_4 = (0,0,1)^{tr} \\ w_5(\underline{x}) = \alpha(1-x_1-x_2-x_3) & , & \underline{r}_5 = (-1,0,0)^{tr} \quad . \end{array} \tag{26}$$

With these $w_j(\underline{x})$ and $\underline{r}_j$ specified we also can express the mean field ODE system and in Kramers-Moyal approximation of the master equation to a Fokker-Planck equation a stochastic differential equation system, as will be shown below. From the SHAR model given as master equation we can now simulate realizations of the stochastic process via the Gillespie algorithm, and in this way obtain a "toy data set"

$$\underline{D} := (H_1, H_2, ..., H_{n_d}) \tag{27}$$

of n_d data points of hospital cases H_ν at times t_ν. On this data set we can then test statistical methods to fit parameters of various models including model comparison in a Bayesian framework, as described for the analytically treatable case in the previous section. We will now try to describe the output of the SHAR model $\underline{D}$ via an effective simpler model, the SIR model with new infection rate $\tilde{\beta}$.

3.3. The effective SIR model

An effective SIR model tries to explain the observed number of infected via a simple infection class I without any further distinction into more complex classes like asymtomatically infected, severe cases or any other possible mechanisms like primary or secondary infection etc. The SIR model is given by

$$\begin{aligned} \frac{d}{dt}S &= \alpha R - \frac{\tilde{\beta}}{N}SI \\ \frac{d}{dt}I &= \frac{\tilde{\beta}}{N}SI - \gamma I \\ \frac{d}{dt}R &= \gamma I - \alpha R \end{aligned} \tag{28}$$

with an initially unknown infection rate $\tilde{\beta}$ to be determined either from data as we will do below, or as was done in [1], by assuming that the effective SIR system has the same stationary state value I^* as a more complex model, here the SHAR model with observed severe infected H^*.

3.3.1. Effective infection rate

In [1] we obtained for the effective infection rate $\tilde{\beta}$ as a function of the parameters of the more complex model, the SHAR model, the expression

$$\tilde{\beta} = \frac{\gamma}{1-\eta\left(1-\frac{1}{1+(\phi-1)\eta}\cdot\frac{\gamma}{\beta}\right)} \tag{29}$$

from assuming the condition that the stationary state I^* of the effective SIR model is the same as the stationary state value of the observed severe cases H^* in the SHAR model, hence $I^*(\tilde{\beta}) = H^*(\beta, \eta, \phi)$. It turned out that also outside stationarity the deterministic simulations of the SHAR model and the effective SIR model with the above determined effective infection rate $\tilde{\beta}$ agree quite well, when using otherwise the same parameters and initial conditions, as far as the SIR system has corresponding parameters in the SHAR model.

Parameters for the initial study of the deterministic matching are given as follows: For the SHAR model we use $N = 1000$, $\gamma = 1$, $\beta = 3{\cdot}\gamma$, $\alpha = 0.1$, $\eta = 3/4$, and for the moment $\phi = 1$. Initial conditions are $S_0 = 200$, $A_0 = 30$, $H_0 = \frac{\eta}{1-\eta}A_0$, a relation which holds in stationarity $H^* = \frac{\eta}{1-\eta}A^*$, and $R_0 = N - S_0 - A_0 - H_0$. For the SIR system we use the same parameters if not otherwise stated, hence initially $\tilde{\beta} = \beta$, and as initial conditions $S_{0,SIR} = (\beta/\tilde{\beta})S_{0,SHAR}$, $I_0 = H_0$ and $R_0 = N - S_0 - I_0$. We use the analytically calculated effective infection rate Eq. (29), which turns with the present $\eta = 3/4$ etc. out to be $\tilde{\beta} = 2 \cdot \gamma$. We also changed the intial condition in the SHAR model to $A_0 = 100$ to observe the effect of initial conditions far away from the equilibrium, and still $I(t)$ is quite well comparable with $H(t)$ outside the condition $I^* = H^*$, see [2].

Next we look at the master equation formulation of the SHAR model to simulate toy data. The stochastic fluctuations of the SHAR model itself go well beyond the differences between the SHAR model and the effective SIR in their deterministic versions. In the present study we want to determine the effective infection rate $\tilde{\beta}$ by comparing stochastic simulations of the effective SIR model with the data $\underline{D}$ obtained from a stochastic simulation of the SHAR model. For this we vary values of $\tilde{\beta}$ and count the number of simulations being close to the data, and the $\tilde{\beta}$ value with maximal number of simulations close to the data is the best estimator of $\tilde{\beta}$ for this data set. Therefore, we now need the stochastic version of the effective SIR model, and best a version which is quick in simulation time, since we need many simulations with varying $\tilde{\beta}$ values.

3.3.2. Stochastic versions of the effective SIR model

The master equation for the effective SIR model is given by the reaction scheme

$$\begin{aligned} S + I &\xrightarrow{\tilde{\beta}} I + I \\ I &\xrightarrow{\gamma} R \\ R &\xrightarrow{\alpha} S \end{aligned} \tag{30}$$

and explicitly for the dynamics of the probabilities we get

$$\begin{aligned}\frac{d}{dt}p(S,I,t) &= \frac{\tilde{\beta}}{N}(S+1)(I-1)\,p(S+1,I-1,t)\\ &\quad +\gamma(I+1)p(S,I+1,t)\\ &\quad +\alpha(N-(S-1)-I)\,p(S-1,I,t)\\ &\quad -\left(\frac{\tilde{\beta}}{N}SI+\gamma I+\alpha(N-S-I)\right)\,p(S,I,t) \quad .\end{aligned} \tag{31}$$

For the master equation in densities $x_1 := S/N$ and $x_2 := I/N$ we have now transitions $w_j(\underline{x})$ and its vectors $\underline{r}_j$ given by

$$\begin{array}{lll} w_1(\underline{x}) = \tilde{\beta}x_1x_2 & , & \underline{r}_1 = (1,-1)^{tr} \\ w_2(\underline{x}) = \gamma x_2 & , & \underline{r}_2 = (0,1)^{tr} \\ w_3(\underline{x}) = \alpha(1-x_1-x_2) & , & \underline{r}_3 = (-1,0)^{tr} \quad . \end{array} \tag{32}$$

We could use now simulations of the master equation directly to measure the effective infection rate $\tilde{\beta}$, but for large population sizes N the simulations become very slow. So we will use the Kramers-Moyal expansion to obtain a Fokker-Planck equation as approximation of the master equation, for which the corresponding stochastic differential equation system can be simulated much faster.

3.4. Fokker-Planck approximation of the effective SIR model

From the master equation in densities, Eq. (25), now for the SIR system, we use Taylor's expansion

$$w_j(\underline{x}+\Delta\underline{x}_j)\cdot p(\underline{x}+\Delta\underline{x}_j,t) = \sum_{\nu=0}^{\infty}\frac{1}{\nu!}\left(\Delta\underline{x}_j\cdot\nabla_{\underline{x}}\right)^{\nu} w_j(\underline{x})\,p(\underline{x},t) \tag{33}$$

giving in Kramers-Moyal approximation to second order in $1/N$ a Fokker-Planck equation

$$\begin{aligned}\frac{\partial}{\partial t}\,p(\underline{x},t) &= -\nabla_{\underline{x}}\left(\sum_{j=1}^{n}(-\underline{r}_j\cdot w_j(\underline{x}))\,p(\underline{x},t)\right)\\ &\quad +\frac{\sigma^2}{2}\sum_{j=1}^{n}(\underline{r}_j\cdot\nabla_{\underline{x}})^2 w_j(\underline{x})\,p(\underline{x},t)\end{aligned} \tag{34}$$

with

$$\nabla_{\underline{x}} = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \end{pmatrix} = \partial_{\underline{x}} \tag{35}$$

or in different notation

$$\frac{\partial}{\partial t}\,p(\underline{x},t) = -\partial_{\underline{x}}\left(\underline{f}(\underline{x})\,p(\underline{x},t)\right) + \frac{\sigma^2}{2}\,\vec{\partial}_{\underline{x}}\left(G^2(\underline{x})\,p(\underline{x},t)\right)\overleftarrow{\partial}_{\underline{x}} \tag{36}$$

using simply a quadratic form $\vec{\partial}_{\underline{x}}\,(G^2(\underline{x})\,p(\underline{x},t))\,\overleftarrow{\partial}_{\underline{x}}$ here with

$$\vec{\partial}_{\underline{x}}\,(G^2\,p)\,\overleftarrow{\partial}_{\underline{x}} = \left(\frac{\partial}{\partial x_1},\frac{\partial}{\partial x_2}\right)\cdot\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22}\end{pmatrix}^2 p(\underline{x},t)\cdot\begin{pmatrix} \overleftarrow{\frac{\partial}{\partial x_1}} \\ \\ \overleftarrow{\frac{\partial}{\partial x_2}}\end{pmatrix} \tag{37}$$

and

$$\begin{aligned} \underline{f}(\underline{x}) &= \sum_{j=1}^{n} \underline{f}_j(\underline{x}) = \sum_{j=1}^{n} (-\underline{r}_j \cdot w_j(\underline{x})) \\ G^2(\underline{x}) &= \sum_{j=1}^{n} G_j^2(\underline{x}) = \sum_{j=1}^{n} \underline{r}_j \cdot \underline{r}_j^{tr} \; w_j(\underline{x}) \quad . \end{aligned} \tag{38}$$

The Fokker-Planck equation gives a stochastic differential equation system with $\sigma = 1/\sqrt{N}$ and in the SIR model the two dimensional Gaussian normal noise vector $\underline{\varepsilon}(t) = (\varepsilon_{x_1}(t), \varepsilon_{x_2}(t))^{tr}$ as

$$\frac{d}{dt}\underline{x} = \underline{f}(\underline{x}) + \sigma G(\underline{x}) \cdot \underline{\varepsilon}(t) \tag{39}$$

and using matrix square root from eigenvalue-eigenvector decomposition $G^2(\underline{x}) = T\Lambda T^{-1}$ as $G(\underline{x}) = T\sqrt{\Lambda}T^{tr}$ to be numerically implemented easily, and much faster than the Gillespie algorithm [38], [39] for the master equation. To further speed up the SDE simulations we can relax the condition of G being a quadratic matrix, and allow a matrix B with $BB^{tr} = G^2$ the expressions in B can become quite simple and fast to compute [45].

Now we can vary $\tilde{\beta}$ and compare in an η-ball method [46], [32] the simulations of the SIR model with the data vector $\underline{D}$, which gives the likelihood $p(\underline{D}|\tilde{\beta})$. The η as a radius of a vicinity around the data vector $\underline{D}$ should not be confused with one of the parameters in the SHAR model, which is also named η, a confusion easily avoided by the very different contexts in which they appear. In case of doubt, we will call η_{SHAR} the model parameter of the SHAR model, and η_1 the η-ball radius around the data, used to evaluate simulations from model M_1, and η_2 from model M_2.

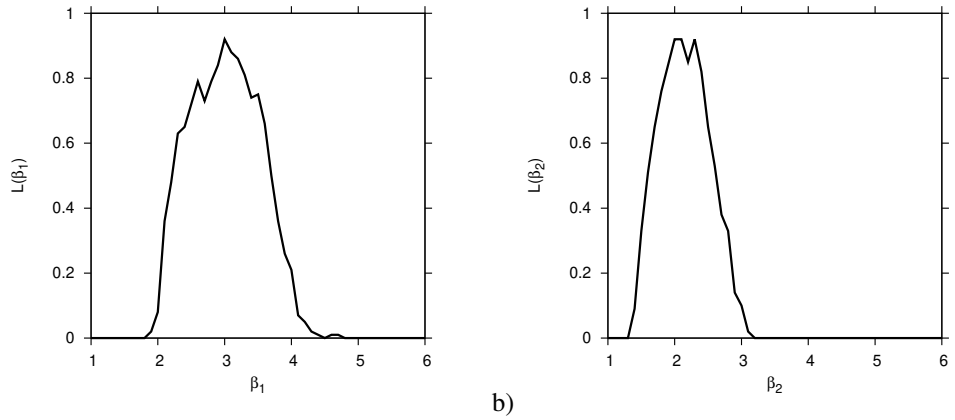


Figure 2: Likelihood functions numerically obtained for a) model M_1 with maximum in $\beta \approx 3$ and b) model $M_2$with maximum in $\beta \approx 2$ as expected from the analytical calculations of the effective $\tilde{\beta}$. We use in both cases the B-matrix version of the stochastic differential equation system to obtain the likelihood functions due to its simulation speed.

3.5. Data and analysis, results and conclusions

From the SHAR model we now generate data sets, hoping that the data already contain so much information about the system to find back parameters with not too large variance. Especially, we take a transient behaviour and a long period in stationarity covered by the time interval of the data. The data are taken by the slow but exact Gillespie algorithm.

Then we compare simulations of the SHAR model, with the data, where now the simulations are performed with the much faster stochastic differential equations obtained by the Fokker-Planck approximation of the master equation with its slow Gillespie algorithm for simulations. We find accurately the best estimate to be

where expected, at $\beta_1 = 3 \cdot \gamma$, and values like $\beta_1 = 2 \cdot \gamma$ or $\beta_1 = 4 \cdot \gamma$ have already quite lower likelihood, this in spite of using different algorithms for data generation and model fitting.

The next result is, when fitting the effective SIR model to the data generated from the SHAR model, again using the stochastic differential equation approach, we obtain a best estimate from the maximum of the likelihood for the effective infection rate $\beta_2 = \tilde{\beta} = 2 \cdot \gamma$, in good agreement with out analytical result, Eq. (29). This is a non-trivial result, since we obtained the analytic expression via comparing stationary states $I^* = H^*$ only, but now obtain the same relation from simulated data outside stationarity with a long and oscillating transient and continued stochastically stabilized oscillations in stationarity. Also remarkable is that for $\beta_2 = \tilde{\beta}$ the value of $3 \cdot \gamma$, which is the one used in the SHAR model, can be statistically excluded, since the likelihood funtion in Fig. 2 b) has values close to zero around $\beta_2 = 3 \cdot \gamma$ as compared to the maximal values near $\beta_2 = 2 \cdot \gamma$.

Finally, we investigate the Bayes factor $k = p(M_1|\underline{D})/p(M_2|\underline{D})$ for a data set $\underline{D}$ generated from M_1, the SHAR model. See for the concepts the previous section again on the Bayes factor and analytic results in simple cases. A Bayes factor of $k = 1$ would indicate equal probability of both models M_1, the SHAR model, and M_2, the effective SIR model, whereas $k \gg 1$ would give much higher probability of model M_1 over model M_2, that the data could lead us to reject model M_2 statistically. In our case, considering the Bayes factor as

$$k = \int \int p(M_1, \beta_1, \eta_{SHAR}|\underline{D})\ d\beta_1 d\eta_{SHAR} \Big/ \int p(M_2, \beta_2|\underline{D})\ d\beta_2 \tag{40}$$

due to the fact that we have to estimate in the SHAR model M_1 not only β_1 but also η_{SHAR} while in the SIR model M_2 we only need β_2 to estimate (neglecting for the moment all other possible insecurities) with a Bayes factor of around $k \approx 1.5$ hence the two models have about equal probability to describe the toy data set. But future studies have to be performed to investigate this point in more detail, including the question of ensembles of realizations and their Bayes factors, hence the fluctuations of the Bayes factor, as could be performed in the simpler analytically treatable case study shown in the previous section.

In the next section we will describe in more detail how to parametrize SHAR-type models, as recently very detailed data on COVID-19 appeared, distinguishing between different severities of the disease, as rarely seen in other diseases.

4. COVID-19 MODELS, SHARUCD

4.1. Introduction

The coronavirus disease 2019, COVID-19, was first identified in China [47] in December 2019, and human to human transmission enabled the virus, called the severe accute respiratory syndrom coronavirus type 2, SARS-CoV-2, to spread around the globe, with already in March 2020 affecting Europe severely, which by then reported more infected cases and death cases than anywhere else in the world outside China [48]. While case fatality ratios vary still a lot from country to country, a global case fatality ratio was reported at the end of April of about 7% [49], and still continues to be around 4 %, where the decrease is mainly explained by increasing testing capacities of infected worldwide.

Italy was the first country hard hit by the pandemic with national lockdown measures taken as soon as beginning of March 2020 [50], while a large increase in Spain followed around 8 days later with lockdown measures in mid March [51] and reinforcement of further measures at the end of March [52], [53]. Rather early on, the autonomous cummunity of the Basque Country, Euskadi, with a population of 2.2 million residents registered the first cases of infected, hospitalized, intensive care unit admitted cases and fatalities around 4th of March 2020 on, and declared a public health emergency before any other part of Spain [54], followed by further restrictions soon after [51], [52], [53], [55]. Only late in April some restrictions of public life started to be loosened up [56] in the Basque country and elsewhere in Spain. In many countries epidemiological evaluations of control startegies were discussed also with input from the academic world, see e.g. [57], [58], [59] besides many others.

For informative mathematical modelling, high quality data is needed and collaborations between modellers and public health authorities are essential [6], [60], [61], [62]. Hence in March 2020 the Basque Government

established multidisciplinary task forces, and among those the Basque Modelling Task Force, BMTF, was initiated to help the local public health managers and the government in evaluating the development of the outbreak in the Basque Country. Among other complementary modellings we developed a SHAR based modelling framework to describe the increasingly available data on positive tested, hospitalized, ICU admitted and deceased cases, and later also recovered, the SHARUCD models with various different versions capturing increasing information from the data and recent research on epidemiological and medical aspects of COVID-19 [63], [64], [65], [66], [67], [68], keeping in the spirit of the above mentioned model parsimony the models as simple as possible, with adjustments mainly given by the available empirical data from the Basque Country. In this way we could describe the exponential phase of the outbreak and continue the analysis into the phase when control measures took effects, mainly by adjusting and measuring momentary growth rates of the observables with then time dependently decreasing infection rates [12], [13].

We use, similar to what was described in the previous sections, stochastic SHARUCD models with apart from the already described variables S, H, A and R, now also ICU admitted U and deceased D, and further for comparison with cumulative cases also cumulative classes for hospitalized C_H, mildly infected C_A, ICU admitted C_U and recovered C_R, counting all incoming cases in the dynamical compartments but no outflows, and in the case of mild/asymptomatic also including a notification ratio.

For the qualitative analysis we also consider the mean field approximation as ordinary differential equation system, but keeping initially all freedom of varying dynamically all variables, including susceptibles, since discussions on eventually soon developing herd immunity were ignited in the research community rather early on, but we could not find any signs of this up to now in the modelling framework, even with large numbers of asymptomatic spreading. For the exponential growth phase we hence can assume that still nearly the whole population was susceptible, but have to be careful with stochastic differential equation systems as described above via Kramers-Moyal approximations of the Markov processes given by the master equation in the initial pre-exponential phase and when control measures took place, in both cases because of low numbers of cases, and now observed localized outbreaks, which are more prone to power law distributions than to Gaussian fluctuations around the mean field solutions [9]. This is because the mean field solutions tend to go to very low numbers while the stochastic models still notice the weak contraction to the mean field solution close to any threshold behaviour and show large fluctuations away from the mean field solution.

4.2. The basic model, SHARUCD model 1

Besides the transitions and variables already introduced in the basic SHAR models we have in addition here the disease induced mortality μ, which affects mainly the severely diseased cases, and ICU admission rate ν, both modelled here in the basic SHARUCD model, also refered to as SHARUCD model 1, as exits from the hospitalized class. As opposed to the above mentioned SHAR models, where e.g. in dengue fever only data on severe and hence hospitalized cases are available e.g. in Thailand [6], we now consider that part of the milder case are detected by testing, hence the number of positive tested infections is larger than the notified hospiltalized cases, the detection ratio denoted by ξ.

The stochastic version of the basic SHARUCD model 1 can be formulated through the master equation in the generic form described above using densities, see Eq. (25), now of all variables $x_1 := S/N$, $x_2 := H/N$, $x_3 := A/N$, $x_4 := R/N$, $x_5 := U/N$, $x_6 := C_H/N$, $x_7 := C_A/N$, $x_8 := C_U/N$ and $x_9 := D/N$ and $x_{10} := C_R/N$, hence state vector $\underline{x} := (x_1, ..., x_{10})^{tr}$, giving the dynamics for the probabilities $p(\underline{x}, t)$ again as

$$\frac{d}{dt}\, p(\underline{x},t) \quad = \quad \sum_{j=1}^{n} \Big(N w_j(\underline{x} + \Delta\underline{x}_j) \cdot p(\underline{x} + \Delta\underline{x}_j, t) \quad - N w_j(\underline{x}) \cdot p(\underline{x}, t) \Big) \tag{41}$$

with $n = 10$ different transitions $w_j(\underline{x})$, as described by the mechanisms above, and small deviation from state $\underline{x}$ as $\Delta\underline{x}_j := \frac{1}{N} \cdot \underline{r}_j$. For the basic SHARUCD model we have explicitly the following transitions $w_j(\underline{x})$

and its shifting vectors $\underline{r}_j$ given by

$$\begin{array}{lll}
w_1(\underline{x}) = \eta\beta x_1(x_2 + \phi x_3 + \varrho) & , & \underline{r}_1 = (1,-1,0,0,0,-1,0,0,0,0)^{tr} \\
w_2(\underline{x}) = \xi(1-\eta)\beta x_1(x_2 + \phi x_3 + \varrho) & , & \underline{r}_2 = (1,0,-1,0,0,0,-1,0,0,0)^{tr} \\
w_3(\underline{x}) = (1-\xi)(1-\eta)\beta x_1(x_2 + \phi x_3 + \varrho) & , & \underline{r}_3 = (1,0,-1,0,0,0,0,0,0,0)^{tr} \\
w_4(\underline{x}) = \gamma x_2 & , & \underline{r}_4 = (0,1,0,-1,0,0,0,0,0,-1)^{tr} \\
w_5(\underline{x}) = (1-\xi)\gamma x_3 & , & \underline{r}_5 = (0,0,1,-1,0,0,0,0,0,0)^{tr} \\
w_6(\underline{x}) = \gamma x_5 & , & \underline{r}_6 = (0,0,0,-1,1,0,0,0,0,-1)^{tr} \\
w_7(\underline{x}) = \nu x_2 & , & \underline{r}_7 = (0,1,0,0,-1,0,0,-1,0,0)^{tr} \\
w_8(\underline{x}) = \mu x_2 & , & \underline{r}_8 = (0,1,0,0,0,0,0,0,-1,0)^{tr} \\
w_9(\underline{x}) = \mu x_5 & , & \underline{r}_9 = (0,0,0,0,1,0,0,0,-1,0)^{tr} \\
w_{10}(\underline{x}) = \xi\gamma x_3 & , & \underline{r}_{10} = (0,0,1,-1,0,0,0,0,0,0,-1)^{tr} \quad .
\end{array} \tag{42}$$

With these $w_j(\underline{x})$ and $\underline{r}_j$ specified we also can express the mean field ODE system. The deterministic version of the model is given by a differential equation system for all classes, including the recording classes of cumulative cases C_H, C_A, C_R and C_U by

$$\begin{aligned}
\frac{d}{dt}S &= -\beta\frac{S}{N}(H + \phi A + \varrho N) \\
\frac{d}{dt}H &= \eta\beta\frac{S}{N}(H + \phi A + \varrho N) - (\gamma + \mu + \nu)H \\
\frac{d}{dt}A &= (1-\eta)\beta\frac{S}{N}(H + \phi A + \varrho N) - \gamma A \\
\frac{d}{dt}R &= \gamma(H + U + A) \\
\frac{d}{dt}U &= \nu H - (\gamma + \mu)U \\
\frac{d}{dt}C_H &= \eta\beta\frac{S}{N}(H + \phi A + \varrho N) \\
\frac{d}{dt}C_A &= \xi\cdot(1-\eta)\beta\frac{S}{N}(H + \phi A + \varrho N) \\
\frac{d}{dt}C_R &= \gamma(H + U + \xi A) \\
\frac{d}{dt}C_U &= \nu H \\
\frac{d}{dt}D &= \mu(H + U)
\end{aligned} \tag{43}$$

in complete form as mean field approximation, also obtainable as the drift part of the diffusion approximation from the master equation to the Fokker-Planck equation as described in detail above in the SHAR/SIR part. We give now a first analysis of the exponential phase of the epidemic, along the lines of our analysis of the COVID-19 data from the Basque Country [12], [13], and will then improve, based on the present analysis extended to the control measure effect phase, this basic SHARUCD model 1 to a new model, SHARUCD model 2, see [12].

For the presentation of the analytic results used in the SHARUCD model 1 and gave rise to the developement of the SHARUCD model 2 we will now first introduce a model with the main features of both models and which is still close to the SIR limiting case, but goes well beyond this limit. This is the SHAR-C model, which includes the SHAR part and the cumulative observables C_H, C_A and C_R, but considering the transitions to ICU admission and to deceased as negligible, hence $\mu,\ \nu \to 0$. The final limit of $\phi \to 1$ towards the SIR-limiting case will be considere only later, since the explicit consideration of $\phi \neq 1$ turned out to be important in analyzing differences between hospitalized and asymptomatic.

Here we show the complete analysis of this model for the first time, and analogously results can be obtained for the various versions of the SHARUCD models we developed over the time during the analysis of the unfolding epidemic. The corresponding results for the here presented two versions of SHARUCD model 1 and 2 will be given as needed during the description of the model development and improvement.

4.3. Analytical calculation of growth rate and reproduction ratio from the active infection classes H and A in the SHAR-C model

In this subsection we summarize some basic notions of epidemiological spreading, using the SIR model as guiding tool, and then analyse the dynamically active compartments H and A of the SHAR-type models in more detail. We concentrate here on the formulation of the SHAR-C model as it has all common features of the SHARUCD model 1 and model 2.

The dynamic equations for the SHAR model in mean field approximation, as obtained from the full SHARUCD models in the limit of small disease induced mortality $\mu \to 0$ and small ICU admission ratio $\nu \to 0$ and not considering for the moment any import ϱ, are given by

$$\begin{aligned} \frac{d}{dt}S &= -\beta\frac{S}{N}(H+\phi A) \\ \frac{d}{dt}H &= \eta\beta\frac{S}{N}(H+\phi A) - \gamma H \\ \frac{d}{dt}A &= (1-\eta)\beta\frac{S}{N}(H+\phi A) - \gamma A \\ \frac{d}{dt}R &= \gamma(H+A) \end{aligned} \tag{44}$$

and we also can calculate the cumulative cases for hospitalized C_H, and asymptomatic C_A for comparison with the data of hospitalized and of positive infected as $I_{cum} := C_H + C_A$, and finally the at some point also available cumulative notified recovered C_R

$$\begin{aligned} \frac{d}{dt}C_H &= \eta\beta\frac{S}{N}(H+\phi A) \\ \frac{d}{dt}C_A &= \xi\cdot(1-\eta)\beta\frac{S}{N}(H+\phi A) \\ \frac{d}{dt}C_R &= \gamma(H+\xi A) \end{aligned} \tag{45}$$

completing the SHAR-C model.

4.3.1. Growth rate

After an introductory phase, which we will analyse in more detail later, the epidemic entered into an exponential growth phase with all observables with essentially the same growth rate, which started in the Basque Country around March 10, 2020, and due to the effects of the imposed control measures has left the exponential growth phase to a slower growth around March 27, 2020. This exponential growth phase is typical for any outbreak with disease spreading in a completely susceptible population, as observed already in the SIR-system, Eq. (28) with omitting the tilde above the infection rate β, from the dynamics of the infected $\frac{dI}{dt} = \left(\beta\frac{S}{N} - \gamma\right)\cdot I$ when $S(t) \approx N$, such that a linear differential equation $\frac{dI}{dt} = (\beta-\gamma)\cdot I =: \lambda\cdot I$ with an exponential growth factor λ is obtained. This growth factor then can be measured again from disease data via $\lambda = \frac{1}{I}\cdot\frac{dI}{dt} = \frac{d}{dt}\,ln(I)$ giving a straight line in a semi-logarithmic plot of the data, see Fig. 3.

For larger compartmental models we obtain similarly an exponential growth factor. For the present basic SHAR-C model we have the active disease classes H and A with the dynamics given by

$$\frac{d}{dt}\begin{pmatrix} H \\ A \end{pmatrix} = \left[\begin{pmatrix} \eta\beta\frac{S}{N} & \phi\eta\beta\frac{S}{N} \\ (1-\eta)\beta\frac{S}{N} & \phi(1-\eta)\beta\frac{S}{N} \end{pmatrix} - \begin{pmatrix} \gamma & 0 \\ 0 & \gamma \end{pmatrix}\right]\cdot\begin{pmatrix} H \\ A \end{pmatrix} \quad . \tag{46}$$

For an epidemic in its initial phase $S(t) \approx N$ we now have constant matrices

$$B = \beta\begin{pmatrix} \eta & \phi\eta \\ (1-\eta) & \phi(1-\eta) \end{pmatrix} \tag{47}$$

for entries into the disease classes and

$$G = \begin{pmatrix} \gamma & 0 \\ 0 & \gamma \end{pmatrix} \tag{48}$$

for exits from the disease classes, where we had infection rate β and recovery rate γ in the SIR case.

With $\underline{x} = (H, A)^{tr}$ we now have with $J = B - G$ the dynamics $\frac{d}{dt}\underline{x} = J\underline{x}$ and its solution via eigenvalue/eigenvector decomposition $JT = T\Lambda$ of the matrix J

$$\underline{x}(t) = Te^{\Lambda(t-t_0)}T^{-1}\underline{x}(t_0) \tag{49}$$

with matrix exponential including the eigenvalue matrix Λ and the transformation matrix $T = (\underline{u}_1, \underline{u}_2)$ from the eigenvectors $\underline{u}_i$ of matrix J. The eigenvalues of the matrix J are given by

$$\lambda_{1/2} = \frac{1}{2}\beta \cdot ((1-\eta)\phi + \eta) - \gamma \pm \frac{1}{2}\beta \cdot ((1-\eta)\phi + \eta) \tag{50}$$

The dominating growth factor is given by the largest eigenvalue $\lambda = \lambda_1$. Hence the eigenvalues are explicitly for the SHAR-C model

$$\lambda_1 = \beta \cdot ((1-\eta)\phi + \eta) - \gamma \tag{51}$$

and

$$\lambda_2 = -\gamma \quad . \tag{52}$$

For the limiting cases of infected being hospitalized, $\eta = 1$, we have

$$\lambda_1 = \beta - \gamma \tag{53}$$

and for zero hospitalization ratio, $\eta = 0$,

$$\lambda_1 = \beta\phi - \gamma \tag{54}$$

hence for $\phi > 1$ we have a larger growths for small hospitalization/severity η than for large η, hence a competition of asymtopmatic/mild for better spreading than the hospitalized. So when we consider for the future mutations, the less pathogenic strains will spread faster and eventually outcompete the overly harmful ones, making new harmful mutations only short-lived accidental pathogens [7], [8], [9], [10].

4.3.2. Complete solution of the dynamic compartments H and A

To calculate the complete time dependent solution of the dynamic HA-system $\frac{d}{dt}\underline{x} = J\underline{x}$ with $\underline{x}(t) = (H(t), A(t))^{tr}$, which is important for the transitory phase called "pre-exponential", since not all vaiables follow initially the same exponential growth with the same growth factor,

$$\underline{x}(t) = e^{J(t-t_0)}\underline{x}(t_0) = Te^{\Lambda(t-t_0)}T^{-1}\underline{x}(t_0) \tag{55}$$

with matrix J, which is now given by

$$J = \begin{pmatrix} \eta\beta - \gamma & \phi\eta\beta \\ (1-\eta)\beta & \phi(1-\eta)\beta - \gamma \end{pmatrix} =: \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{56}$$

we need to calculate the eigenvector matrix T with $JT = T\Lambda$. With the eigenvalues $\lambda_1 = \beta\cdot((1-\eta)\phi + \eta) - \gamma$ and $\lambda_2 = -\gamma$ we have

$$T = \begin{pmatrix} \lambda_1 - a_{22} & \lambda_2 - a_{22} \\ a_{21} & a_{21} \end{pmatrix} = (1-\eta)\beta \begin{pmatrix} \frac{\eta}{(1-\eta)} & -\phi \\ 1 & 1 \end{pmatrix} \tag{57}$$

but then removing the constant in front and

$$\tilde{T} = \begin{pmatrix} c_1 \cdot \begin{pmatrix} u_{11} \\ u_{21} \end{pmatrix}, & c_2 \cdot \begin{pmatrix} u_{12} \\ u_{22} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \eta & -\phi \\ 1-\eta & 1 \end{pmatrix} \tag{58}$$

with convenient factors $c_1 := 1 - \eta$ and $c_2 := 1$ give equally valid transformations, diagonalizing J. Then $\tilde{T}$ has the inverse with $det(\tilde{T}) = ((1-\eta)\phi + \eta) =: \kappa$

$$\tilde{T}^{-1} = \frac{1}{(1-\eta)\phi + \eta} \begin{pmatrix} 1 & \phi \\ \eta - 1 & \eta \end{pmatrix} \quad . \tag{59}$$

Hence we have

$$\begin{aligned}\begin{pmatrix} H(t) \\ A(t) \end{pmatrix} &= \tilde{T} e^{\Lambda(t-t_0)} \tilde{T}^{-1} \begin{pmatrix} H_0 \\ A_0 \end{pmatrix} \\ &= \frac{1}{(1-\eta)\phi+\eta} \begin{pmatrix} \eta & -\phi \\ 1-\eta & 1 \end{pmatrix} \begin{pmatrix} e^{\lambda_1(t-t_0)} & 0 \\ 0 & e^{\lambda_2(t-t_0)} \end{pmatrix} \cdot \begin{pmatrix} 1 & \phi \\ \eta-1 & \eta \end{pmatrix} \begin{pmatrix} H_0 \\ A_0 \end{pmatrix} \\ &= \frac{1}{(1-\eta)\phi+\eta} \begin{pmatrix} \eta e^{\lambda_1(t-t_0)} & -\phi e^{\lambda_2(t-t_0)} \\ (1-\eta)\, e^{\lambda_1(t-t_0)} & e^{\lambda_2(t-t_0)} \end{pmatrix} \cdot \begin{pmatrix} H_0+\phi A_0 \\ (\eta-1)\, H_0+\eta A_0 \end{pmatrix}\end{aligned}$$

with the result now for the composition of the exponentials of both eigenvalues

$$\begin{pmatrix} H(t) \\ A(t) \end{pmatrix} = \frac{1}{(1-\eta)\phi+\eta} \left[\begin{pmatrix} \eta(H_0+\phi A_0) \\ (1-\eta)(H_0+\phi A_0 \end{pmatrix} \cdot e^{\lambda_1(t-t_0)} + \begin{pmatrix} (1-\eta)\phi H_0 - \eta\phi A_0) \\ \eta A_0 - (1-\eta) H_0 \end{pmatrix} \cdot e^{\lambda_2(t-t_0)} \right] \tag{60}$$

as explicit result for the general solution of the form

$$\begin{pmatrix} H(t) \\ A(t) \end{pmatrix} = \begin{pmatrix} K_{H1} \\ K_{A1} \end{pmatrix} \cdot e^{\lambda_1(t-t_0)} + \begin{pmatrix} K_{H2} \\ K_{A2} \end{pmatrix} \cdot e^{\lambda_2(t-t_0)} \quad . \tag{61}$$

For the time limiting cases the solution gives for the intial conditions at $t = t_0$

$$\begin{pmatrix} H(t_0) \\ A(t_0) \end{pmatrix} = \frac{1}{(1-\eta)\phi+\eta} \begin{pmatrix} (\eta+(1-\eta)\phi) H_0 \\ ((1-\eta)\phi+\eta) A_0 \end{pmatrix} = \begin{pmatrix} H_0 \\ A_0 \end{pmatrix} \tag{62}$$

as expected and for large times, hence formally for $t \to \infty$, long after the pre-exponential phase, where still both eigenvalues might have mattered,

$$\begin{pmatrix} H(t \to \infty) \\ A(t \to \infty) \end{pmatrix} = \frac{1}{(1-\eta)\phi+\eta} \begin{pmatrix} \eta(H_0+\phi A_0) \\ (1-\eta)(H_0+\phi A_0) \end{pmatrix} \cdot e^{\lambda_1(t-t_0)} \quad . \tag{63}$$

4.3.3. Solutions for cumulative cases from the general solution and explicit calculation for the SHAR-C model

With the general solution, Eq. (61), we can now also calculate cumlative cases, e.g. the cumulative hospitalized C_H for a general case close to the complete SHARUCD model via

$$\frac{d}{dt} C_H = \eta\beta(H + \phi A) \tag{64}$$

giving

$$C_H(t) - C_H(t_0) = \eta\beta \int_{t_0}^{t} (H(\tilde{t}) + \phi A(\tilde{t}))\, d\tilde{t} \tag{65}$$

with the result

$$C_H(t) = C_H(t_0) + \eta\beta(K_{H1} + \phi K_{A1}) \cdot \frac{1}{\lambda_1} \left(e^{\lambda_1(t-t_0)} - 1 \right) + \eta\beta(K_{H2} + \phi K_{A2}) \cdot \frac{1}{\lambda_2} \left(e^{\lambda_2(t-t_0)} - 1 \right) \tag{66}$$

giving the structure of the solution as

$$C_H(t) = C_H(t_0) + \eta\beta \cdot \kappa_1 \cdot f(\lambda_1) + \eta\beta \cdot \kappa_2 \cdot f(\lambda_2) \tag{67}$$

with $\kappa_i := (K_{Hi} + \phi K_{Ai})$ and $f(\lambda_i) := \frac{1}{\lambda_i} \left(e^{\lambda_i(t-t_0)} - 1 \right)$ for $i \in \{1, 2\}$ and similarly for other cumulative variables like e.g. C_A and C_R. It is interesting to observe that for large positive λ_1 e.g. the function $\frac{1}{\lambda_1} \left(e^{\lambda_1(t-t_0)} - 1 \right) \approx \frac{1}{\lambda_1} e^{\lambda_1(t-t_0)}$ the exponential phase approximation holds as we described above, always assuming $\lambda_1 \gg \lambda_2$. On the other hand in a controlled situation with negative growth $0 > \lambda_1 > \lambda_2$ we have due to the change of sign of $\lambda_1 = -|\lambda_1|$ the dynamic behaviour $\frac{1}{\lambda_1} \left(e^{\lambda_1(t-t_0)} - 1 \right) = \frac{1}{|\lambda_1|} (1 - e^{-|\lambda_1|(t-t_0)}) \approx \frac{1}{|\lambda_1|} = const.$, such that the cumulative cases become constant, not increasing any more, as to be expected.

For the present SHAR-C model we have explicitly given, with abbreviation $\kappa := (1-\eta)\phi + \eta$,

$$\begin{pmatrix} K_{H1} \\ K_{A1} \end{pmatrix} = \frac{1}{\kappa} \begin{pmatrix} \eta(H_0 + \phi A_0) \\ (1-\eta)(H_0 + \phi A_0) \end{pmatrix} \tag{68}$$

and

$$\begin{pmatrix} K_{H2} \\ K_{A2} \end{pmatrix} = \frac{1}{\kappa} \begin{pmatrix} (1-\eta)\phi H_0 - \eta A_0 \\ \eta A_0 - (1-\eta) H_0 \end{pmatrix} \tag{69}$$

such that we obtain

$$\kappa_1 = K_{H1} + \phi K_{A1} = H_0 + \phi A_0 \tag{70}$$

and

$$\kappa_2 = K_{H2} + \phi K_{A2} = \frac{1}{\kappa}\eta(\phi - 1)A_0 = \frac{\eta(\phi-1)}{(1-\eta)\phi + \eta} A_0 \quad . \tag{71}$$

Hence for the SIR-limiting case $\phi = 1$ with have $\kappa_2 = 0$, such that λ_2 has no influence on the cumulative cases. This reflects the fact that the SIR system has only one growth factor λ and not two different eigenvalues. So any chance to obtain information about λ_2 only is possible for $\phi \neq 1$.

For the cumulative mild/asymptomatic cases C_A with dynamics

$$\frac{d}{dt} C_A = \xi(1-\eta)\beta(H + \phi A) \tag{72}$$

taking explicitly a detection ratio ξ into account we have analogously to the calculations for C_H now the solution

$$C_A(t) = C_A(t_0) + \xi(1-\eta)\beta \cdot \kappa_1 \cdot f(\lambda_1) + \xi(1-\eta)\beta \cdot \kappa_2 \cdot f(\lambda_2) \tag{73}$$

with the same constants κ_i and eigenvalue dependent functions $f(\lambda_i)$ as in C_H. While for the initial condition of the cumulative hospitalized $C_H(t_0)$ we can start with zero as for the dynamically active prevalence value $H(t_0)$ at an initial stage of the epidemic, for $C_A(t_0)$ we have more possible choices, depending on the epidemiological situation. One choice is simply the proportion of detected to overall asymptomatic, hence $C_A(t_0) = \xi \cdot A_0$. On the other hand, if a public health system only starts to respond with testing when the first severe cases appear in hospital, hence at a time t_H, with $C_H(t_H) = 1$, we would have $C_A(t_0) = 0$ for any time $t < t_H$, hence no recording starting before the first severe case.

Then really recorded are the cumulative positive cases $I_{cum}(t) := C_H(t) + C_A(t)$, such that finally we only can consider ratios $I_{cum}(t)/C_H(t)$ for further informations at vaious times t, e.g. $t = t_0$ or $t = t_H$ in the initial phase of the outbreak or later in the exponential phase at a time t where we only have the exponential function of the positive eigenvalue λ_1 active, hence in our linearly approximated system for $t \to \infty$.

Finally, for the notified recovered cases C_R we have because of the dynamic equation

$$\frac{d}{dt} C_R = \gamma(H + \xi A) \tag{74}$$

now the solution

$$C_R(t) = C_R(t_0) + \gamma \cdot \tilde{\kappa}_1 \cdot f(\lambda_1) + \gamma \cdot \tilde{\kappa}_2 \cdot f(\lambda_2) \tag{75}$$

with the same eigenvalue dependent functions $f(\lambda_i)$ as in C_H, but different initial condition dependent constants $\tilde{\kappa}_i$. Namely, we have

$$\tilde{\kappa}_1 = K_{H1} + \xi K_{A1} = \frac{1}{\kappa}\Big[\ ((1-\eta)\xi + \eta)(H_0 + \phi A_0)\Big] \tag{76}$$

and

$$\tilde{\kappa}_2 = K_{H2} + \xi K_{A2} = \frac{1}{\kappa}\Big[\ (\phi - \xi)(1-\eta)H_0 + (\xi - 1)\eta A_0)\Big] \tag{77}$$

completing the calculations for the whole SHAR-C model in approximation of $S \approx N$ and time independent parameters.

Hence we obtain that the growth factor of all really observable quantities C_H, C_A and C_R is $\lambda = \lambda_1$, since the term $\left(e^{\lambda_2(t-t_0)} - 1\right) = (1 - e^{-|\lambda_2|(t-t_0)}) \to 1$ for all cumulative cases, giving no further increment

of cases after the short "pre-exponential" phase until the initial condition information becomes irrelevant as $e^{-|\lambda_2|(t-t_0)} \to 0$ while $\left(e^{\lambda_1(t-t_0)} - 1\right) \to e^{\lambda_1(t-t_0)}$ increases exponentially as long as $\lambda_1 > 0$, as we observe for the dynamical varaibles

$$H(t) \to K_{H1} \cdot e^{\lambda_1(t-t_0)} \qquad , \qquad A(t) \to K_{A1} \cdot e^{\lambda_1(t-t_0)} \tag{78}$$

for large times t. Hence with this complete analysis of the SHAR-C model we have described the dynamics as well for the "pre-exponential phase", where the initial conditions still matter in the interplay between λ_1 and λ_2, and for the "exponential phase", when only the common growth factor $\lambda = \lambda_1$ matters for all variables. We will use the present notions below for further numerical studies of the currently best SHARUCD model 2, which will be developed below, again based on the analysis of growth factors, for examining especially the pre-exponential initial phase, which has not been characterized well up to now.

4.3.4. SIR-limiting case

For cross checks and the basic understanding of the dynamics of the epidemic we now investigate simplifications to obtain an SIR-type system, where now we assume $I := H + A$, hence no impact of eventual recording problems of detecting mild/asymptomatically infected. We can immediately construct a limiting SIR system by considering $\phi \to 1$, which gives

$$B = \beta \begin{pmatrix} \eta & \eta \\ (1-\eta) & (1-\eta) \end{pmatrix} \tag{79}$$

for entries into the disease classes and

$$G = \gamma \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{80}$$

for exits with eigenvalues for $J = B - G$ given by

$$\lambda_{1/2} = \frac{1}{2} \cdot tr \pm \sqrt{\frac{1}{4} \cdot tr^2 - det} \tag{81}$$

with the parameter dependent trace

$$tr = \beta - 2\gamma \tag{82}$$

and determinant

$$det = -\beta\gamma + \gamma^2 \tag{83}$$

such that the eigenvalues are in this SIR-limiting case

$$\lambda_{1/2} = \frac{1}{2}\beta - \gamma \pm \frac{1}{2}\beta \tag{84}$$

giving

$$\lambda_1 = \beta - \gamma \tag{85}$$

and

$$\lambda_2 = -\gamma \quad . \tag{86}$$

From the above system we obtain now a plain SIR system, when considering in this case $I := H + A$ and $dI/dt = dH/dt + dA/dt = \beta(S/N)(H + A) - \gamma(H + A) = \beta(S/N)I - \gamma I$, with its only growth factor $\lambda = \beta - \gamma$ which is the largest eigenvalue λ_1 of the HA-system in the SIR-limit.

4.4. Reproduction ratio

Since the concept of reproduction numbers of infected are in the main public media by now often refered to we describe the concept breifly here, with the caution that as single number cannot give as good information as the analysis of all growth rates of the available observables, as we frequently used in the BMTF in the Basque Country to monitor the situation after the effects of the lockdown became visible, and since we enter now in an again more stochastic phase of isolated outbreaks.

Starting point is the basic reproduction number R_0, the number of secondary cases I_s from a primary case I_p during its infectiveness before recovering in a completely susceptible population. In its simplest version for SIR models a primary case, $I_p(t_0) = 1$, recovers via

$$\frac{dI_p}{dt} = -\gamma I_p \tag{87}$$

hence

$$I_p(t) = I_p(t_0)e^{-\gamma(t-t_0)} \quad . \tag{88}$$

The number of secondary cases from the primary case is given by

$$\frac{dI_s}{dt} = \beta \frac{S}{N} I_p(t) \tag{89}$$

with $I_s(t_0) = 0$, a simple inhomogeneous linear differential equation in case of a entirely susceptible population $S(t) = N$. The solution is

$$I_s(t) = \frac{\beta}{\gamma} \cdot I_p(t_0) \left(1 - e^{-\gamma(t-t_0)}\right) + I_s(t_0) \tag{90}$$

and gives the total number of secondary cases from a primary case as the long time limit as $I_s(t \to \infty) = \frac{\beta}{\gamma}$, hence the basic reproduction number is simply $\mathcal{R}_0 = \frac{\beta}{\gamma}$. So we have the relation between $\mathcal{R}_0$ and the growth rate λ here as $\mathcal{R}_0 = \frac{\beta}{\gamma} = 1 + \frac{\lambda}{\gamma}$. Generalized to a population of primary cases $I_p(t_0) > 1$, the reproduction ratio is given by

$$r := I_s(t \to \infty)/I_p(t_0) = \frac{\beta}{\gamma} \tag{91}$$

as the ratio of secondary cases produced by primary cases during their infectiousness.

This concept can be also extended to larger compartmental models, with the notions of matrices B and G as introduced above, replacing the single parameters β and γ in the simple SIR model. For any primary cases H_p or A_p we have with $\underline{x}_p = (H_p, A_p)^{tr}$ the decay dynamics

$$\frac{d}{dt}\underline{x}_p = -G \cdot \underline{x}_p \tag{92}$$

with solution

$$\underline{x}_p(t) = e^{-G(t-t_0)} \cdot \underline{x}_p(t_0) \tag{93}$$

using again the matrix exponential. For secondary cases H_s and A_s we have the dynamics of $\underline{x}_s = (H_s, A_s)^{tr}$ given by

$$\frac{d}{dt}\underline{x}_s = B \cdot \underline{x}_p(t) \tag{94}$$

with solution analogously to the SIR case as

$$\underline{x}_s(t) - \underline{x}_s(t_0) = BG^{-1}\left(1 - e^{-G(t-t_0)}\right)\underline{x}_p(t_0) \tag{95}$$

with

$$F := BG^{-1} \tag{96}$$

the next generation matrix, since

$$\underline{x}_s(t \to \infty) = F \cdot \underline{x}_p(t_0) \tag{97}$$

or from generation $\underline{x}_n$ to generation $\underline{x}_{n+1}$ the discrete iteration

$$\underline{x}_{n+1} = F \cdot \underline{x}_n \quad . \tag{98}$$

For the present case we have the next generation matrix given as

$$F = \frac{\beta}{\gamma} \cdot \begin{pmatrix} \eta\beta & \phi\eta \\ (1-\eta) & \phi(1-\eta) \end{pmatrix} \tag{99}$$

with its dominant eigenvalue for the basic SHARUCD-model

$$r_1 = (\eta + (1-\eta)\phi) \cdot \frac{\beta}{\gamma} \tag{100}$$

and the other one being zero. In the limiting case of a simple SIR-type model, with $\phi \approx 1$ and $\mu, \nu \ll \gamma$, we obtain again $r_1 = \beta/\gamma$ as can be easily seen.

This concept of the reproduction ratio can be extended into the phase when effects of the control measures become visible and parameters slowly change. The momentary reproduction ratios $r(t)$ can be analyzed, as frequently done for the COVID-19 epidemics, but often called “basic reproduction number”. While the momentary growth rate follows directly from the time continuous data at hand, the momentary reproduction ratio depends on the notion of a generation time γ^{-1}. In many countries now the reproduction ratios are discussed in the wider public in the context of COVID-19, but using various often quite different generation times, and hence the reproduction ratios are as single numbers not always the most informative measures. The growth factors of various variables as used here often give a better picture, especially when increasing testing capacities suggest larger reproduction ratios but severe disease, like hospitalization, ICU admission and deceased case are less affected. For deceased e.g. there is no ”reproduction” or ”generation time”, but well a growth factor.

4.5. Growth rate of the full SHARUCD model 1

We now have for the full SHARUCD model 1 the constant matrices

$$B = \beta \begin{pmatrix} \eta & \phi\eta \\ (1-\eta) & \phi(1-\eta) \end{pmatrix} \tag{101}$$

for entries into the disease classes and

$$G = \begin{pmatrix} (\gamma+\mu+\nu) & 0 \\ 0 & \gamma \end{pmatrix} \tag{102}$$

for exits from the disease classes. With $J = B - G$ the eigenvalues are given by

$$\lambda_{1/2} = \frac{1}{2} \cdot tr \pm \sqrt{\frac{1}{4} \cdot tr^2 - det} \tag{103}$$

with the parameter dependent trace

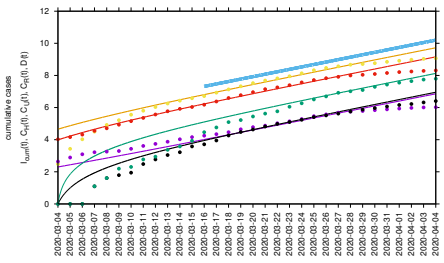
$$tr = (\eta + \phi(1-\eta)) \cdot \beta - (2\gamma + \mu + \nu) \tag{104}$$

and determinant

$$det = \gamma(\gamma+\mu+\nu) - ((\gamma+\mu+\nu)\phi(1-\eta) + \gamma\eta) \cdot \beta \quad . \tag{105}$$

In the limiting case of a simple SIR-type model we obtain again $\lambda_1 \approx \beta - \gamma = \lambda$ where $\lambda \approx \beta - \gamma$ is not a bad initial guess also for SHARUCD models, but deviates slightly from parallel behaviour of all variables during their exponential phases in the semi-logarithmic plots shown in Fig. 3. As light blue line in Fig. 3 is the growth rate, λ_1 in Eq. (103), of the complete model given.

The concept of the growth rate can be extended into the phase when effects of the control measures become visible and parameters slowly change, such that for short times the above analysis holds as for constant parameters [12]. The momentary growth rates are analyzed below.



a)

Figure 3: Semilogarithmic plot of all variables of the exponential growth phase of the epidemic in the Basque Country, namely positive tested (yellow), hospitalized cases (red), ICU admitted (violet), deceased cases (black) and recovered (green), and in comparison the growth factor λ as light blue line.

4.6. Momentary growth rates and momentary reproduction ratios and improvement of the basic SHARUCD model 1

To obtain the momentary growth rates from the COVID-19 data at hand available for the Basque Country we use $\lambda = \frac{d}{dt}\, ln(I)$ at first applied to the cumulative tested positive cases $I_{cum}(t)$ obtaining, via a smoothing window, the new cases after time τ as

$$I_{new,\tau}(t) := I_{cum}(t) - I_{cum}(t-\tau) \tag{106}$$

and hence, the growth rate

$$\lambda = \frac{1}{\Delta t}\Big(ln(I_{new,\tau}(t)) - ln(I_{new,\tau}(t-\Delta t))\Big) \quad . \tag{107}$$

From the growth rate, the reproduction ratio is calculated with the recovery period γ^{-1} obtained from underlying models and recent literature about SARS-CoV-2 interaction with human hosts [67], [68].

The concept of growth rates can be extended to the other measured variables, hospitalizations, deceased cases, recovered cases and ICU admitted cases. The sigmoidal shape of decreasing growth rates is well visible in the hospitalized and the tested positive infected, whereas the deceased and the recovered are following only later with a delay of about 8 to 10 days, and a much slower sigmoidal curve or near to linear decline. Notably, the ICU admitted cases follow the sharp sigmoidal decline of growth rate of hospitalized and tested positive cases rather than the growth rate of deceased and the recovered [12]. These results let to information about how to refine the model in order to capture the dynamics of the ICU admissions, better than in the present model. We now describe the refinements to the SHARUCD model 2.

4.7. The refined model, SHARUCD model 2

The refined model is given by changes in the transition rates of the stochastic version, see [12], such that now the transition to ICU admission is in synchrony with the transition to hospital. Given the observed synchronization of the ICU admission cases with the cumulative tested positive cases and hospitalizations, the SHARUCD model is hence refined as follows. We change the transition into ICU admissions from the previous assumption of hospitalized patients recovering with recovery rate γ, being admitted to ICU facilities with rate ν or dying with disease induced death rate μ by the assumption that ICU admissions are consequence of disease severity prone to hospitalization analogously to rate η.

The updated transitions are changed from the previously used form

$$\begin{array}{lll} w_1(\underline{x}) = \eta\beta x_1(x_2 + \phi x_3 + \varrho) & , & \underline{r}_1 = (1,-1,0,0,0,-1,0,0,0,0)^{tr} \\ w_7(\underline{x}) = \nu x_2 & , & \underline{r}_7 = (0,1,0,0,-1,0,0,-1,0,0)^{tr} \end{array} \tag{108}$$

into

$$\begin{array}{lll} w_1(\underline{x}) = \eta(1-\nu)\beta x_1(x_2 + \phi x_3 + \varrho) & , & \underline{r}_1 = (1,-1,0,0,0,-1,0,0,0,0)^{tr} \\ w_7(\underline{x}) = \eta\nu\beta x_1(x_2 + \phi x_3 + \varrho) & , & \underline{r}_7 = (1,0,0,0,-1,-1,0,-1,0,0)^{tr} \end{array} \tag{109}$$

with the parameter ν adjusted from the ICU-admission rate in units of d^{-1} into an ICU-admission ratio $\nu \in [0,1]$. By changing and fixing $\nu = 0.1$, we obtain immediately a very good agreement between the available cumulative ICU data and model simulations, with only small deviations in the other variables [12]. The deterministic version of the model is given by a differential equation system for all classes, including the recording classes of cumulative cases C_H, C_A, C_R and C_U by

$$\begin{aligned}
\frac{d}{dt}S &= -\beta\frac{S}{N}(H+\phi A+\varrho N)\\
\frac{d}{dt}H &= \eta(1-\nu)\beta\frac{S}{N}(H+\phi A+\varrho N)-(\gamma+\mu)H\\
\frac{d}{dt}A &= (1-\eta)\beta\frac{S}{N}(H+\phi A+\varrho N)-\gamma A\\
\frac{d}{dt}R &= \gamma(H+U+A)\\
\frac{d}{dt}U &= \nu\eta\beta\frac{S}{N}(H+\phi A+\varrho N)-(\gamma+\mu)U\\
\frac{d}{dt}C_H &= \eta\beta\frac{S}{N}(H+\phi A+\varrho N)\\
\frac{d}{dt}C_A &= \xi\cdot(1-\eta)\beta\frac{S}{N}(H+\phi A+\varrho N)\\
\frac{d}{dt}C_R &= \gamma(H+U+\xi A)\\
\frac{d}{dt}C_U &= \nu\eta\beta\frac{S}{N}(H+\phi A+\varrho N)\\
\frac{d}{dt}D &= \mu(H+U)
\end{aligned} \tag{110}$$

which gives the improved SHARUCD model 2. With this improved model we could describe the dynamics of the ICU admissions as well as the ones for hospitalized and deceased cases, which was not possible earlier with the basic model in that quality [12].

4.7.1. Analytical results for SHARUCD model 2 for the exponential phase

For the improved model, SHARUCD model 2, we have now

$$B = \beta\begin{pmatrix} \eta(1-\nu) & \phi\eta(1-\nu)\\ (1-\eta) & \phi(1-\eta)\end{pmatrix} \tag{111}$$

and

$$G = \gamma\begin{pmatrix} (1+\mu/\gamma) & 0\\ 0 & 1\end{pmatrix} \tag{112}$$

with the eigenvalues in a form resembling the SIR-limiting case, see above. It is again $\lambda_{1/2} = \frac{1}{2}\cdot tr \pm \sqrt{\frac{1}{4}\cdot tr^2 - det}$ with the parameter dependent trace $tr = \beta(\eta(1-\nu)+\phi(1-\eta))-(2\gamma+\mu)$ and determinant $det = -\beta(\gamma\eta(1-\nu)+(\gamma+\mu)\phi(1-\eta))+\gamma(\gamma+\mu)$ such that the eigenvalues are

$$\lambda_{1/2} = \frac{1}{2}\beta((1-\eta)\phi+\eta(1-\nu))-(\gamma+\frac{\mu}{2})\pm\frac{1}{2}\beta\cdot\sqrt{c} \tag{113}$$

with a small perturbation in case of parameters close to $\mu\to 0$, $\nu\to 0$, and $\phi\to 1$ given by

$$\sqrt{c} := \sqrt{\Big((1-\eta)\phi+\eta(1-\nu)\Big)^2+2\frac{\mu}{\beta}((1-\eta)\phi-\eta(1-\nu))+4\frac{\nu}{\beta}\gamma\eta+\left(\frac{\nu}{\beta}\right)^2} \tag{114}$$

hence with the structure of the solution well visible in the limiting case of the SIR system, Eq. (84), for $\mu = 0$, $\nu = 0$, and $\phi = 1$.

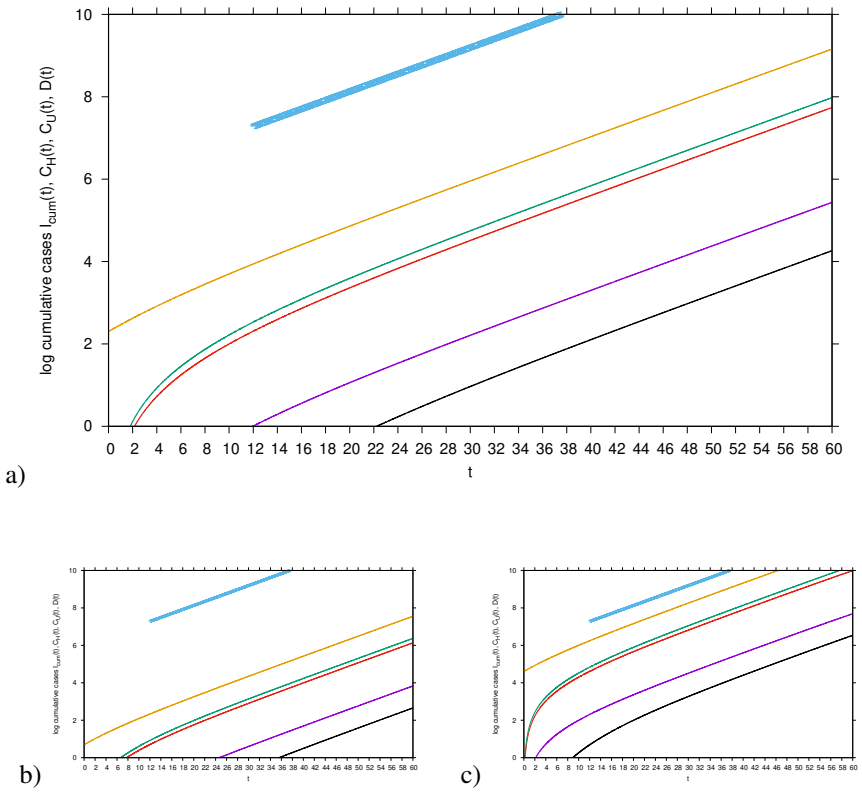


Figure 4: Semilogarithmic plots of the analysis of the initial phase of the epidemic before all observables enter into the same exponential growth. Parameter values and zero initial conditions for all but one variable, the mild/asymptomatic infected are given in the text. In all graphs we plot observables $I_{cum} = C_H + C_A + C_U$ (yellow), C_R (green), C_H (red), C_U (violet) and D (black), and in blue the growth factor λ. In a) we assume intially $A_0 = 10$ mild/asymptomatic infected and all mild/asymptomatic infected detected $C_A(t_0) = A_0$. The first severe/hospitalized infected appears after two days, ICU admitted after 12 days and deceased cases after 22 days. For reference we also show the results for b) $A_0 = 2$ and c) $A_0 = 100$, observing when the various observables cross from below to in mean 1 case, i.e. $ln(1) = 0$.

4.7.2. Numerical analysis of full system including the pre-exponential phase

For any numerical perturbational analysis we use the parameters close to the ones obtained from the data analysis of the Basque COVID-19 data, but occasionally stay closer to the limiting SIR system, hence $\mu \ll \gamma$, $\nu \ll 1$ and $\phi \approx 1$. Hence we have $\gamma = 0.05\ d^{-1}$ and $\beta = 3 \cdot \gamma$ during the onset of the epidemic, close to what was estimated during the exponential phase. For a clear distinction between mild/asymptomatic and smaller numbers of hospitalized we take $\eta = 1/4$ slightly lower than the value obtained from the Basque data. For any stochastic considerations we take the population size $N = 2 \cdot 10^6$ close to the actual population of the Basque Country of 2.2 million inhabitants. For the perturbational parameters we take $\mu = (1/10) \cdot \gamma$ which is a bit lower than the actually measured mortality, whereas for $\nu = 0.1$ we have the actual value, and finally we have $\phi = 1.1$ which is lower than what to expect from the data, but closer to the perturbational situation of a value of 1. Of course the actually best values from the data can then be used as well, ones the directions of perturbational effects are explored. The first results in the semilogarithmic plots with intially A_0 mild/asymtomatic infected are shown in Fig. 4 to investigate the pre-exponential phase. Of special interest is the onset of severe infection for various levels of initial mild/symptomatic infected.

Hence with results like the ones shown in Eq. (66) we have characterized for SHARUCD model 2 the complete onset of the epidemic, with Eq. (61) also holding for the complete model starting its dynamics from initial conditions e.g. as some mild or asymptomatic index cases $A_0 > 0$ and no as yet infected severe

cases $H_0 = 0$ as well as all the other compartments. This would be a situation of introduction of the disease before the first known cases in the pre-exponential phase. We still would have to consider in a stochastic framework the role of import creating randomly initial infected which either might recover without further transmission or cause a fist outbreak which by chance might terminate on its own or spark the exponential phase with explosive growths. Such random fluctuations would become most pronounced with large variations of system size in an eventual re-entry phase after lockdown and lifting to a situation of $\beta \approx \gamma$ or momentary reproduction ratio of around 1, and due to increasing travel randomly incoming index cases, see [9], [10] e.g. for the description of such critical fluctuations and similar scenarios in vaccinated populations hovering around the vaccination threshold [44]. Controlled systems in general seem to tend to operate around situations where the control just works but does not harm too much other aspects of life affected by interventions. However, this stage is most vulnerable to large fluctuations, hence critical, and the system evolved to it.

4.8. Conclusions and discussion

We investigated after the general discussion on model comparison of different models on the same data and the introduction and qualitative behaviour of basic SHAR models, including their description by effective SIR models, the possibilities to actually parametrize well such models with severe/potentially hospitalized and mild/asymptomatic and potentially undetected infecteds with sufficient detail to account for all relevant data features but not introduce uninformed mechanisms where we do not have data to support in the present and unprecedented pandemic of COVID-19. Uppon empirical data model refinements could be performed and now such currently best minimalistic models can be used for further analysis of various management questions, as exercised here for re-entry of infection or entry into new areas of the world after analysing the initial pre-exponential phase of the epidemic in a given territory.

Preparation for a possible second wave of transmission of COVID-19 are on the way, as the influence of seasonality and the extend of acquired immunity and its duration against SARS-CoV-2 are not clear nor well measurable yet. This will be specially important when the lockdown is completely being lifted, under the danger of re-emergence, likely starting with localized outbreaks of various sizes.

ACKNOWLEDGEMENTS

Maíra Aguiar is receiving funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 792494, and Nico Stollenwerk is receiving funding from Trento University Mathematics Dpartment Guest Lecturer Programs and from GNAMPA, Rome. Data on COVID-19 were provided by the Basque Health Department. We thank for the huge efforts of the whole COVID-19 Basque Modelling Taks Force (BMTF), specially to Eduardo Millán for collecting and preparing the data sets used in this study. We thank Adolfo Morais Ezquerro, Vice Minister of Universities and Research of the Basque Goverment, Joseba Bidaurrazaga Van-Dierdonck, Basque Public Health Department, and Javier Mar, Servicio Vasco de Salud-Osakidetza, Debagoiena Integrated Healthcare Organisation Research Unit, for many enlightening discussions on the COVID-19 development in the Basque Country as it unfolded, and especially Luís Mateus, Lisbon, and Raquel Filipe, London, for many interesting discussions on Bayesian statistics, model selection and SHAR-type models. Further we thank Andrea Pugliese and Mattia Sensi, Trento, for general discussions on COVID-19 in Italy and elsewhere, and Jose Antonio Lozano, Scientific Director of BCAM, Bilbao, for the fruitful discussions on various complementary modellings. Many of the here presented aspects have been developed over the past years at Lisbon University and are revisited in context of the present pandemic during recent lectures and presentations given at Trento University and in on-line seminars and discussions by the authors.

REFERENCES

[1] Filipe, R., Stollenwerk, N., Mateus, L., Ghaffari, P., Halstead, S., Aguiar, M., Effective infection rate in SIR-type models from models with symptomatic and asymptomatic infection, Proceedings of the 16th International Conference on Mathematical Methods in Science and Engineering - CMMSE 2015, Cádiz, Spain, ISBN: 978-84-608-6082-2, edited by Jesus Vigo et al., 2016.

[2] Stollenwerk, N., Filipe, R., Mateus, L., Ghaffari,P., Kooi, B., Halstead, S., and Aguiar, M., Effective parameters, likelihoods and Bayesian model selection in application to epidemiological models: from SHAR to effective SIR models, Proceedings of the 17th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2017, Cádiz, Spain, ISBN: 978-84-617-8694-7, edited by Jesus Vigo et al., 2017.

[3] Aguiar, M., Ballesteros, S., Kooi, B.W., and Stollenwerk, N., 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, 181–196, 2011.

[4] Kooi, W. B., Aguiar, M., and Stollenwerk, N., Bifurcation analysis of a family of multi-strain epidemiology models, Journal of Computational and Applied Mathematics, 252, 148–158, 2013.

[5] Kooi, W.B., Aguiar, M., and Stollenwerk, N., Analysis of an asymmetric two-strain dengue model, Mathematical Biosciences, 248, 128–139, 2014.

[6] Aguiar, M., Paul, R., Sakuntabhai, A., and Stollenwerk, N., Are we modeling the correct data set? Minimizing false predictions for dengue fever in Thailand, Epidemiology and Infection, 142, 2447–2459, 2014.

[7] Stollenwerk, N., and Jansen, V.A.A., Meningitis, pathogenicity near criticality: the epidemiology of meningococcal disease as a model for accidental pathogens, Journal of Theoretical Biology, 222, 347–359, 2003.

[8] Stollenwerk, N., Maiden, M.C.J., and Jansen, V.A.A., Diversity in pathogenicity can cause outbreaks of menigococcal disease, Proc. Natl. Acad. Sci. USA 101, 10229–10234, 2004.

[9] Stollenwerk, N., and Jansen, V., Population Biology and Criticality: From critical birth–death processes to self-organized criticality in mutation pathogen systems, (Imperial College Press, World Scientific, London), 2011.

[10] Stollenwerk, N., and Jansen, V.A.A., Evolution towards criticality in an epidemiological model for meningococcal disease. Physics Letters A 317, 87–96, 2003.

[11] Pollán, M., et al., Prevalence of SARS-VoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet*, published online July 6, 2020. https://doi.org/10.1016/S0140-6736(20)31483-5.

[12] Aguiar, M., Millán Ortundo, E., Bidaurrazaga Van-Dierdonck, J., Mar, J., and Stollenwerk, N., Modelling COVID-19 in the Basque Country: from introduction to control measure response, medRxiv preprint, 2020. https://doi.org/10.1101/2020.05.10.20086504.

[13] Aguiar, M., Van-Dierdonck, J. B., Stollenwerk, N., Reproduction ratio and growth rates: measures for an unfolding pandemic, medRxiv preprint and accepted for publication in PloS One, 2020. https://doi.org/10.1101/2020.05.18.20105528

[14] Mateus, L., Stollenwerk, N., and Zambrini, J.C., Stochastic Models in Population Biology: From Dynamic Noise to Bayesian Description and Model Comparison for Given Data Sets, Int. Journal. Computer Math., 90, 2161–2173, 2013.

[15] Lavezzo, E., et al., Suppression of COVID-19 outbreak in the municpality of Vó, Italy, medRxiv preprint, 2020. https://doi.org/10.1101/2020.04.17.20053157.

[16] van Kampen, N. G., Stochastic Processes in Physics and Chemistry, North-Holland, Amsterdam, 1992.

[17] Gardiner, C. W., Handbook of stochastic methods, Springer, New York, 1985.

[18] Honerkamp, J., Stochastic Dynamical Systems: Concepts, Numerical Methods and Data Analysis, VCH Publishers, Heidelberg, New York, 1993.

[19] MacKay, D., Information theory, inference, and learning algorithms, Cambridge University Press, Cambridge, 2005.

[20] Sivia, D.S., Data analysis: A Bayesian tutorial, Oxford University Press, 1996.

[21] Stollenwerk, N., Aguiar, M., Ballesteros, S., Boto, J., Kooi, W. B., and Mateus, L., Dynamic noise, chaos and parameter estimation in population biology, Interface, Focus, 2, 156–169, 2012.

[22] van Noort, S., and Stollenwerk, N., From dynamical processes to likelihood functions: an epidemiological application to influenza, in Proceedings of 8th Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2008, ISBN 978-84-612-1982-7, 2008.

[23] Aguiar, M., and Stollenwerk, N., A new chaotic attractor in a basic multi-strain epidemiological model with temporary cross-immunity, arXiv:0704.3174v1 [nlin.CD], 2007. Available at http://arxive.org.

[24] Aguiar, M., Kooi, B.W., and Stollenwerk, N., Epidemiology of dengue fever: A model with temporary cross-immunity and possible secondary infection shows bifurcations and chaotic behaviour in wide parameter regions, Math. Model. Nat. Phenom., 3, pp. 48–70, 2008.

[25] Aguiar, M., Stollenwerk, N., and Kooi, B.W., Torus bifurcations, isolas and chaotic attractors in a simple dengue model with ADE and temporary cross immunity, Int. Journal of Computer Math., 86, pp. 1867–1877, 2009.

[26] Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. H., Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, J. R. Soc. Interface, 6, 187–202, 2009.

[27] Sun, L., Lee, C., and Hoeting, J.A., Parameter inference and model selection in deterministic and stochastic dynamical models via approximate Bayesian computation: modeling a wildlife epidemic, arXiv:1409.7715v2, 2015.

[28] Stollenwerk, N., Drepper, F., and Siegel, H., Testing nonlinear stochastic models on phytoplankton biomass time series, Ecological Modelling, 144, 261–277, 2001.

[29] Halstead, S.B., Immune enhancement of viral infection. Progress in Allergy, 31, 301–364, 1982.

[30] Halstead, S.B., Neutralization and antibody-dependent enhancement of dengue viruses, Advances in Virus Research, 60, 421–467, 2003.

[31] Duong, V., Lambrechts, L., Paul, R. E., Ly, S., Laya, R. S., Long, K. C., Huy, R., Tarantola, A., Scott, T. W., Sakuntabhai, A., and Buchy, P., Asymptomatic humans transmit dengue virus to mosquitoes, Proc. Nat. Acad. Science, 112, 14688–14693, 2015.

[32] Stollenwerk, N., Mateus, L., Rocha, F., Skwara, U., Ghaffari, P., Aguiar, M., Prediction and predictability in population biology: Noise and chaos, Math. Model. Nat. Phenom., 10, 142–164, 2015.

[33] Aguiar, M., Stollenwerk, N., and Halstead, S., The impact of the newly licensed dengue vaccine in endemic countries, accepted for publication in "PLOS Neclected Tropical Diseases", published online December 21, 2016.

[34] Aguiar, M., Stollenwerk, N., and Halstead, S., The risks behind Dengvaxia recommendation, The Lancet Infectious Diseases, 16, 882–883, 2016.

[35] Aguiar, M., Halstead, S., and Stollenwerk, N., Consider stopping dengvaxia administration without immunological screening, accepted for publication in "Expert Review of Vaccines", published online December 23, 2016, http://dx.doi.org/10.1080/14760584.2017.1276831.

[36] Aguiar, M., Stollenwerk, N., and Kooi, W. B., Scaling of stochasticity in dengue hemorrhagic fever epidemics, Math. Model. Nat. Phenom., 7, 1–11, 2012.

[37] Aguiar, M., Kooi, W. B., Rocha, F., Ghaffari, P., and Stollenwerk, N., How much complexity is needed to describe the fluctuations observed in dengue hemorrhagic fever incidence data?, Ecological Complexity, 16, 31–40, 2013.

[38] Gillespie, D.T., A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, Journal of Computational Physics, 22, 403–434, 1976.

[39] Gillespie, D.T., Monte Carlo simulation of random walks with residence time dependent transition probability rates, Journal of Computational Physics, 28, 395–407, 1978.

[40] Gang, Hu, Stationary solution of master equations in the large-system-size limit, Physical Review A 36 5782–5790, 1987.

[41] Billings, L., Mier-y-Teran-Romero, L., Lindley, B., and Schwartz, I., Intervention-based stochastic disease eradication, 2013. arXiv:1303.5614v1.

[42] Stollenwerk, N, Fuentes Sommer, P., Kooi, B., Mateus, L., Ghaffari, P., and Aguiar, M., Hopf and torus bifurcations, torus destruction and chaos in population biology, Ecological Complexity, 30, 91–99, 2017.

[43] Ghaffari, P., Jansen, V., and Stollenwerk, N., Evolution towards critical fluctuations in a system of accidental pathogens, Numerical Analysis and Applied Mathematics ICNAAM 2011, Chalkidiki, AIP Conf. Proc., 1389, 1263–1266, 2011. doi: 10.1063/1.3637837 Copyright 2011 American Institute of Physics 978-0-7354-0956-9).

[44] Jansen, V.A.A., Stollenwerk, N., Jensen, H.J., Ramsay, M.E., Edmunds, W.J., and Rhodes, C.J., Measles outbreaks in a population with declining vaccine uptake, Science, 301, 804, 2003.

[45] Allen, E.J., Allen L.J.S, Arciniega, A., Greenwood, P., Construction of equivalent stochastic differential equation models, Stochastic Analysis and Application, 26, 274–297, 2008.

[46] Stollenwerk, N., and Briggs, K.M., Master equation solution of a plant disease model, Physics Letters, A 274, 84–91, 2000.

[47] World Health Organization, Emergencies preparedness, response, Novel Coronavirus – China, 2020. Retrieved from https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/

[48] World Health Organization, WHO announces COVID-19 outbreak a pandemic, 2020. Retrieved from http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic

[49] World Health Organization, Coronavirus disease (COVID-2019) situation reports,2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200420-sitrep-91-covid-19.pdf?sfvrsn=fcf0670b_4

[50] Governo Italiano Presidenza del Consiglio dei Ministri, March 9th, 2020. Retrieved from http://www.governo.it/it/articolo/firmato-il-dpcm-9-marzo-2020/14276

[51] Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática, March 14th, 2020. Retrieved from http://noticias.juridicas.com/base_datos/Laboral/661797-rd-463-2020-de-14-mar-estado-de-alarma-para-la-gestion-de-la-situacion-de.html

[52] Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática, March 27, 2020. Retrieved from http://noticias.juridicas.com/base_datos/Admin/662751-real-decreto-476-2020-de-27-de-marzo-por-el-que-se-prorroga-el-estado-de.html

[53] Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática, March 29, 2020. Retrieved from http://noticias.juridicas.com/base_datos/Laboral/662759-rd-ley-10-2020-de-29-mar-regulacion-de-un-permiso-retribuido-recuperable.html

[54] Boletín oficial: BOPV (País Vasco), DECRETO 6/2020, de 13 de marzo, del Lehendakari, Retrieved from https://www.legegunea.euskadi.eus//eli/es-pv/d/2020/03/13/6/dof/spa/html/x59-contfich/es/

[55] Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática, April 10th, 2020. Retrieved from https://www.boe.es/eli/es/rd/2020/04/10/487

[56] Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Democrática, April 10th, 2020. Retrieved from https://www.boe.es/boe/dias/2020/04/25/pdfs/BOE-A-2020-4665.pdf

[57] Anderson, R. M., Heesterbeek, H., Klinkenberg, D., Déirdre Hollingsworth, How will country-based mitigation measures influence the course of the COVID-19 epidemic?, The Lancet, 2020. doi: https://doi.org/10.1016/S0140-6736(20)30567-5

[58] Imperial College COVID-19 Response Team, Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries, March 30th, 2020. Retrieved from https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-13-europe-npi-impact/

[59] Kissler, S. M., et al., Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period, Science, 14 Apr 2020. eabb5793, DOI: 10.1126/science.abb5793

[60] Anastassopoulou, C., Russo, L., Tsakris, A. and Siettos, C., Data-based analysis, modelling and forecasting of the COVID-19 outbreak, PLoS One, 15, e0230405, 2020.

[61] Michaud, J., Kates, J., and Levitt, L., COVID-19 Models: Can They Tell Us What We Want to Know?, April 2020. Retrieved from https://www.kff.org/coronavirus-policy-watch/covid-19-models/

[62] Giordano, G., et al., Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine, 2020. doi: https://doi.org/10.1038/s41591-020-0883-7

[63] Yi, Y.,, Lagniton, P.N.P., Ye, S., Li, E., Xu, R-H., COVID-19: what has been learned and to be learned about the novel coronavirus disease, Int. J. Biol. Sci., 16(10), 1753–1766, 2020.

[64] Kramer, M., et al., Epidemiological data from the nCoV-2019 Outbreak: Early Descriptions from Publicly Available Data, 2020.Retrieved from http://virological.org/t/epidemiological-data-from-the-ncov-2019-outbreak-early-descriptions-from-publicly-\available-data/337

[65] Wu. P., et al., Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. Euro Surveill., 25, 2000044, 2020.

[66] Wei-jie Guan et al. Clinical Characteristics of Coronavirus Disease 2019 in China, New England Medicine Journal, 382, 1708–1720, 2020.

[67] Lauer, S. A., et al., The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application, Annals of Internal Medicine, 2020. doi: https://doi.org/10.7326/M20-0504.

[68] Liu, Y., et al., Viral dynamics in mild and severe cases of COVID-19, The Lancet Infectious Diseases, 2020. doi: https://doi.org/10.1016/S1473-3099(20)30232-2.