A Bayesian Framework for Parameter Estimation in Dynamical Models with Applications to Forecasting

Flávio Codeço Coelho[1,*], Cláudia Torres Codeço[2],
1 Flávio Codeço Coelho Instituto Gulbenkian de Ciência, Oeiras, Portugal
2 Cláudia Torres Codeço Scientific Computing Program, Oswaldo Cruz Foundation, Rio de Janeiro, RJ, Brazil
∗ **E-mail: fcoelho@igc.gulbenkian.pt**

Abstract

Mathematical models in Biology are powerful tools for the study and exploration of complex dynamics. Nevertheless, bringing theoretical results to an agreement with experimental observations involves acknowledging a great deal of uncertainty intrinsic to our theoretical representation of a real system. Proper handling of such uncertainties, is key to the successful usage of models to predict experimental or field observations. This problem has been addressed over the years by many tools for model calibration an parameter estimation. In this article we present a general framework for uncertainty analysis and parameter estimation which is designed to handle uncertainties associated with the modeling of dynamic biological systems while remaining agnostic as to the type of model used. We apply the framework to two Influenza transmission models: one deterministic and the other stochastic. The results show that the framework can be applied without modifications to the two types of models and that it performs equally well on both. We also discuss the application of the framework to calibrate models for forecasting purposes.

Author Summary

A mathematical model is by definition a simplified and idealized representation of a real system. How well a model's dynamics reproduce the real system's depends in part on how well we can set the parameters of the model to correspond to their real-world counterparts. To add to this problem, precise information about all the details of a model may not be available or the system may be intrinsically unpredictable or stochastic. These facts contribute to what we call the uncertainty about a model. A standard way to deal with uncertainty is to represent it as a probability distribution. In our paper, we present a comprehensive tool to use experimental or field data to reduce the uncertainties in models, by updating those probability distributions. Its usage can lead to more precise undertanding of the systems being modelled as well as more reliable forecasting of systems such the spread of infectious diseases.

Introduction

Mathematical models have long played a key role in understanding infectious disease epidemiology [1] as well other biological dynamical systems. Their ability to combine established theory and data to predict empirical observation is unique and cannot be easily achieved by other methods [2]. In such models, data in the form of rate parameters and time-series, and theory in the form of the model formulation, interact to provide insight about each other. Parameter estimation and model selection techniques, allow us to improve theory with the help of data (model selection) and estimate data which cannot be directly observed, with the help of theory (parameter estimation).

Proper representation of the intrinsic uncertainty associated with dynamic models of biological systems, has been under increasing scrutiny through the develop of a number of methods for parameter estimation and model calibration. Such methods, to be effective, must strive to be as comprehensible as possible in the treatment of all identifiable sources of uncertainty related to a given mathematical

representation of a biological system [3]. In practice, however, many uncertainty analysis methods fall short of this ideal. Some of the work in the recent literature focus on developing exact methods for parameter estimation, requiring, for instance, the derivation of the full likelihood function for the model at hand. Exact methods, however, tend to be closely coupled to a specific model or class of models, being less generally applicable [4–7].

The effort to develop appropriate estimation methods far exceed the (not in the least trivial) effort put in developing the model to it is to be applied, thus calling for the development of more general methods which will free modellers to concentrate in the modelling while having a good uncertainty analysys tool at their disposal.

In this paper we introduce a generic framework for parameter estimation in dynamic models and apply it to both deterministic and stochastic models of the spread of Influenza. The framework builds on different tools from the fields of Bayesian inference and sequential Monte Carlo (SMC) and is aimed at the fitting of dynamic models to available data. The fitting process generates posterior probability distributions for both the model's parameters and outputs. The original contribution of this paper is in the way Bayesian and SMC computational methods available in the literature [8–10] are combined to maximize generality and efficacy in the incorporation of available information into the estimation process. At the core of the framework, sits an extension of the Bayesian melding method initially proposed by Poole and Raftery [10] and successuffly applied for parameter estimations in a variety of applications [8, 9, 11].

The framework's generality stems from the fact that the dynamic model being fit to data, from the point of view of the inference machinery, is treated as a "black box" with inputs (parameters) and outputs (time-series), and the full uncertainty about each of these elements can be included in the form of prior distributions which will get updated based on observational data. The SMC approach to fitting also contributes to the framework's generality by allowing for a piece-wise fitting of the model to data. Model comparison and selection can also be easily done since the dynamic model is just a pluggable component of the framework.

The framework is meant to be used in a recursive fashion for continuous parameter estimation in situation where data becomes available gradually. This inference workflow is adequate for model-based forecasting applications which must include all available data up to a given point in time in order to better predict subsequent observations. The recursive character of the analysis permit the estimation of time-varying parameters, without requiring prior definition of a functional form for their variation. Thus, externally forcing functions affecting the model's parameters would simply be captured in the temporal variations of their estimates. This is particurlarly useful in epidemiology, where forecasting based on dynamic models often has to deal with time-varying transmission rates [12–15] .

Recursive Bayesian estimation of key parameters governing epidemic dynamics have been proposed before [13, 14, 16, 17]. Most of these studies, focus on estimating the effective reproductive number, R, of the epidemic, considering remaining parameters known. Our approach, in contrast, aims for the simultaneous estimation of all relevant parameters in an arbitrary dynamic model in a manner which is conducive to forecasting studies.

For this work, an open-source software library [18] was developed which allows for the immediate application of the framework proposed here on other models. The library can be used from within a Sage worksheet [19], requiring little programming knowledge. Example Sage Worksheets, with the models described in this paper, are provided as supplementary material.

Methods

The core of the analytical framework proposed is mainly based of the Bayesian Melding methodology [10] with modifications to make it work with dynamic models, that is, work with time-series as model outputs. The Bayesian Melding method pioneered in providing a formal inferential framework that took into full account information available about a model's inputs and outputs. Our framework's workflow is depicted

in figure 1. It treats the dynamic model (M) as a black box which take as inputs, a set of parameter values (Θ) on which we want to do inference on, and generates a set of time-series as output(Φ).

$$\Phi = M(\Theta) \tag{1}$$

The inferential problem, consists in finding the posterior probability distribution of Θ or a subset of it, and of Φ, given existing data. Data most frequently enter the inference in the form of time-series corresponding to the models variables but it can also be available for the models parameters be used to update Θ's joint prior probability distribution. The observed data ($\mathfrak{D}$) used to fit the model may be refer to only a subset of the model's outputs (Φ). The inference proceeds recursively, on an arbitrary number of time windows, W. Observed data for a given time window, $\mathfrak{D}_w$, is used to calculate the likelihood of the model's outputs:

$$L(\Phi_w) = P(\mathfrak{D}_w \mid \Phi_w) \tag{2}$$

where $w = 0, \ldots, W$. From equations 1 and 2, we have that the likelihood of the model's inputs is given by:

$$L(\Theta_w) = P(\mathfrak{D}_w \mid M(\Theta_w)) \tag{3}$$

In practice this means that the most likely sets of parameters (θ) will be the ones which generated the most likely outputs (ϕ).

The posterior of Θ is update recursively according to equation 4.

$$\pi(\Theta_w) \propto q(\Theta_w) L(\Theta_w) \tag{4}$$

where,

$$q(\Theta_w) = \begin{cases} q(\Theta_0) & if\ w = 0 \\ \pi(\Theta_{w-1}) & if\ w > 0 \end{cases} \tag{5}$$

Since outputs are dependent on inputs, information on outputs (data) can be used to restrict the input parameter space. The mechanism by which this dependency is generated can be deterministic or stochastic. The accuracy of the analysis will depend on the system's identifiability, i. e. different θ generate different ϕ.

The pluggable nature of the model, allows for a simple way to compare multiple models and select which one fits best the available data. Model comparison and selection techniques are, however, not discussed in this paper but can be found in the literature [20].

Prior Information: Before the first window of data can be processed to start the inference, prior probability distributions for the parameters in Θ, $q(\Theta_0)$ (eq. 5), need to be defined as well as the initial conditions for the model. If prior information about the distribution of the outputs are available, they can be defined here and they will be pooled with the induced prior on the outputs as described by Poole and Raftery [10]. For subsequent inference windows ($w \geq 1$), the initial values of $\{S, E, I, A, R\}$ are set to match observed data, and the probability distribution of Θ, $q(\Theta_w)$ will be $\pi(\Theta_{w-1})$, the posterior of the previous window (eq. 4).

Sequential Monte Carlo: The posterior distributions for the Θ and Φ are estimated by sampling-importance-resampling [21]. This sequential Monte Carlo procedure involves running the model k times, and then resampling the set $\{\Theta_{w,i}\}_{i=1\ldots k}$ with probability proportional to $L(\Theta_{w,i})$ fit. The l best fitting $\phi_{w,i}$, determine which $\theta_{w,i}$ are retained. Fitting procedure is described in more detail below. We will use small ϕ and θ to denote a single simulation result and a single parameter sample, respectively.

The computational steps involved in obtaining posterior distributions for the Θ_m and Φ_m, on every time window w, are summarized below. It assumes that we have observations of at least one of the outputs of the models for the time window.

1. If $w = 0$, specify prior distributions for the input parameters, $q(\Theta_w)$, and initial values for the state-variables as described above. If $w \geq 1$, set $q(\Theta_w) = \pi(\Theta_{w-1})$.
2. Obtain k samples the joint probability distribution of parameters, $q(\Theta_w)$ by means of Latin-hypercube-sampling. Regular random sampling is also possible but less efficient.
3. Run the model k times from $t = 0, \ldots, T$ where T is the length of the time window, with initial values adjusted to match observed data.
4. If there are priors for Φ (not the case in the examples presented), form the pooled prior distribution of the model outputs according to Poole and Raftery. [10]
5. Calculate the likelihood of each set of outputs ($L(\Phi_{w,i})$) of the model conditioned on the observed time-series available ($\mathfrak{D}_w$).
6. Form the resampling weight vector, $R_{w,i}$ which is the product of the Pooled Prior on the outputs and $L(\Phi_{w,i})$.
7. Obtain l samples from the joint prior distribution of the inputs, with probability given by $R_{w,i}$ to get the posterior distribution of the input parameters, $\pi(\Theta_w)$.
8. Run the model l times from the posterior distribution of the input parameters, to obtain the posterior distribution of the model's outputs.

For the applications presented in this paper, the distribution of $\phi_i[t]$ is assumed to be Normal. Thus $L(\Phi_{w,i})$ is a Normal likelihood function with fixed variance s^2. Other parametric forms for the likelihood function can be adopted. Figure 2 illustrates likelihood calculation. Parameters ($\Theta_{w,i}$) are resampled with probability proportional to the likelihood of $M(\Theta_{w,i})$, as given by:

$$L(\Phi_{w,i}) = \prod_{t=1}^{T} P(\mathfrak{D}_w[t] \mid \Phi_{w,i}[t]) \tag{6}$$

For larger models than the ones presented here, the computation of the likelihoods may become too expensive. In such cases, in step 5 and 6 above, another fit metric could be applied instead, namely, Approximate Bayes Computation methods(ABC) [22]. In such approach, a distance function is used to quantify how close are are each of the $\Phi_{w,i}$ to the observed data, $\mathfrak{D}_w$. For example, the root-mean-square deviation (equation 7) to compare the simulated ($\Phi_{w,i}$) and the observed time-series.

$$RMSD(\phi_{w,i}) = \sqrt{\frac{\sum_{t=1}^{T}(\phi_{w,i}[t] - \mathfrak{D}_w[t])^2}{T}} \tag{7}$$

Using ABC, parameter sets ($\Theta_{w,i}$) are resampled with probability proportional to $1 - \frac{RMSD(\phi_{w,i})}{\sum_i RMSD(\phi_{w,i})}$. For the sake of conciseness, results using ABC were not included in this paper but the method is available in the software library [18] implementing this analytical framework. The Sage worksheet available in the supplementary material, can also be easily modified to re-run the examples presented here with an ABC fitting scheme.

Running Workflow: Once the model is defined and the SMC and inference parameters are set, all is left is to run the analysis, for every w in W. Window width, T, is determined by the availability of data when data is still being collected at the time of the fitting, or by the modeler when all data is available from the start. Figure 3 illustrates the moving window inference. in which the model is fit to data in a piecewise fashion. This workflow is very conducive to forecasting since at the end of each window we have new Θ_w estimates to feed into longer forecasting simulations. The moving window approach, relax the assumption that input parameters are constant, and new estimates for the parameters are generated conditioned only on the state of the system at the beginning of the inference window (estimated in the previous window) and current data. This allows for a good fiit even when underlying model parameters (Θ) are not stationary.

Deterministic SEIAR model: In order to test the parameter estimation methodology, we setup a fairly detailed deterministic model for seasonal Influenza [23]. From this model we generate data and use it to try to estimate some key parameters of the model starting from vague priors. The chosen model structure is based on an SEIR (Susceptible-Exposed-Infectious-Recovered) with the Infectious compartment divided into symptomatic (I) and asymptomatic (A) individuals and allowing for a small probability of reinfection (σ). The model is implemented as a set of ordinary differential equations (see table 1 for parameters):

$$
\begin{aligned}
\lambda &= \beta(I + \eta A) \\
\frac{dS}{dt} &= \mu - (\lambda + \mu)S, \\
\frac{dE}{dt} &= \lambda S - (\epsilon + \mu)E \\
\frac{dI}{dt} &= \alpha\epsilon E + \alpha\sigma\lambda R - (\tau + \mu)I \\
\frac{dA}{dt} &= (1-\alpha)\epsilon E + (1-\alpha)\sigma\lambda R - (\tau + \mu)A \\
\frac{dR}{dt} &= \tau I + \tau A - R(\sigma\lambda + \mu)
\end{aligned}
$$

For the deterministic SEIAR model, we chose to estimate three parameters, $\Theta = \{\beta, \alpha, \sigma\}$, keeping the rest constant (see table 1 for parameter values). Although all parameters could be estimated, we decided to restrict the inference to the parameters that are most likely to be unique to a given epidemics.

In all experiments, simulations where started (in the first window, w_0) from these initial values: 0.999, 0, 0.001, 0, 0 and 0, for S, E, I, A, R and Ia, respectively.

Stochastic SEIAR Model

To demonstrate the suitability of the analytical framework to fitting stochastic models, we implemented a slightly simpler version of the deterministic model (no reinfection) as a system of stochastic differential equations,

$$
\begin{aligned}
P(dS_t = -1, dE = 1 \mid H_t) &= \beta S(I_t + A_t)\, dt \\
P(dE_t = -1, dI_t = 1 \mid H_t) &= \alpha\epsilon E_t\, dt \\
P(dE_t = -1, dA_t = 1 \mid H_t) &= (1-\alpha)\epsilon E_t\, dt \\
P(dI_t = -1, dR_t = 1 \mid H_t) &= \tau I_t\, dt \\
P(dA_t = -1, dR_t = 1 \mid H_t) &= \tau A_t\, dt
\end{aligned}
$$

This model is simulated using the Gillespie Direct exact algorithm [24]. Since here we are modeling discrete populations sizes, initial values for the compartments were set to $[490, 0, 10, 0, 0]$ respectively for S, E, I, A, R. In this model we estimate all four input parameters, so $\Theta = \{\alpha, \beta, \epsilon, \tau\}$. On each window w, the model was run 2000 times (k), the best 1000 (l) of them were retained. To stabilize the model's output, each of the k runs is actually an average of five runs with the same θ_i.

Results and Discussion

For testing the analytical framework proposed, simulated datasets were generated from both the deterministic and stochastic SEIAR models. This simulated dataset was then segmented in 20 1-week long segments which were "fed" into the analysis as described in the methods section.

Independent uniform prior distributions were defined for all the parameters being estimated (tables 2 and 3). These priors were defined so as to include the true parameter values in their support. Remaining parameters were kept constant throughout the simulations. When fitting real data, where true parameter values are not known, wider priors can be used. Moreover, misspecified priors which don't include the true parameter value in their support can be diagnosed by posterior distributions which get progressivelly squashed towards one end of the prior range. No priors where attributed to the model's outputs (Φ), since we wanted to imposed no direct expectation on Φ. Although vague (uniform) priors with no correlation structure were specified, in a scenario where one has empirical evidence about parameters this evidence can be incorporated (Bayesianly or not), to yield a more informative joint prior distribution ($q(\Theta)$).

The fitting of the deterministic SEIAR model to simulated data with moving 7-days window was able to consistently estimate the original parameter values with increasing precision with each time-window processed (figure 4). We also obtained a very good matching of the posterior prevalence (I and A) to data (figure 5), which further attests to the unbiasedness of the estimates (table 2). Each week, the model was run $k = 3000$ times, generating week long (7 days) time-series. Of these runs, the $l = 1000$ best runs were retained and their Θ_i formed the posterior distribution $\pi(\Theta)$ used to calculated the posterior time series, $\pi(\Phi)$(figure 5).

To simulate a forecasting setting, parameter estimates at the end of each week, $\pi(\Theta_w)$, were used to generate forecasts for the following 7 days (figure 6). These predictions were generated after adjusting the two infectious classes, I and A (The ones we are interested in forecasting) to the last observed data value. The mass-balance of the model was maintained by making adjustments to the susceptible class to maintain the population size constant.

To further examine the robustness of the inference, we added underreporting noise to the simulated data against which the model would be fit. This noise was added by multiplying each observed daily prevalence (I) by a uniformly distributed random number between 0.9 and 1. Tables 2 and 3 show the effects of underreporting noise on parameter estimates for the deterministic and stochastic models, respectively. this intensity of underreporting noise was found not to degrade inference accuracy in any substantial way.

Close inspection of the weekly parameter estimates for the deterministic model (figure 4) tells us about the temporal variation of the uncertainty in the model. For exmple we can see that the uncertainty about β, the transmission coefficient, is greatly reduced after the first 6 weeks of data are analyzed. This indicated that reasonably reliable estimates of the transmission rate are possible when the epidemic is just starting. In contrast, the reinfection parameter, σ can only be reliably estimated after the tenth week or so, which makes perfect sense, because only then we will have enough recovered individuals passive of reinfection.

From the Θ estimates ot the stochastic SEIAR model, a different set of insights about the models dynamics become evident (figure 8).The recursive nature of the inference, makes it easier to pinpoint issues of identifiability in the model structure by looking for bi- or multi-modality in posterior parameter distributions. As an example, we can look at the appearance of bimodal posterior estimates for τ, ϵ and

alpha at the beginning of the epidemic. As the dynamics progresses, however, this ambiguity is resolved. In the case of α, the fraction of asymptomatics, its estimate converge to one of the previous modes as soon as the I and A start to diverge, at around the third week.

These signals may also indicate the potential existence of more than one independent process underlying the observed dynamics and thus better inform modellers and policy makers.

The inference framework worked as efficiently for the Stochastic SEIAR model as it did for the deterministic model, even with underreporting noise (figure 7). The inference was able to generate good estimates of the true parameter values (figure 8) in both fitting scenarios: with and without underreporting noise (table 3). Consequently, the posterior parameter distributions at each week yielded reasonable predictions for the following week(figure 9).

The overall efficiency of the analysis and precision of the estimates, indicate that this method can be a very promising tool for forecasting epidemics. On the stochastic SEIAR, It is remarkable how the Θ estimates stabilize around their true value, long before the epidemic peaks (around the tenth week). This suggests that reliable predictions about rest of the epidemic could be made as early as in the 5th or 6th weeks.

It is also worth mentioning that the computational time to run each inference cycle (1 week window) is less than a minute (on a 2-core, 2.1GHz processor) for the samples sizes chosen. This makes the evaluation of multiple models in parallel (for model comparison purposes) as well as running more replicates for better results, a perfectly feasible task. The software was designed to take advantage of multiple cores on a computer, when available, giving scalability to the methodology.

In real-world situations, availability of data is much more restricted and may compromise the quality of the estimates, in such cases one may want to resort to simpler models with fewer degrees of freedom in order to get better results. Nevertheless, preliminary results by our research group using real surveillance data of seasonal Influenza, are confirmating the eficacy of this framework to predict real outbreaks and epidemics. These results will be subject of a future publication.

Estimating parameters for arbitrary types of dynamical models based on available data is an extremely challenging problem. There will always be a compromise to be made between inference precision and generality of the estimation methods. Based on the results presented here, we are convinced that a good balance can be striken between generality and the generation of quality estimates.

Acknowledgments

The authors acknowledge Gabriela Gomes for stimulating discussions during the preparation of this manuscript.

References

1. Anderson RM, May RM (1979) Population biology of infectious diseases: Part i. Nature 280: 361–367.

2. Ness RB, Koopman JS, Roberts MS (2007) Causal system modeling in chronic disease epidemiology: a proposal. Annals of Epidemiology 17: 564–8.

3. Coelho FC, Codeço CT, Struchiner CJ (2008) Complete treatment of uncertainties in a model for dengue r_0 estimation. Cadernos De Saúde Pública / Ministério Da Saúde, Fundação Oswaldo Cruz, Escola Nacional De Saúde Pública 24: 853–61.

4. Vyshemirsky V, Girolami M (2008) BioBayes: a software package for bayesian inference in systems biology. Bioinformatics 24: 1933–1934.

5. Golightly A, Wilkinson DJ (2006) Bayesian sequential inference for stochastic kinetic biochemical network models. Journal of Computational Biology 13: 838851.

6. Golightly A, Wilkinson DJ (2008) Bayesian inference for nonlinear multivariate diffusion models observed with error. Computational Statistics and Data Analysis 52: 16741693.

7. Lecca P, Palmisano A, Ihekwaba A, Priami C (2009) Calibration of dynamic models of biological systems with KInfer. European Biophysics Journal .

8. Alkema L, Raftery AE, Brown T (2008) Bayesian melding for estimating uncertainty in national HIV prevalence estimates. Sex Transm Infect 84: i11–16.

9. Ionides EL, Bret C, King AA (2006) Inference for nonlinear dynamical systems. Proceedings of the National Academy of Sciences of the United States of America 103: 18438–43.

10. Poole D, Raftery AE (2000) Inference for deterministic simulation models: The bayesian melding approach. Journal of the American Statistical Association 95: 1244–1255.

11. Ševčiková H, Raftery AE, Waddell PA (2007) Assessing uncertainty in urban simulations using bayesian melding. Transportation Research Part B 41: 652669.

12. Alkema L, Raftery AE, Clark SJ (2007) Probabilistic projections of HIV prevalence using bayesian melding. The Annals of Applied Statistics 1: 229–248.

13. Bettencourt LMA, Ribeiro RM (2008) Real time bayesian estimation of the epidemic potential of emerging infectious diseases. PLoS ONE 3: e2185.

14. Finkenstdt BF, Grenfell BT (2000) Time series modelling of childhood diseases: a dynamical systems approach. Journal Of The Royal Statistical Society Series C 49: 187–205.

15. Cauchemez S, Ferguson NM (2008) Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in london. Journal of the Royal Society, Interface / the Royal Society 5: 885–897.

16. Chowell G, Nishiura H, Bettencourt LM (2007) Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. Journal of The Royal Society Interface 4: 155–166.

17. Cauchemez S, Boelle P, Donnelly CA, Ferguson NM, Thomas G, et al. (2006) Real-time estimates in early detection of SARS. Emerging Infectious Diseases 12: 110–113.

18. Coelho FC (2009). bayesian-inference - project hosting on google code. URL `http://code.google.com/p/bayesian-inference/`.

19. Stein W, et al. (2009) Sage Mathematics Software (Version 4.1.1). The Sage Development Team. `http://www.sagemath.org`.

20. Kass RE, Raftery AE (1995) Bayes factors. Journal of the American Statistical Association 90: 773–795.

21. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical Science 7: 457–472.

22. Marjoram P, Molitor J, Plagnol V, Tavaré S (2003) Markov chain monte carlo without likelihoods. Proceedings of the National Academy of Sciences of the United States of America 100: 15324–15328.

23. Águas R, Oliveira NM, Gomes MGM (2009) Unveiling the contribution of reinfection to influenza dynamics. Submitted.

24. Gillespie D (1977) Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81: 2361, 2340.

Figure Legends

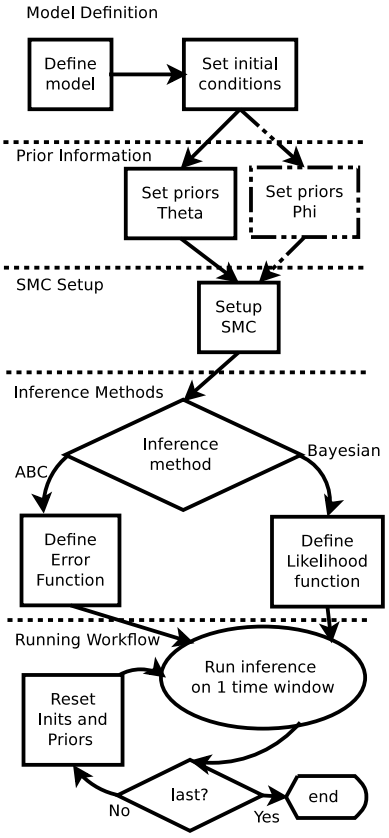

Figure 1. Analytical workflow. The inferential framework follows the digram above. The bottom section, running workflow, is the actual recursive inference cycle. The above sections correspond to the initialization of the analysis and is explained in the methodology. Dash dotted blocks indicate optional step.

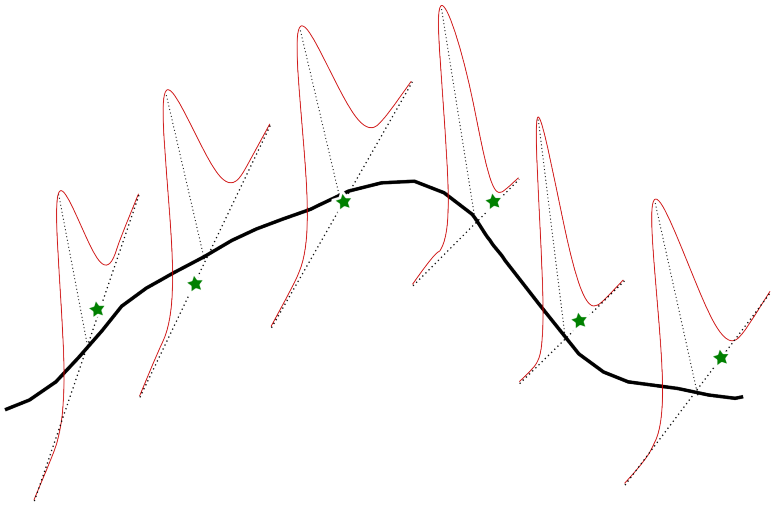

Figure 2. Likelihood calculation. This diagram describes how likelihoods for each realization of the model outputs are calculated. The full likelihood is the product of the individual likelihoods calculated for every data point on every time-series generated by the model. Green stars represent data points and the solid black line represents the model's output (ϕ_i). Gaussian curves represent the parametric form of the likelihood function.

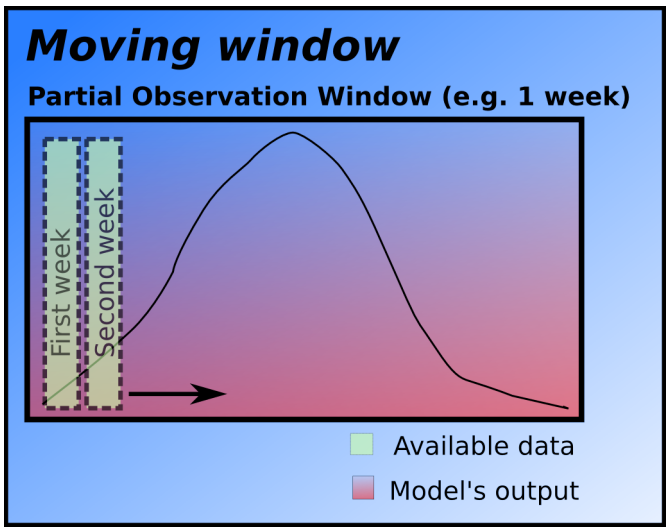

Figure 3. Moving-window Inference. Illustration of the moving window recursive inference process. Data is fed into the model in successive steps and posterior distributions on the parameters and state variables become the priors for the following step.

Tables

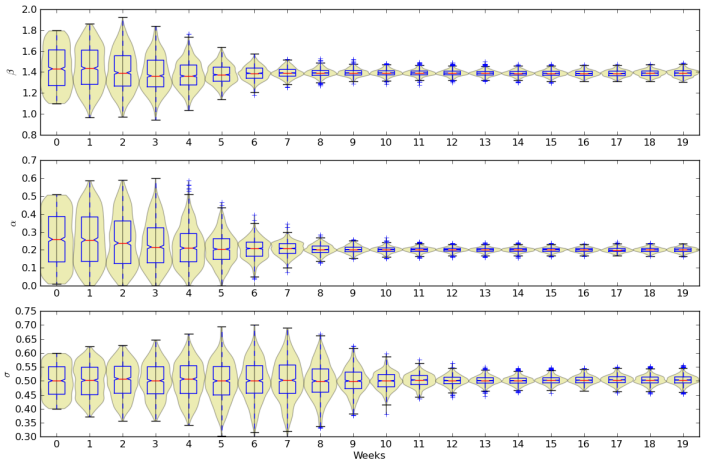

Figure 4. Temporal evolution of θ estimates, Deterministic SEIAR model. These plots represent the posterior estimates of β, α and σ at the end of each week. Yellow areas represent probability density.

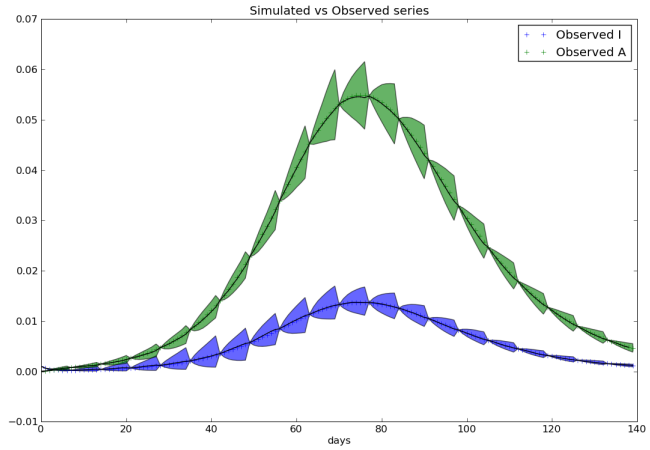

Figure 5. Posteriors for I and A, Deterministic SEIAR. Posterior Distributions of symptomatics, I and Asymptomatics, A. Black curve is the median and colored area is the 95% credible interval. Note that these are not continuous time series but rather a juxtaposition of the posterior curves at each week.

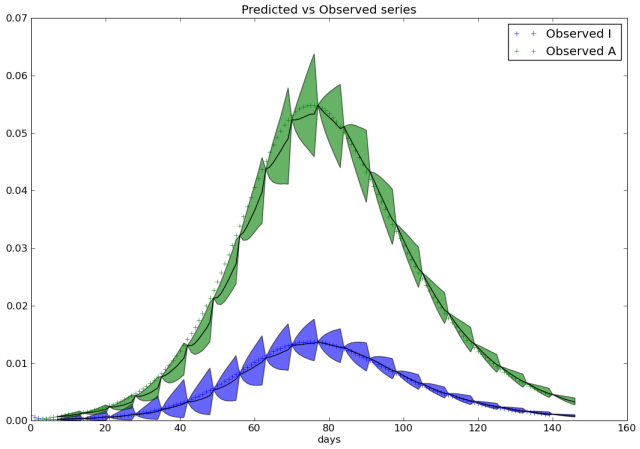

Figure 6. Predicted I and A, Deterministic SEIAR. Forecast of symptomatics, I and asymptomatics, A. Black curve is the median and colored area is the 95% interval. Curves for each week are generated by sampling parameter sets (θ_i) from the previous week posteriors.

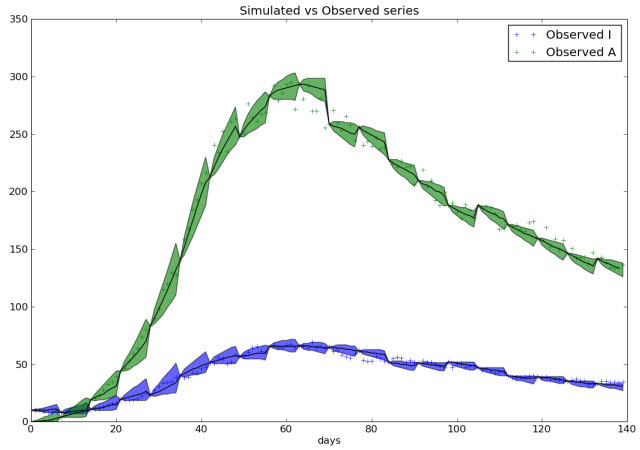

Figure 7. Posteriors for I and A, Stochastic SEIAR. Posterior Distributions of symptomatics, I and Asymptomatics, A. Black curve is the median and colored area is the 95% credible interval. In this run the fit is done against artificially underreported data, to illustrate the ability of the inference framework to deal with two levels of stochasticity.

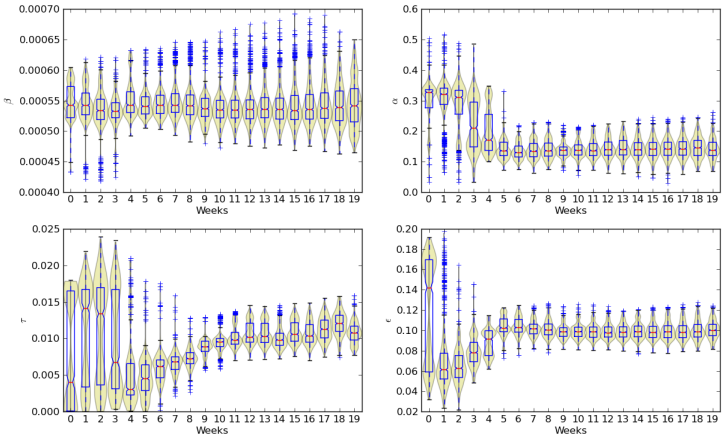

Figure 8. Temporal evolution of θ estimates, Stochastic SEIAR. These plots represent the posterior estimates of β, α, τ and ϵ at the end of each week. Yellow areas represent probability density.

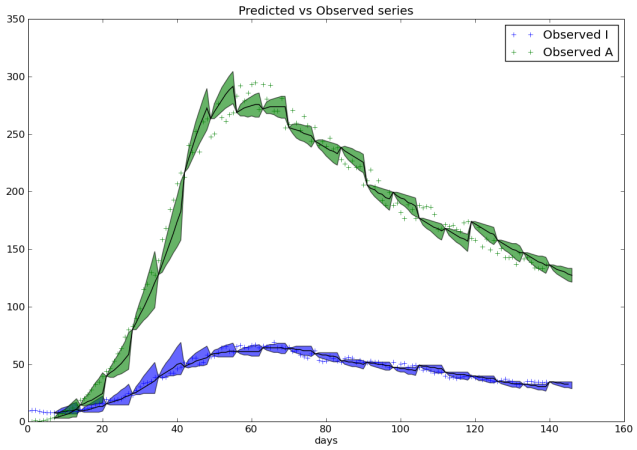

Figure 9. Predicted I and A, Stochastic SEIAR. Predictions of symptomatics, I and Asymptomatics, A. Black curve is the median and colored area is the 95% interval. Curves for each week are generated by sampling parameter sets (θ_i) from the previous week posteriors. In this run the fit is done against artificially underreported data.

Table 1. Model Parameters

Name	True value deterministic model	True value stochastic model	Meaning
μ	0	–	Birth/Death rate
β	1.4	0.00058	Transmission coefficient
η	0.5	–	Infectivity reduction for asymptomatics
ϵ	0.2	0.1	Incubation rate
α	0.2	0.2	Fraction of symptomatic
σ	0.5	–	Reinfection rate
τ	0.5	0.01	Infectious period

Parameters of the SEIAR model for both deterministic and stochastic versions, their meanings and values used in the simulations.

Table 2. Parameters Priors and Posteriors. Deterministic SEIAR

Parameter	Prior	Posterior wo/ underreporting†	Posterior w/ underreporting‡
β	$U[1.1, 1.7]$	$1.3897[1.3039, 1.4891]$	$1.4462[1.3879, 1.5193]$
α	$U[0.01, 0.51]$	$0.2000[0.1615, 0.2361]$	$0.1809[0.1559, 0.2083]$
σ	$U[0.4, 0.6]$	$0.5030[0.4527, 0.5575]$	$0.5067[0.4764, 0.5545]$

Probability distributions for the deterministic SEIAR model. Priors are Uniform distributions: $U[min, max]$ The posteriors correspond to final estimate of the parameters, after the last week of data has been processed($median[min, max]$). †fitted to simulated data. ‡Fitted to simulated data modified by under-reporting noise as described in the text.

Table 3. Parameters Priors and Posteriors. Stochastic SEIAR

Parameter	Prior	Posterior wo/ underreporting†	Posterior w/ underreporting‡
β	$U[1e-5, 6e-4]$	$0.0006[0.0004, 0.008]$	$0.0006[0.0004, 0.0007]$
α	$U[0.01, 0.5]$	$0.1907[0.0131, 0.3793]$	$0.1958[0.0688, 0.3216]$
ϵ	$U[0, 0.2]$	$0.0985[0.0737, 0.1247]$	$0.0983[0.0798, 0.1200]$
τ	$U[0.0, 0.02]$	$0.0102[0.0070, 0.0140]$	$0.0086[0.0047, 0.0128]$

Probability distributions for the Stochastic SEIAR model. Priors are Uniform distributions: $U[min, max]$ The posteriors correspond to final estimate of the parameters, after the last week of data has been processed ($median[min, max]$).†fitted to simulated data. ‡Fitted to simulated data modified by under-reporting noise as described in the text.