

**Approximating optimal SMC proposal distributions
in individual-based epidemic models**

Lorenzo Rimella, Christopher Jewell and Paul Fearnhead

Lancaster University

Supplementary Material

S1 Preliminaries

Given the vector \mathbf{a} we use $\mathbf{a}^{(i)}$ for the (i) -th element of \mathbf{a} (vectors are always assumed to be column vectors), given the matrix \mathbf{A} we use $\mathbf{A}^{(i,j)}$ for the (i, j) -th element of \mathbf{A} and we use $\mathbf{A}^{(i,\bullet)}$ (or $\mathbf{A}^{(\bullet,j)}$) to represent the column vector given by the (i) -th column (or the (j) -th row) of matrix \mathbf{A} . We use \mathbf{A}^T for the transpose of matrix \mathbf{A} (similarly for vectors). The dot product between two matrices \mathbf{A}, \mathbf{B} is denoted by \mathbf{AB} (similarly for vectors). Sums or differences between vectors or matrices have to be thought of as elementwise, similarly when using vectors or matrices and a scalar. We employ the convention \mathbb{R}_+ for the positive real numbers. We use the

notation $\mathbb{I}_a(b)$ for the indicator function, with a, b being scalars, vectors, or matrices.

Given the probability mass (or density) function of a random variable $\mathcal{D}(x|\theta)$ evaluated in x with parameters θ as in table 1, we use $X \sim \mathcal{D}(\bullet|\theta)$ to say the random variable X is drawn from the corresponding probability random variable and we denote with $\mathbb{E}_{\mathcal{D}(X|\theta)}[X]$ the corresponding expectation.

Distribution	Categorical	Bernoulli	Binomial	Gaussian	Uniform	Multinomial
Notation	$\mathcal{Cat}_M(i \mathbf{p})$	$\mathcal{Be}(i q)$	$\mathcal{Bin}(i N, q)$	$\mathcal{N}(a \mu, \sigma^2)$	$\mathcal{Unif}(q a, b)$	$\mathcal{Mult}(\mathbf{c} N, \mathbf{p})$
Domain	$[1 : M]$	$\{0, 1\}$	$[0 : N]$	\mathbb{R}	$[a, b]$	$\{\mathbf{c} : \sum_{i \in [1:M]} \mathbf{c}^{(i)} = N\}$

Table 1: Notation table for probability mass and density functions along with domains.

S2 Introduction to compartmental models

In epidemiology, compartmental models are used to describe the dynamics of an epidemic in a population, where the compartments represent different stages of the disease (Brauer, 2008). A closed population stochastic compartmental model is fully defined by: the number of compartments M , the population size N , the initial probability of being assigned to a compartment \mathbf{p}_0 , and the probability of transition from one compartment to the other \mathbf{K}_\bullet , which is a stochastic transition matrix $\mathbf{c} \rightarrow \mathbf{K}_\mathbf{c}$ defined as a func-

tion of the compartments' state \mathbf{c} . The dynamics of a compartmental model is generally described by a discrete-time Markov chain $(\mathbf{c}_t)_{t \geq 0}$ counting the number of individuals in each compartment, i.e. \mathbf{c}_t is an M -dimensional vector with $\mathbf{c}_t^{(i)}$ being the number of individuals in compartment i (Keeling and Rohani, 2011). The evolution of \mathbf{c}_t can be specified from an individual perspective by defining the discrete-time Markov chain $(\mathbf{x}_t)_{t \geq 0}$, where \mathbf{x}_t is an N -dimensional vector representing the state of the population at time t (Whiteley and Rimella, 2021), i.e. $\mathbf{x}_t^{(n)}$ is the compartment of individual n at time t . The evolution of $(\mathbf{x}_t)_{t \geq 0}$ and $(\mathbf{c}_t)_{t \geq 0}$ can be then represented as the following flow:

Time 0: $\mathbf{x}_0^{(n)} \sim \mathcal{Cat}_M(\bullet | \mathbf{p}_0)$ for $n \in [1 : N]$ and

$$\mathbf{c}_0^{(i)} = \sum_{n=1}^N \mathbb{I}_{\mathbf{x}_0^{(n)}}(i) \text{ for } i \in [1 : M];$$

Time t : $\mathbf{x}_t^{(n)} | \mathbf{x}_{t-1} \sim \mathcal{Cat}_M \left(\bullet \mid \mathbf{K}_{\mathbf{c}_{t-1}}^{(\mathbf{x}_{t-1}^{(n)}, \bullet)} \right)$ for $n \in [1 : N]$ and

$$\mathbf{c}_t^{(i)} = \sum_{n=1}^N \mathbb{I}_{\mathbf{x}_t^{(n)}}(i) \text{ for } i \in [1 : M];$$

where we first make all the individuals move (simulate \mathbf{x}_t) and then count the individuals in each compartment (compute \mathbf{c}_t).

SIS example The susceptible-infected-susceptible model (SIS) is a well-known compartmental model used to model the spread of a disease in a population where herd immunity is not possible, i.e. the individuals can be

re-infected. A stochastic SIS with closed population N can be represented in the previous framework by: $M = 2$, \mathbf{p}_0 in the 2-dimensional simplex (with $\mathbf{p}_0^{(1)}$ probability of being susceptible at time 0) and \mathbf{K}_\bullet a 2by2-dimensional stochastic transition matrix. A popular choice of \mathbf{K}_\bullet is:

$$\mathbf{K}_c = \begin{bmatrix} e^{-\beta \frac{\mathbf{c}^{(2)}}{N}} & 1 - e^{-\beta \frac{\mathbf{c}^{(2)}}{N}} \\ 1 - e^{-\gamma} & e^{-\gamma} \end{bmatrix}$$

with $\beta, \gamma \in \mathbb{R}_+$ transmission and recovery parameters.

Focusing on the compartments' state automatically assumes homogeneous individuals, which is a significant simplification of the real world, where each individual often has their own covariates, e.g. age. individual-based compartmental models or simply individual-based models relax the homogeneity assumption and look at the disease from an individual perspective. The main difference resides in the individual-specific $(\mathbf{p}_{n,0})_{n \in [1:N]}$ and $(\mathbf{K}_{n,\bullet})_{n \in [1:N]}$ representing the heterogeneous dynamic of the individuals.

S2.1 Considerations on non-granular observation models

Compartmental models in epidemiology are generally treated as latent and accompanied by an observation model representing the conditional distribution of the observations given the current compartments' state often refer as the emission distribution. Popular choices are the Negative binomial

distribution (Fintzi et al., 2021) and the binomial distribution (Lekone and Finkenstädt, 2006), with the latter being used by Ju et al. (2021) in an individual-based model framework. The main limitation of these emission distributions is that they are formulated over the aggregated population, while for an individual-based model, it is fair to assume individual-based observations.

S3 Hidden Markov models

In an HMM $(\mathbf{x}_t, \mathbf{y}_t)_{t \geq 1}$ the recursive computation of $p(\mathbf{x}_t | \mathbf{y}_{[1:t]}, \theta)$ and $p(\mathbf{y}_{[1:t]} | \theta)$ is known as forward algorithm and it consists of the following steps:

Time 0: $p(\mathbf{x}_0 | \theta)$ from the initial distribution;

Time t : $p(\mathbf{x}_t | \mathbf{y}_{[1:t-1]}, \theta) = \sum_{\mathbf{x}_{t-1} \in [1:M]^N} p(\mathbf{x}_t | \mathbf{x}_{t-1}, \theta) p(\mathbf{x}_{t-1} | \mathbf{y}_{[1:t-1]}, \theta)$ and

$$p(\mathbf{y}_t | \mathbf{y}_{[1:t-1]}, \theta) = \sum_{\mathbf{x}_t \in [1:M]^N} p(\mathbf{y}_t | \mathbf{x}_t, \theta) p(\mathbf{x}_t | \mathbf{y}_{[1:t-1]}, \theta) \text{ and}$$

$$p(\mathbf{x}_t | \mathbf{y}_{[1:t]}, \theta) = \frac{p(\mathbf{y}_t | \mathbf{x}_t, \theta) p(\mathbf{x}_t | \mathbf{y}_{[1:t-1]}, \theta)}{p(\mathbf{y}_t | \mathbf{y}_{[1:t-1]}, \theta)} \text{ and}$$

$$p(\mathbf{y}_{[1:t]} | \theta) = p(\mathbf{y}_{[1:t-1]} | \theta) p(\mathbf{y}_t | \mathbf{y}_{[1:t-1]}, \theta).$$

A close-form solution for the filtering is often not available and an SMC algorithm can be employed to compute particle estimates of both $p(\mathbf{x}_t | \mathbf{y}_{[1:t]}, \theta)$ and $p(\mathbf{y}_{[1:t]} | \theta)$. A general SMC algorithm is presented in 1 and it consists of: resampling the previous particles according to the resampling

scheme and correcting the weights accordingly, proposing new particles with the proposal distribution, update the weights. Note that it is required $\sum_{p \in [1:P]} \tilde{w}_s^p = 1$ with $\tilde{w}_s^p \propto w_s^{i_s^p} / r_s(i_s^p)$, with i_s resampled indices, which is then updated according to the transition kernel, emission distribution and proposal distribution. The likelihood estimate is then given by $\hat{p}(y_{[1:s]}|\theta) \leftarrow \hat{p}(y_{[1:s-1]}|\theta) \sum_{p \in [1:P]} w_s^p$, where w_s^p are the weights after update.

Algorithm 1 A general sequential Monte Carlo algorithm

Require: $P, \theta, \mathbf{y}_{[1:t]}, (q(\mathbf{x}_s|\mathbf{x}_{s-1}, \mathbf{y}_{[1:t]}))_{s \in [0:t]}, (r_s(i))_{s \in [0:t]}$

- 1: Sample $\mathbf{x}_0^p \sim q(\bullet|\mathbf{y}_{[1:t]})$
 - 2: Compute $w_0^p \leftarrow \frac{p(\mathbf{x}_0^p|\theta)}{q(\mathbf{x}_0^p|\mathbf{y}_{[1:t]})}$ for $p \in [1 : P]$
 - 3: **for** $s = 1, \dots, t$ **do**
 - 4: Resample $i_{s-1}^p \sim r_{s-1}(\bullet)$ and set $\tilde{\mathbf{x}}_{s-1}^p \leftarrow \mathbf{x}_{s-1}^{i_{s-1}^p}$ for $p \in [1 : P]$
 - 5: Correct the weights $\tilde{w}_{s-1}^p \propto \frac{w_{s-1}^{i_{s-1}^p}}{r_{s-1}(i_{s-1}^p)}$ and normalize for $p \in [1 : P]$
 - 6: Propose $\mathbf{x}_s^p \sim q(\bullet|\tilde{\mathbf{x}}_{s-1}, \mathbf{y}_{[1:t]})$
 - 7: Compute $w_s^p \leftarrow \tilde{w}_{s-1}^p \frac{p(\mathbf{x}_s^p|\tilde{\mathbf{x}}_{s-1}^p, \theta)p(\mathbf{y}_s|\mathbf{x}_s^p, \theta)}{q(\mathbf{x}_s^p|\tilde{\mathbf{x}}_{s-1}^p, \mathbf{y}_{[1:t]})}$ for $p \in [1 : P]$
 - 8: $\hat{p}(y_{[1:t]}|\theta) \leftarrow \prod_{s \in [1:t]} \sum_{p \in [1:P]} w_s^p$
-

In the HMM terminology the initial distribution $p(\mathbf{x}_0|\theta)$, the transition kernel $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \theta)$ and the emission distribution $p(\mathbf{y}_t|\mathbf{x}_t, \theta)$ for our

individual-based model with granular observation are given by:

$$p(\mathbf{x}_0|\theta) = \prod_{n \in [1:N]} \mathbf{p}_{n,0}^{(\mathbf{x}_0^{(n)})}, \quad p(\mathbf{x}_t|\mathbf{x}_{t-1}, \theta) = \prod_{n \in [1:N]} \mathbf{K}_{n, \mathbf{c}_{t-1}}^{(\mathbf{x}_{t-1}^{(n)}, \mathbf{x}_t^{(n)})},$$

$$p(\mathbf{y}_t|\mathbf{x}_t, \theta) = \prod_{n \in [1:N]} \left(\mathbf{q}_t^{(\mathbf{x}_t^{(n)})} \right)^{\mathbb{I}_{\mathbf{y}_t^{(n)}(\mathbf{x}_t^{(n)})}} \left(1 - \mathbf{q}_t^{(\mathbf{x}_t^{(n)})} \right)^{\mathbb{I}_{\mathbf{y}_t^{(n)}(0)}}.$$

These definitions are useful to see how the transition kernel and the emission distribution can be used to simplify the formulation of $p(\mathbf{y}_t|\mathbf{x}_{t-1}, \theta)$:

$$p(\mathbf{y}_t|\mathbf{x}_{t-1}, \theta) = \sum_{\mathbf{x}_t \in [1:M]^N} \prod_{n \in [1:N]} \mathbf{K}_{n, \mathbf{c}_{t-1}}^{(\mathbf{x}_{t-1}^{(n)}, \mathbf{x}_t^{(n)})} \left(\mathbf{q}_t^{(\mathbf{x}_t^{(n)})} \right)^{\mathbb{I}_{\mathbf{y}_t^{(n)}(\mathbf{x}_t^{(n)})}} \left(1 - \mathbf{q}_t^{(\mathbf{x}_t^{(n)})} \right)^{\mathbb{I}_{\mathbf{y}_t^{(n)}(0)}}$$

$$= \prod_{n \in [1:N]} \sum_{\mathbf{x}_t^{(n)} \in [1:M]} \mathbf{K}_{n, \mathbf{c}_{t-1}}^{(\mathbf{x}_{t-1}^{(n)}, \mathbf{x}_t^{(n)})} \left(\mathbf{q}_t^{(\mathbf{x}_t^{(n)})} \right)^{\mathbb{I}_{\mathbf{y}_t^{(n)}(\mathbf{x}_t^{(n)})}} \left(1 - \mathbf{q}_t^{(\mathbf{x}_t^{(n)})} \right)^{\mathbb{I}_{\mathbf{y}_t^{(n)}(0)}}.$$

Algorithm 2 Multinomial approximation by Whiteley and Rimella (2021)

Require: $(\bar{\mathbf{p}}_{n,0})_{n \in [1:N]}$, $(\bar{\mathbf{K}}_{\bullet})_{n \in [1:N]}$, $(\mathbf{q}_s)_{s \in [t]}$, $(\mathbf{o}_s)_{s \in [t]}$

- 1: $\mathbf{m}_{0|0} \leftarrow \bar{\mathbf{p}}_{n,0}$ ▷ Forward step
 - 2: **for** $s = 1, \dots, t$ **do**
 - 3: $\mathbf{m}_{s-1|s} \leftarrow \left(\mathbf{m}_{s-1|s-1}^T \bar{\mathbf{K}}_{\mathbf{m}_{s-1}} \right)^T$
 - 4: $\mathbf{m}_{s|s} \leftarrow \frac{\mathbf{o}_s}{N} + \left(1 - \frac{\mathbf{1}_M^T \mathbf{o}_s}{N} \right) \frac{\mathbf{m}_{s-1|s} \circ (\mathbf{1}_M - \mathbf{q}_s)}{1 - \mathbf{m}_{s-1|s}^T \mathbf{q}_s}$
 - 5: **for** $s = t-1, \dots, 0$ **do** ▷ Backward step
 - 6: $\mathbf{L}_s \leftarrow \left\{ \left[(\mathbf{m}_{s|t} \mathbf{1}_M^T) \circ \bar{\mathbf{K}}_{\mathbf{m}_s} \right] / \left[\mathbf{1}_M (\mathbf{m}_{s|t}^T \bar{\mathbf{K}}_{\mathbf{m}_s}) \right] \right\}^T$
 - 7: $\mathbf{m}_{s|t} \leftarrow \left(\mathbf{m}_{s+1|t}^T \mathbf{L}_s \right)^T$
-

S4 A priori estimate of the compartments' state

The approximate homogeneous dynamic considered in the main paper is:

Time 0: $\mathbf{x}_0^{(n)} \sim \mathcal{Cat}_M(\bullet | \bar{\mathbf{p}}_{n,0})$ for $n \in [1 : N]$ and

$$\mathbf{c}_0^{(i)} = \sum_{n=1}^N \mathbb{I}_{\mathbf{x}_0^{(n)}}(i) \text{ for } i \in [1 : M];$$

Time t : $\mathbf{x}_t^{(n)} | \mathbf{x}_{t-1} \sim \mathcal{Cat}_M\left(\bullet | \bar{\mathbf{K}}_{\mathbf{c}_{t-1}}^{(\mathbf{x}_{t-1}^{(n)}, \bullet)}\right)$ for $n \in [1 : N]$ and

$$\mathbf{c}_t^{(i)} = \sum_{n=1}^N \mathbb{I}_{\mathbf{x}_t^{(n)}}(i) \text{ for } i \in [1 : M].$$

The full algorithm by (Whiteley and Rimella, 2021) is reported in algorithm 2.

S5 Additional experiments

S5.1 Susceptible-infected-susceptible

Outcomes from the SIS model are shown in figure 1. The full table on the marginal likelihood standard deviation is reported in table 2. Complete marginal likelihood contour plot are reported in figure 2 and figure 3 in log scale and normalized to have their max in zero.

We measure the sensitivity to \mathbf{q} of APF and our method when $h = 1, 5, 10, 20$ when $P = 512$. We study the performance of the considered methods when \mathbf{q} varies, precisely we choose a scenario where data are generated according to $\mathbf{q} = [0.8, 0.8]^T$, but the algorithms use $\mathbf{q} = [i, i]^T$ with

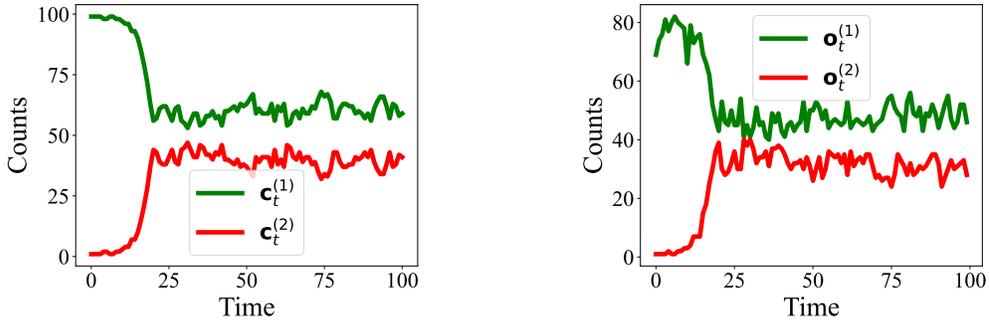


Figure 1: SIS epidemic, on the left the unobserved process, on the right the observed data.

	APF	0.7s	h=5	2.5s	h=10	3.94s	h=20	6.61s
	DGP	NDGP	DGP	NDGP	DGP	NDGP	DGP	NDGP
P	std	std	std	std	std	std	std	std
64	7.02	10.97	0.4	1.26	0.48	1.35	0.46	1.19
128	4.99	9.89	0.3	0.92	0.31	1.0	0.37	0.89
256	5.24	8.3	0.27	0.72	0.24	0.67	0.35	0.63
512	4.01	6.66	0.17	0.48	0.18	0.49	0.18	0.48
1024	3.42	7.17	0.15	0.34	0.21	0.33	0.15	0.35
2048	2.83	6.23	0.11	0.25	0.11	0.22	0.11	0.22

Table 2: Table reporting standard deviation for APF and our method when $h = 5, 10, 20$ under DGP and NDGP with $P = 64, 128, 256, 512, 1024, 2048$. The mean computational cost is reported in the first row with the name of the algorithm.

$i \in [0.1, \dots, 0.9]$. As for the previous experiment we compute the standard deviation and ESS bands by running 100 times algorithm 1 for each frame-

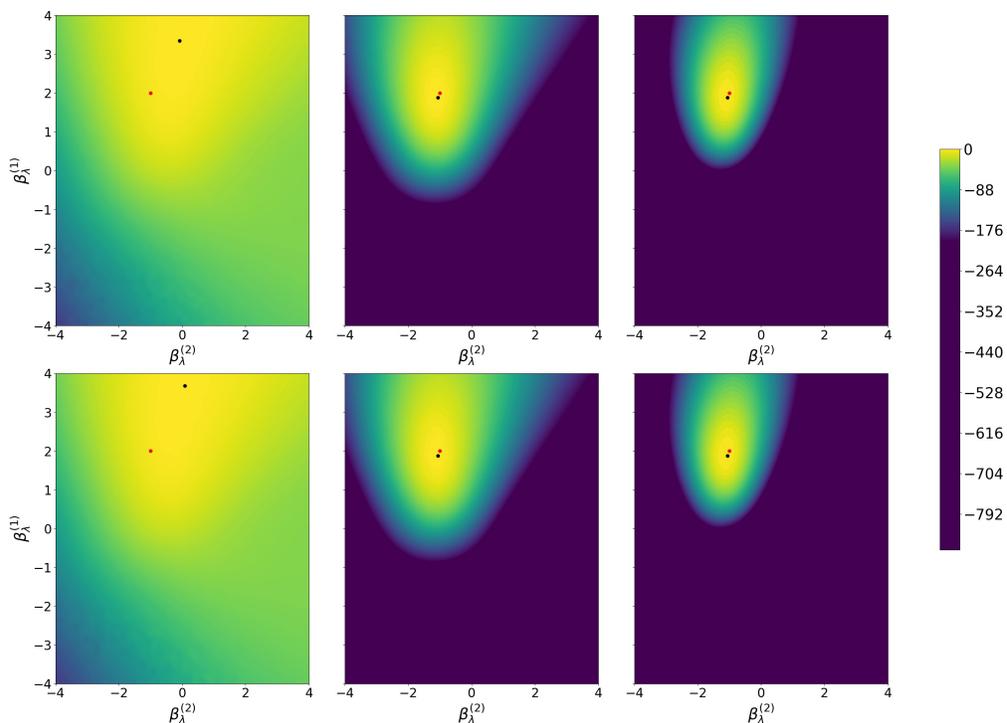


Figure 2: Marginal likelihood contour plots on a β_λ grid in log-scale. Columns refer to $t = 10, 50, 100$ from left to right. Rows refer to $h = 5, 10$ from top to bottom. The colorbar is common across the plots and in each plot, the maximum is set to 0. In red is the DGP and in black is the MLE on the grid.

work. Results are reported in figure 4. APF is always associated with a higher standard deviation of the marginal likelihood estimate and to a lower ESS. As for the previous experiments, we have a significant improvement in both standard deviation and ESS when choosing $h = 1$, this gets even better when $h \geq 5$, especially from an ESS perspective.

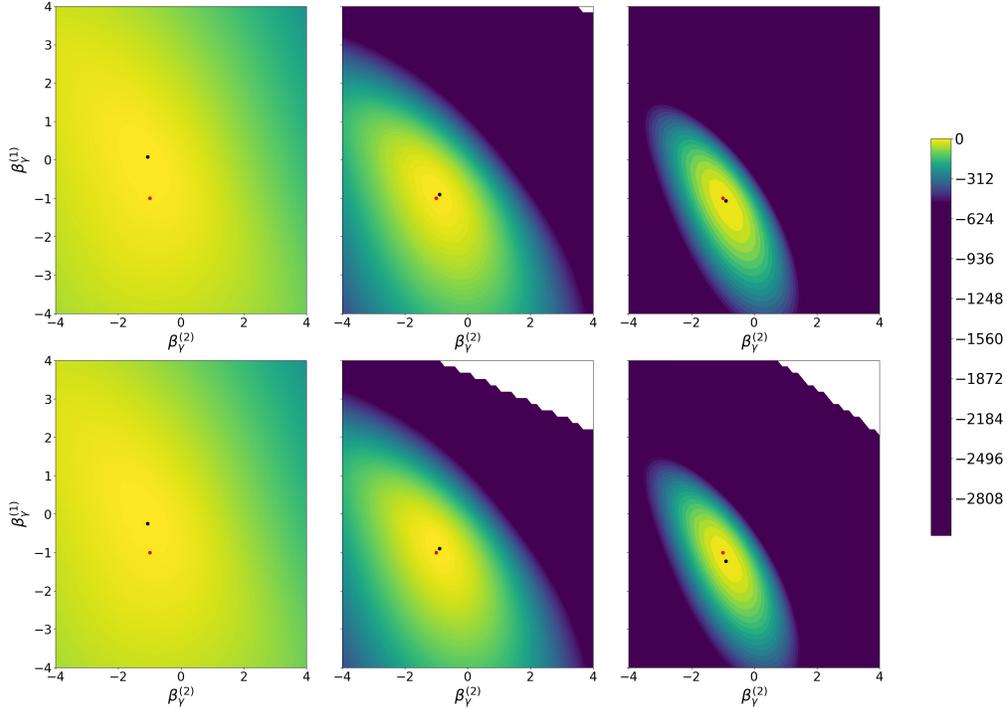


Figure 3: Marginal likelihood contour plots on a β_γ grid in log-scale. Columns refer to $t = 10, 50, 100$ from left to right. Rows refer to $h = 5, 10$ from top to bottom. The colorbar is common across the plots and in each plot, the maximum is set to 0. In red is the DGP and in black is the MLE on the grid.

We now infer $(\beta_0, \beta_\lambda, \beta_\gamma, \mathbf{q})$ from data generated from DGP. For this experiment, we employ a Particle marginal Metropolis-Hastings (PMMH) (Andrieu et al., 2010) using an SMC with our proposal distribution and resampling scheme along with the following set of priors: $\mathcal{N}(\beta_0^{(1)}|0, 3)$, $\mathcal{N}(\beta_0^{(2)}|0, 3)$, $\mathcal{N}(\beta_\lambda^{(1)}|0, 3)$, $\mathcal{N}(\beta_\lambda^{(2)}|0, 3)$, $\mathcal{N}(\beta_\gamma^{(1)}|0, 3)$, $\mathcal{N}(\beta_\gamma^{(2)}|0, 3)$, $\text{Unif}(\mathbf{q}^{(1)}|0, 1)$,

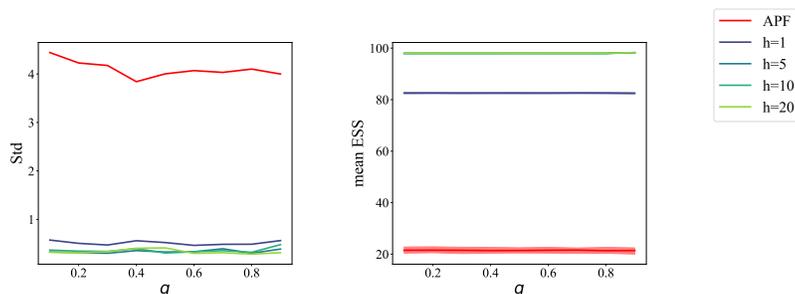


Figure 4: Effective sample size and marginal likelihood standard deviation sensitivity to \mathbf{q} . The first column shows the marginal likelihood standard deviation, while the second column reports ESS where mode, 5%-quantile, and 95%-quantile are reported.

$\mathcal{U}nif(\mathbf{q}^{(2)}|0, 1)$. As proposal distribution for the parameters, we use a Gaussian random walk on the log-parameters with variance chosen to match the optimal acceptance rate of 23%.

We run the PMMH for 100000 iterations and we used a burn-in period of 10000 and thinning. Marginal posterior distributions are reported in figure 5, we can notice that posterior are “peaky” and close to the DGP, with the exception of β_0 which has a flatter posterior, which is more difficult to learn given that the only information is derived from \mathbf{y}_1 . To push our study one step further we also plot $R_0^{(n)}$ distribution and posterior predictive in figure 6. We can observe that the PMMH learned the $R_0^{(n)}$ distribution on a global level (first column) and on an individual level (second column), with $R_0^{(n)}$ from the DGP almost indistinguishable. For the posterior predictive we

S5. ADDITIONAL EXPERIMENTS

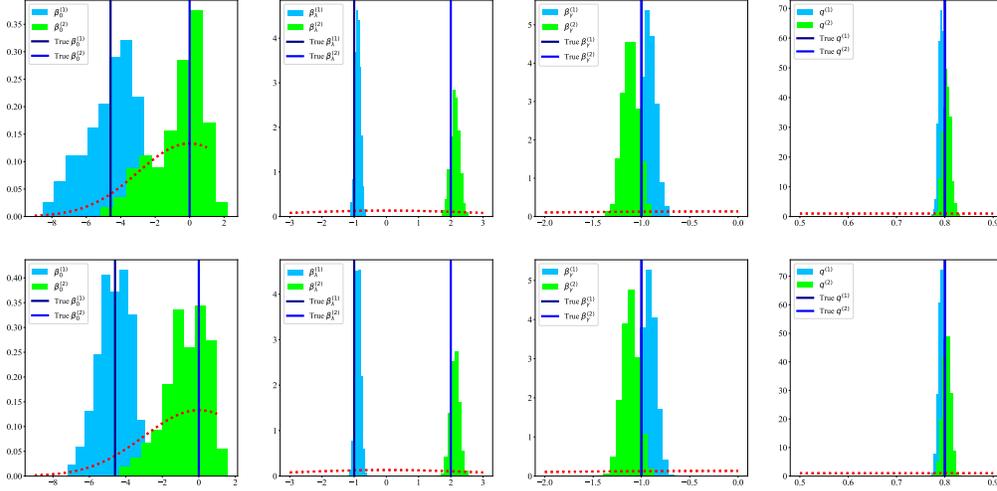


Figure 5: Posterior distribution over the parameters of interest. On the columns $(\beta_0, \beta_\lambda, \beta_\gamma, \mathbf{q})$ from left to right. On the rows $h = 5, 10$ from top to bottom. Priors are reported in red dotted lines. Vertical lines are the DGP.

sample 200 times from the posterior and we then simulate 200 epidemics with those parameters. Figure 6 shows good coverage of the aggregated data (fourth column) and of the aggregated latent data (third column). The latter is not available during the inference process, but it is stored at simulation time to add an additional level of comparison.

S5.2 Susceptible-exposed-infected-removed

We can make the SEIR model heterogeneous by considering $d \in \mathbb{N}$ and by defining $(\mathbf{w}_n)_{n \in [1:N]}$ as the collection of d -dimensional vectors collect-

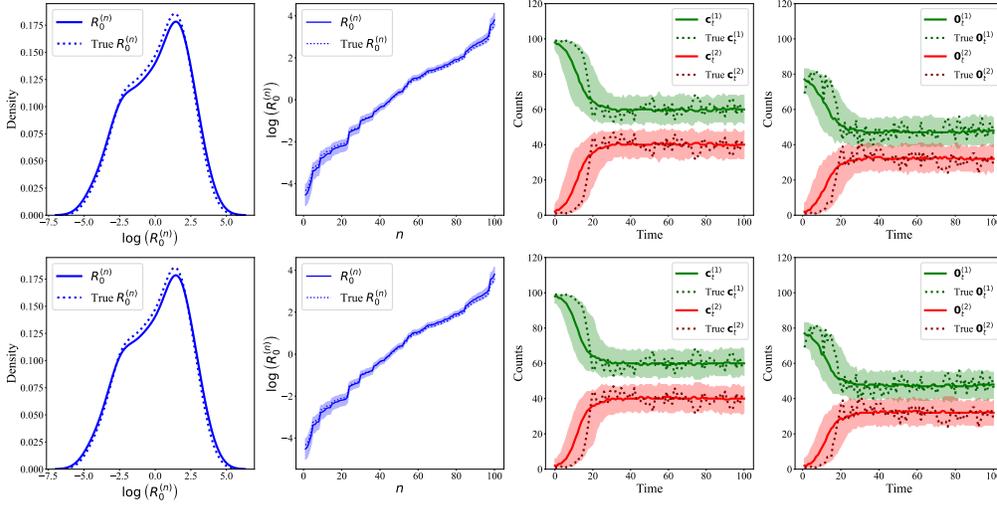


Figure 6: R_0 distribution and posterior predictive. First column distribution of the mean $R_0^{(n)}$. Second column $R_0^{(n)}$ and credible bands for each individual (sorted for increasing $R_0^{(n)}$). The third and fourth columns are posterior predictive over \mathbf{x}_t and \mathbf{y}_t . Rows from top to bottom $h = 5, 10$.

ing the individual-specific covariates. From \mathbf{w}_n we can define the initial distribution:

$$\mathbf{p}_{n,0} = \begin{bmatrix} 1 - \frac{1}{1 + \exp(-\beta_0^T w_n)} \\ 0 \\ \frac{1}{1 + \exp(-\beta_0^T w_n)} \\ 0 \end{bmatrix} \quad \text{for } n \in [1 : N] \text{ and } \beta_0 \in \mathbb{R}^d$$

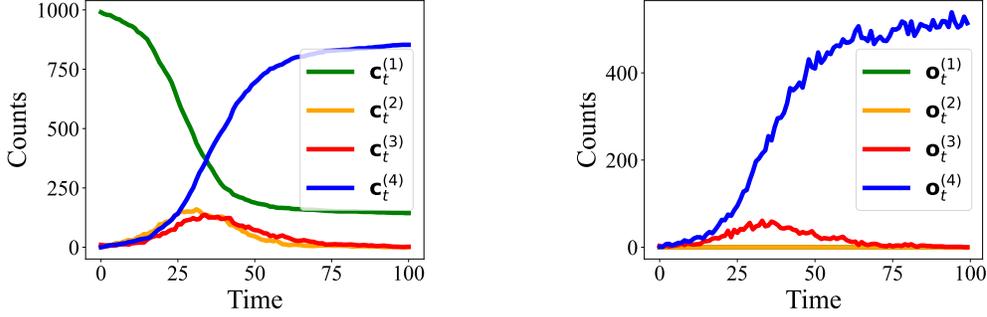


Figure 7: SEIR epidemic, on the left the unobserved process, on the right the observed data.

	h=5	0.9s	h=10	3.5s	h=20	5.45s	h=50	9.03s
	DGP	NDGP	DGP	NDGP	DGP	NDGP	DGP	NDGP
P	std	std	std	std	std	std	std	std
64	49.1	29.59	23.36	34.47	8.44	18.13	8.48	13.93
128	58.18	68.2	20.47	32.6	9.59	18.32	6.93	11.71
256	43.99	72.48	21.3	31.35	7.56	16.07	6.77	11.16
512	48.23	74.78	18.37	28.64	6.39	15.76	6.23	10.72
1024	45.64	66.12	19.36	29.75	5.71	14.1	5.38	10.5
2048	42.7	58.37	15.03	24.68	5.69	13.25	4.57	10.45

Table 3: Table reporting standard deviation for our method when $h = 5, 10, 20$ under DGP and NDGP with $P = 64, 128, 256, 512, 1024, 2048$. The mean computational cost is reported in the first row with the name of the algorithm.

and calculate $(\mathbf{K}_{n,\bullet})_{n \in [1:N]}$ as:

$$\mathbf{K}_{n,c} = \begin{bmatrix} 1 - \frac{1}{1 + \exp(-\beta_\lambda^\top w_n)} \frac{\mathbf{c}^{(3)}}{N} & \frac{1}{1 + \exp(-\beta_\lambda^\top w_n)} \frac{\mathbf{c}^{(3)}}{N} & 0 & 0 \\ 0 & \exp(-\rho) & 1 - \exp(-\rho) & 0 \\ 0 & 0 & 1 - \frac{1}{1 + \exp(-\beta_\gamma^\top w_n)} & \frac{1}{1 + \exp(-\beta_\gamma^\top w_n)} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

for $n \in [1 : N]$ and with $\beta_\lambda, \beta_\gamma \in \mathbb{R}^d$ and $\rho \in \mathbb{R}_+$. In this model, we have individual-specific probabilities of infection and recovery and a homogeneous latent period of $1/\rho$. Outcomes from the model are shown in figure

7.

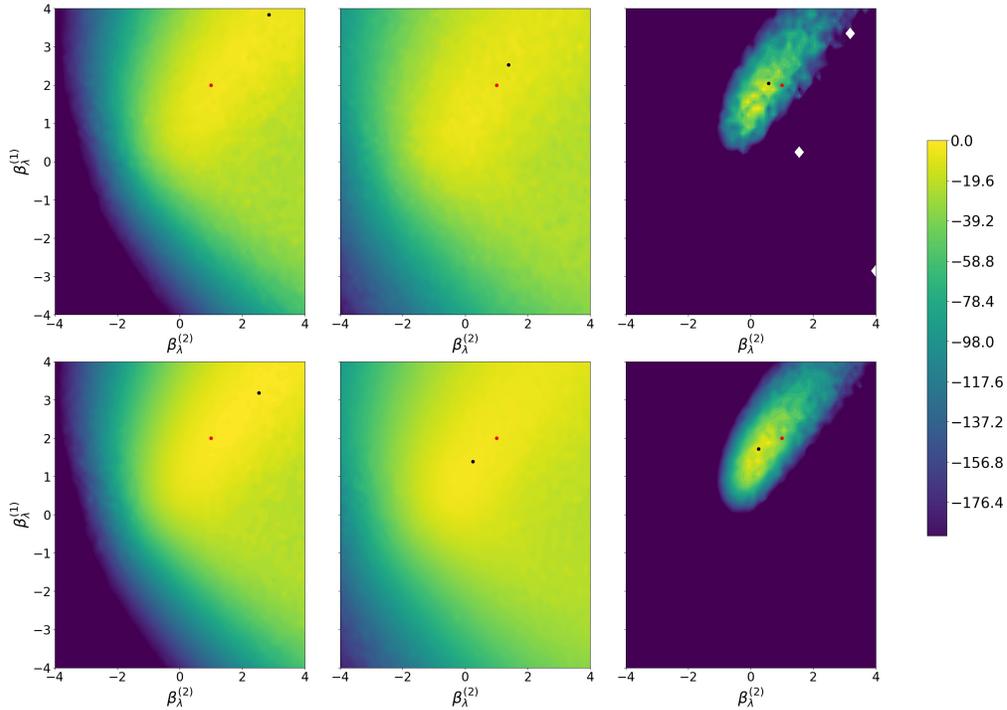


Figure 8: Marginal likelihood contour plot on a β_λ grid in log-scale. Columns refer to $t = 25, 50, 100$ from left to right. Rows refer to $h = 10, 20$ from top to bottom. The colorbar is common across the plots and in each plot, the maximum is set to 0. In red is the DGP and in black is the MLE on the grid.

The full table on the marginal likelihood standard deviation is reported

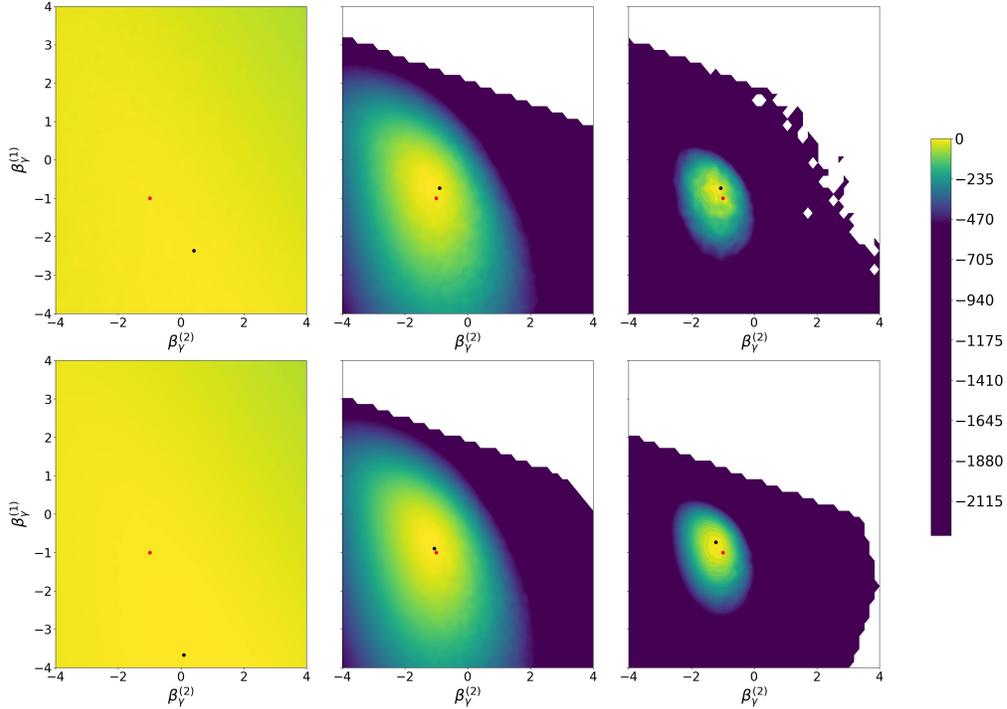


Figure 9: Marginal likelihood contour plot on a β_γ grid in log-scale. Columns refer to $t = 25, 50, 100$ from left to right. Rows refer to $h = 10, 20$ from top to bottom. The colorbar is common across the plots and in each plot, the maximum is set to 0. In red is the DGP and in black is the MLE on the grid.

in table 3. Complete marginal likelihood contour plots are reported in figure 8 and figure 9 in log scale and normalized to have their max in zero.

Sensitivity to \mathbf{q} is measured as for the SIS case, with data generated with $\mathbf{q} = [0, 0, 0.4, 0.6]^T$ and our algorithm is run using $P = 512$ and $\mathbf{q} = [0, 0, i, i]^T$ with $i \in [0.1, \dots, 0.9]$. Figure 10 reports the results for different

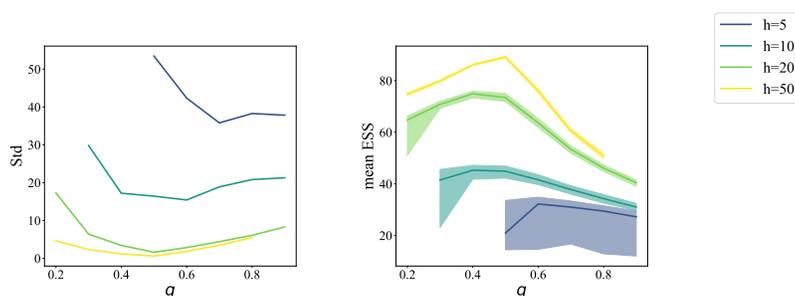


Figure 10: Effective sample size and marginal likelihood standard deviation sensitivity to q . The first column shows the marginal likelihood standard deviation, while the second column reports ESS where mode, 5%-quantile, and 95%-quantile are reported.

choices of h . As expected an increase in h is associated with a smaller standard deviation and a bigger mean ESS. It is also important to mention small values of h are more likely to fail even if they are close to the DGP, see $h = 5$, while higher values of h report $-\infty$ when they are far from the DGP, see $h = 50$ on the values 0.1, 0.9.

REFERENCES

References

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods.

Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72(3), 269–342.

Brauer, F. (2008). Compartmental models in epidemiology. In *Mathematical Epidemiology*, pp.

19–79. Springer.

Fintzi, J., J. Wakefield, and V. N. Minin (2021). A linear noise approximation for stochastic

epidemic models fit to partially observed incidence counts. *Biometrics*.

Ju, N., J. Heng, and P. E. Jacob (2021). Sequential Monte Carlo algorithms for agent-based

models of disease transmission. arXiv:2101.12156.

Keeling, M. J. and P. Rohani (2011). Modeling infectious diseases in humans and animals. In

Modeling infectious diseases in humans and animals. Princeton university press.

Lekone, P. E. and B. F. Finkenstädt (2006). Statistical inference in a stochastic epidemic SEIR

model with control intervention: Ebola as a case study. *Biometrics* 62(4), 1170–1177.

Whiteley, N. and L. Rimella (2021). Inference in stochastic epidemic models via multinomial

approximations. In *International Conference on Artificial Intelligence and Statistics*, pp.

1297–1305. PMLR.

Department of Mathematics and Statistics, Lancaster University, UK

E-mail: l.rimella@lancaster.ac.uk

Department of Mathematics and Statistics, Lancaster University, UK

Lorenzo Rimella, Christopher Jewell and Paul Fearnhead

E-mail: c.jewell@lancaster.ac.uk

Department of Mathematics and Statistics, Lancaster University, UK

E-mail: p.fearnhead@lancaster.ac.uk