

## S1 Text – Supporting text

### S1.1 Text

#### The Gillespie Algorithm

In the section “Predicting epidemic behavior”, we discussed that to use Eq. (11) to project the expected behavior of an epidemic outcome  $Z$  using the accrued observation  $Y_i$  at time  $t_i$ , one needs access to a stochastic simulator to sample from the random variable  $Z|\nu_i$  for a given set of parameter values  $\theta$ . We use the Gillespie algorithm [40] for this purpose. To describe the steps of the Gillespie algorithm, we consider an epidemic model with  $K$  compartments (e.g. in the model described in § ”Design of the Performance Analysis”,  $K = 4$ ). Let  $\nu(t) = (\nu^1(t), \nu^2(t), \dots, \nu^K(t))$  denote the state of the epidemic at time  $t$ , where  $\nu^k(t)$  denotes the number of individuals in compartment  $k \in \{1, 2, \dots, K\}$  at time  $t$ .

The state of an epidemic (i.e.  $\nu(t)$ ) may change due to the occurrence of specific events, such the transmission of the disease to a susceptible or the recovery from the disease. We use  $Q$  to denote the number of such events. In the model shown in Fig 3, these epidemic events include Infection, Seeking Treatment, and Recovery (and hence  $Q = 3$ ). Let  $\Lambda_q(\nu(t), \theta)$  denote the rate at which the event  $q$  is occurring when the epidemic is at state  $\nu(t)$  and with parameter values  $\theta$  that is a transformation of the parameter  $\theta$ .

Note that the stochastic model usually uses a volume dependent transformation of the parameter  $\theta$  from the ODE model. We do not need this transformation as our ODE description is volume independent.

In the model described in § ”Design of the Performance Analysis”, the force of infection is  $\theta_1 \nu^S(t) \frac{\nu^I(t)}{N(t)}$ , the rate of seeking treatment is  $\theta_2 \nu^I(t)$ , and the rate of recovery is  $\theta_3 \nu^T(t)$ , where  $\nu^S(t)$ ,  $\nu^I(t)$  and  $\nu^T(t)$  denote, respectively, the number of individuals in the compartments Susceptible, Infective, and Treatment (see Fig 3). The Gillespie algorithm proceeds as follows.

Assuming that the epidemic is at state  $\nu(t)$  at time  $t$ :

1. Calculate the sum of rates:  $\Lambda_0 = \sum_{q=1}^Q \Lambda_q(\nu(t), \theta)$  .
2. Sample the time to the next event:  $\tau \leftarrow -\frac{1}{\Lambda_0} \ln u_1$  , where  $u_1$  is a random number drawn from a uniform distribution  $\mathcal{U}((0, 1])$ .
3. Determine the event that just realized: Select the event  $\tilde{q}$  such that

$$\sum_{q=1}^{\tilde{q}-1} \frac{\Lambda_q(\nu(t), \theta)}{\Lambda_0} < u_2 \leq \sum_{q=\tilde{q}}^Q \frac{\Lambda_q(\nu(t), \theta)}{\Lambda_0}$$

where  $u_2$  is another random number drawn from a uniform distribution  $\mathcal{U}((0, 1])$ .

4. Update epidemic state given the realized event  $\tilde{r}$ .
5. Until a desired time condition is satisfied, increment time  $t \leftarrow t + \tau$  and move to Step 1.

**Figure SI 1.** Pseudo-code of Gillespie algorithm for simulating stochastic compartmental models.

To obtain simulation trajectories displayed in Fig 4 A and Figs S1 Fig A to S8 Fig A,

we use the Gillespie algorithm (Fig SI 1) to simulate the SITR compartmental model shown in Fig 1.

These SITR trajectories are used to calculate the number of new diagnoses in week  $i$  which is equal to  $T_i + R_i - T_{i-1} - R_{i-1}$ . If for a simulated trajectory the conditions on attack rate (30% - 50% for the mild scenario, 50%-70% for the severe scenario and 70%-100% for the extreme scenario) and timing of peak (between 10 and 20) are fulfilled, we keep the trajectory for the respective scenarios or we otherwise we reject it.

### S1.2 Text

#### Equations of the SEITR model

The Susceptible - Exposed - Infected - Treatment - Recovered model contains one more compartment than the SITR model, namely Exposed. Disease transmission can be modeled using the following ODE model:

$$\begin{aligned}\frac{d}{dt}x_S(t) &= -k_1x_S(t)\frac{x_I(t)}{N(t)}, \\ \frac{d}{dt}x_E(t) &= k_1x_S(t)\frac{x_I(t)}{N(t)} - k_2x_E(t), \\ \frac{d}{dt}x_I(t) &= k_2x_E(t) - \theta_2x_I(t), \\ \frac{d}{dt}x_T(t) &= \theta_2x_I(t) - \theta_3x_t(t), \\ \frac{d}{dt}x_R(t) &= \theta_3x_T(t)\end{aligned}\quad (17)$$

where  $k_1 = k_2 = 2\theta_1$  and  $\theta_1$  is the disease transmission rate,  $\theta_2$  is the rate of seeking treatment while infectious, and  $\theta_3$  is the rate of recovering.

In our analysis, we assume that at time  $t = 0$  one population member becomes infected. The model initial condition can therefore be defined as  $x_0 = (x_S(0), x_E(0), x_I(0), x_T(0), x_R(0)) = (N - 1, 1, 0, 0)$ . We furthermore assume a constant population size  $N(t) = x_S(t) + x_E(t) + x_I(t) + x_T(t) + x_R(t)$  denote the population size at time  $t$ .

### S1.3 text

**Formulas to calculate integrated relative errors (IRE) for different metrics**

The exact form of  $f_{M_i|\theta}(m|\theta)$  of the IREs in Eq. (16) depends on the metric of interest.

**$R_0$  and duration of infectiousness:** For  $R_0$  or the mean duration of infectiousness, we have  $f_{M_i|\theta}(m|\theta) = 1$ , as  $R_0$  and the mean duration of infectiousness only depend on the parameter vector sampled from the posterior  $\pi_i$ .

**$R_{\text{eff}}$ :** For  $R_{\text{eff}}$  as the performance target,  $M_i$  is defined as  $M_i = R_0 \frac{S_i}{N}$ . As  $S$  is the only state that the target depends on, the function  $f_{M_i|\theta}(m|\theta)$  needs to yield the probability for this state. This is accomplished by integrating over the remaining states:

$$f_{M_i|\theta}(m|\theta) = \int_{(\nu_i^{(S)}, \nu_i^{(I)}, \nu_i^{(T)}, \nu_i^{(R)}) \in \Omega_i} \Pi_i \left( (\nu_i^{(I)}, \nu_i^{(T)}, \nu_i^{(R)}) | Y_i; \theta \right)$$

**Infection prevalence:** If the target is the infection prevalence:  $M_i = \nu_i^{(S)}$ , the function  $f_{M_i|\theta}(m|\theta)$  needs to yield the probability for this state and, hence, is calculated by integrating over the remaining states.

$$f_{\nu_i^{(S)}|\theta}(\nu_i^{(I)}|\theta) = \int_{(\nu_i^{(S)}, \nu_i^{(T)}, \nu_i^{(R)}) \in \Omega_i} \Pi_i \left( (\nu_i^{(S)}, \nu_i^{(T)}, \nu_i^{(R)}) | Y_i; \theta \right)$$

As the I.Poi benchmark does not calculate a belief state for time  $t_i$ , we use the simulation model to carry out simulations with parameters according to  $\pi_i(\theta|Y_i)$  and consider these simulations as a sample from the belief state  $\Pi_i$ .

**All Predictions:** In case of prediction for one week, three weeks (cumulative and specific) or the attack rate, it holds

$$f_{M_i|\theta}(m|\theta) = P(m|Y_i)$$

with  $P$  as in Eq. (9).

**Median integrated relative error (mIRE):** Next, we will calculate a median relative error (mIRE) over the 50 trajectories as

$$mIRE(M_i) = \text{median} \left( IRE(M_i^{(1)}), \dots, IRE(M_i^{(50)}) \right).$$

### S1.4 text

#### Computational Effort

As mentioned in the Discussion, the most computationally challenging parts are the ODE system integration for MSS, PF and EnKF and the stochastic simulation for I.Poi. We analyzed the computational time for the severe scenario with a population size of 10000. Carrying out the 1000 stochastic simulations for benchmark I.Poi for the 1000 parameters takes about 100 minutes. The additional time for evaluating the likelihood approximation is rather short with about only 1 second, see table 1 for numbers.

The computational times are shorter before the peak and longer after the peak, when less or more data is taken into account. The computational time will only vary for

the I.Poi. when changing the population size as the other methods do not require  
758  
stochastic simulations.  
759

We note that the implementation was not carried out in a way to maximize speed, so  
760  
there is opportunity for additional speed up with more efficient implementation.  
761  
However, these numbers demonstrate that all the algorithms can be used on a personal  
762  
computer in real time.  
763

**Table 1. Computational time for MSS and benchmarks:**

time to peak	MSS	I.Poi.	PF	EnKF
8 weeks to peak	$85.2 \pm 32.8$	$100^* \text{ min} + 0.6 \pm 0.3$	$7.7 \pm 4.1$	$10.6 \pm 5.9$
4 weeks to peak at peak	$125.2 \pm 33.2$ $157.6 \pm 31.6$	$100^* \text{ min} + 1 \pm 0.3$ $100^* \text{ min} + 1.4 \pm 0.3$	$12.5 \pm 4.1$ $17.4 \pm 4.$	$17.8 \pm 6.$ $26.4 \pm 6.3$
4 weeks after peak	$196.2 \pm 33.8$	$100^* \text{ min} + 1.7 \pm 0.3$	$22.6 \pm 4.$	$34.5 \pm 6.6$
8 weeks after peak	$228.8 \pm 35.$	$100^* \text{ min} + 2.1 \pm 0.3$	$27.5 \pm 4.1$	$41.8 \pm 6.3$

Computational time in seconds for the calibration of one time course in average for the  
severe scenario in the population with 10000 individuals.

\* The 100 min are the computational time to carried out 1000 stochastic simulations for  
each parameter with the Gillespie algorithm.

**S1.5 text**

764

**Pseudo code for the benchmark method A (I.Poi)**

765

1. Initialization
  - (a) Choose an initial prior probability function  $\pi_0(\theta)$ .
  - (b) Choose the set  $\Theta$  that includes the values of parameters  $\theta$  for which the approximate likelihood function (10) should be calculated.
  - (c) Set  $\ln L^{(I.Poi)}(Y_0; \theta) \leftarrow 0$  for every  $\theta \in \Theta$ .
2. Calibration: For each observation  $y_i$ ,  $i = 1, \dots, n$ ,
  - (a) Assume  $\mathcal{P}^{(I.Poi)}(\cdot|\theta) \sim Poisson(\mu_i)$  and use simulated trajectories to estimate the mean  $\mu_i$ , for every  $\theta \in \Theta$ .
  - (b) Update the likelihood function:  $L^{(I.Poi)}(Y_i; \theta) \leftarrow L^{(I.Poi)}(Y_{i-1}; \theta) \times \mathcal{P}^{(I.Poi)}(y_i|\theta)$  for every  $\theta \in \Theta$ .
  - (c) Update the parameter posterior distribution:  $\pi_i(\theta|Y_i) \leftarrow L^{(I.Poi)}(y_i|\theta) \pi_{i-1}(\theta, Y_{i-1})$  for every  $\theta \in \Theta$ .

**Figure SI 2.** Pseudo code for the benchmark method A (I.Poi) [23].

**S1.6 text**

766

**Pseudo code for the Particle Filter**

767

1. Initialization
  - (a) Choose an initial prior probability function  $\pi_0(\theta)$ .
  - (b) Choose the set  $\Theta$  that includes the values of parameters  $\theta$  for which the likelihood function (1) should be calculated with the specifications in Eq. (12) and name the specific likelihood approximation  $L^{(PF)}$ .
  - (c) Set  $\ln L^{(PF)}(Y_0; \theta) \leftarrow 0$  for every  $\theta \in \Theta$ .
2. Calibration: For each observation  $y_i$ ,  $i = 1, \dots, n$ ,
  - (a) Calculate the probability  $\mathcal{P}^{(PF)}(y_i|Y_{i-1}; \theta)$  for every  $\theta \in \Theta$ .
  - (b) Update the likelihood function:  $L^{(PF)}(Y_i; \theta) \leftarrow L^{(PF)}(Y_{i-1}; \theta) \times \mathcal{P}^{(PF)}(y_i|Y_{i-1}; \theta)$  for every  $\theta \in \Theta$ .
  - (c) Update the parameter posterior distribution:  $\pi_i(\theta|Y_i) \leftarrow L^{(PF)}(y_i|\theta) \pi_{i-1}(\theta, Y_{i-1})$  for every  $\theta \in \Theta$ .

**Figure SI 3.** Pseudo code for the Particle Filter [27]

**S1.7 text**

768

**Pseudo code for the Ensemble Kalmann filter**

769

1. Initialization

- (a) Choose an initial prior probability function  $\pi_0(\theta)$ .
- (b) Choose the set  $\Theta$  that includes the values of parameters  $\theta$  for which the likelihood function (1) should be calculated with the specifications in Eq. (12).
- (c) Choose an initial set of particles  $\Psi$ . Its elements  $\psi \in \Psi$  contain values for the parameters and states:  $\psi = (\theta, \nu)$ .

2. Calibration: For each observation  $y_i, i = 1, \dots, n$ ,

- (a) Propagate each particle by setting  $\psi \leftarrow (\theta, \nu_{prior})$  with  $\nu_{prior} = x(t_i - t_{i-1}, \nu; \theta)$ .
- (b) Calculate the prior for the observations:  $\psi^{(y)} = \nu_{prior}^{(T)} + \nu_{prior}^{(R)} - \nu^{(T)} - \nu^{(R)}$ .
- (c) Calculate the variance  $\sigma_{obs,i}^2$  as in Eq. (13).
- (d) Calculate the prior variance of the observed quantity as  $\sigma_{prior}^2 = \text{Variance } (\psi^{(y)})$ .
- (e) Calculate the prior co-variance of the observed quantity and each unobserved component of the particle  $\psi$  as  $\sigma_m = \text{co-variance } (\psi^{(y)}, \psi^{(m)})$ , where  $m$  denotes the unobserved component.
- (f) Update  $\psi \leftarrow \frac{\sigma_m}{\sigma_{prior}^2} \delta$  with

$$\delta = \frac{\sigma_{obs}^2}{\sigma_{obs}^2 + \sigma_{prior}^2} \psi_{prior}^y + \frac{\sigma_{prior}^2}{\sigma_{obs}^2 + \sigma_{prior}^2} \tilde{y}_i$$

and  $\tilde{y}_i = y_i + \xi_i$ , with  $\xi_i \sim N(0, \sigma_{obs,i}^2)$ :

- (g) Set  $(\theta, \nu) \leftarrow \psi$  for the next iteration.

**Figure SI 4.** Pseudo code for the Ensemble Kalmann filter [27].

### S1.8 text

#### Detailed pseudo code for our MSS for SITR

##### 1. Input:

- ODE solution  $x(t, x_0; \theta)$  for an ODE  $x'$  integrated for time  $t$  with initial value  $x_0$  and parameter  $\theta$  such as in equation (5). 773
- Data set  $y_1, \dots, y_n$  775
- Number of ensemble members  $N_{grid}$ . 776
- Total populations size  $N$ . 777

##### 2. Initialization:

```

FOR j=1,...,Ngrid 778
    θ(j) = (R0(j), mdi(j), γ(j)) ~ U(R0(range) × mdi(range) × γ(range)) 779
        % Draw a parameter sample from an uninformative prior; mdi - mean duration of infectiousness; 780
        % γ - additional parameter for the number of weeks, the epidemics started before the first observation 781
        % in our case R0(range) = [1, 3], mdi(range) = [1, 20], γ(range) = [0, 5]. 782
    S-γ(j) = N - 1 % initialize states 783
    I-γ(j) = 1 % we assume one infected at start, but no knowledge when epidemic started 784
    T-γ(j) = R-γ(j) = 0 785
    ν̂-γ(j) = (S-γ(j), I-γ(j), T-γ(j), R-γ(j)) 786
    L(Y-γ | θ(j)) = 1 787
    y-γ = y-γ+1 = ... = y0 = 0 % auxiliary observations to model the time before the first case 788
END FOR 789

```

##### 3. Loop:

```

FOR i = -γ, ..., n % epidemic stared at -γ and data available until n 790
    FOR j = 1, ..., Ngrid 791
        % % % Forward integration
        ν̂i(j) = x(ti - ti-1, ν̂i-1(j); θ(j)) % ODE integration 792
        Covi(j) = Σ(ti - ti-1, ν̂i-1(j); θ(j)) % LNA co-variance integration 793
        Covi,(k,k)(j) = max(Covi,(k,k)(j), α) for k ∈ {S, I} % ensure that no negative co-variance, we use α = 10-6 794
        % % % calculate state estimate in case of observation noise with standard deviation σ2
        % The state estimate is calculated by maximizing the probability in equation 7. 795
        % LNA ⇒ p normal and P(yi | νi, ν̂i-1; θ) normal as gaussian observation noise 796
        % taking into account that ν̂i(j,T) + ν̂i(j,R) - ν̂i-1(j,T) + ν̂i-1(j,R) would correspond to a noise free observation 797
        % ⇒ maxνi (2π)-ndim/2 (det(Covi(j)))-1/2 × exp(-(ν̂i(j) - ν̂i(j))(Covi(j))-1(ν̂i(j) - ν̂i(j))) 798
        % % % calculate state estimate in case of observation noise with standard deviation σ2
        % As the 1st, 2nd and 4th factor do not contain νi and max is equal to min(-log(.)) , define h as 799

```

$$h(\hat{\nu}_i^{(j)}) = (\hat{\nu}_i^{(j)} - \tilde{\nu}_i^{(j)}) \cdot (Cov_i^{(j)})^{-1} \cdot (\hat{\nu}_i^{(j)} - \tilde{\nu}_i^{(j)}) + \frac{((\hat{\nu}_i^{(j,T)} + \hat{\nu}_i^{(j,R)} - \hat{\nu}_{i-1}^{(j,T)} + \hat{\nu}_{i-1}^{(j,R)}) - y_i)^2}{\sigma^2}$$
805  
 % and minimize by solving derivative equals 0.  
806  

$$\begin{cases} \frac{\partial h(\hat{\nu}_i^{(j)})}{\partial \hat{\nu}_i^{(j,S)}} = 0 \\ \frac{\partial h(\hat{\nu}_i^{(j)})}{\partial \hat{\nu}_i^{(j,I)}} = 0 \\ \frac{\partial h(\hat{\nu}_i^{(j)})}{\partial \hat{\nu}_i^{(j,T)}} = 0 \end{cases}$$
% solve the optimization problem analytically for  $\hat{\nu}_i^{(j,S)}$ ,  $\hat{\nu}_i^{(j,I)}$  and  $\hat{\nu}_i^{(j,T)}$   
807

%% calculate state estimate in case of no observation noise  
808  
 % The state estimate is calculated by maximizing the probability in equation 7.  
809  
 % LNA  $\Rightarrow p$  normal  
810  
 % taking into account that  $y_i = \hat{\nu}_i^{(j,T)} + \hat{\nu}_i^{(j,R)} - \hat{\nu}_{i-1}^{(j,T)} + \hat{\nu}_{i-1}^{(j,R)}$  (noise free observation), reduces one  
811  
 dimension of the optimization problem  
812  

$$h(\hat{\nu}_i^{(j)}) = (g(\hat{\nu}_i^{(j)}) - \tilde{\nu}_i^{(j)}) \cdot (Cov_i^{(j)})^{-1} \cdot (g(\hat{\nu}_i^{(j)}) - \tilde{\nu}_i^{(j)})$$
% define function h  
813  

$$g(\hat{\nu}_i^{(j)}) := \begin{cases} N - \hat{\nu}_{i-1}^{(j,T)} - \hat{\nu}_i^{(j,S)} - \hat{\nu}_{i-1}^{(j,R)} - y_i \\ \hat{\nu}_i^{(j,T)} \end{cases}$$
% function g taking into account the relation of  $y_i$   
814  

$$\begin{cases} \frac{dh(\hat{\nu}_i^{(j)})}{d\hat{\nu}_i^{(j,S)}} = 0 \\ \frac{dh(\hat{\nu}_i^{(j)})}{d\hat{\nu}_i^{(j,T)}} = 0 \end{cases}$$
% solve the optimization problem analytically for  $\hat{\nu}_i^{(j,S)}$  and  $\hat{\nu}_i^{(j,T)}$   
815

%% Calculate state specific bounds in case of observation noise with standard deviation  $\sigma^2$   
816  

$$\hat{\nu}_i^{(j,S)} = \min \left\{ \max \left\{ 0, \hat{\nu}_i^{(j,S)} \right\}, N \right\}$$
% within the bounds of 0 and population size  
817  

$$\hat{\nu}_i^{(j,R)} = \min \left\{ \max \left\{ 0, \hat{\nu}_i^{(j,R)} \right\}, N \right\}$$
  
818  

$$\hat{\nu}_i^{(j,I)} = \min \left\{ \max \left\{ 0, \hat{\nu}_i^{(j,I)} \right\}, N \right\}$$
  
819  
 %% Calculate state specific bounds in case of no observation noise  
820  

$$\hat{\nu}_i^{(j,S)} = \min \left( \hat{\nu}_i^{(j,S)}, N - 1, N - \hat{\nu}_{i-1}^{(j,T)} - \hat{\nu}_{i-1}^{(j,R)} - y_i - \alpha, N - \sum_{l=1}^i y_l - 1 \right)$$
% we use  $\alpha = 10^{-6}$   
821  
 % upper bounds: population size, population size - already diagnosed cases - current cases,  
822  
 % population size - all cases; whichever is lowest  
823  

$$\hat{\nu}_i^{(j,T)} = \min \left( \max \left( 0, \hat{\nu}_i^{(j,T)} \right), \sum_{l=1}^i y_l \right)$$
% no more people in T than already diagnosed cases  
824  

$$\hat{\nu}_i^{(j,I)} = N - \hat{\nu}_{i-1}^{(j,T)} - \hat{\nu}_i^{(j,S)} - \hat{\nu}_{i-1}^{(j,R)} - y_i$$
  
825

%% calculate likelihood  
826  
 % need to integrate over state space  $\rightarrow$  sampling based  
827  
 FOR  $l = 1, \dots, n_r : z_l^{(j)} \sim \mathcal{N}(\hat{\nu}_i^{(j)}, Cov_i^{(j)})$  END FOR  
%  $n_r = 10000$   
828  

$$newlys_{(i,l)}^{(j)} = z_l^{(j,T)} - \hat{\nu}_{i-1}^{(j,T)} + (N - z_l^{(j,S)} - z_l^{(j,I)} - z_l^{(j,T)}) - \hat{\nu}_{i-1}^{(j,R)}$$
  
829  

$$dist = \mathcal{N}_{truncated(-10^{-10}, N)} \left( \text{mean} \left( newlys_{(i,l=1,\dots,n_r)}^{(j)} \right), \text{Variance} \left( newlys_{(i,l=1,\dots,n_r)}^{(j)} \right) + \sigma^2 \right)$$
  
830  

$$L(y_i | \theta^{(j)}) = PDF(dist, y_i)$$
  
831  
 END FOR  
832

$$L^{(\text{MSS})}(Y_i | \theta^{(j)}) = L^{(\text{MSS})}(Y_{i-1} | \theta^{(j)}) \times L^{(\text{MSS})}(y_i | \theta^{(j)})$$
  
833

END FOR

4. **Output:** sample of parameters  $\theta^{(j)}$  with corresponding likelihood values  $L(Y_n|\theta^{(j)})$  and belief states.

834

835

**S1.9 text**

836

**Detailed pseudo code for I.Poi. for SITR**

837

**1. Input:**

838

- Stochastic simulation model  $H(t, x_0; \theta)$  simulating a trajectory until time  $t$  with an initial value  $x_0$  and a parameter  $\theta$ .  
839
- Data set  $y_1, \dots, y_n$   
840
- Number of ensemble members  $N_{grid}$ .  
841
- Total populations size  $N$ .  
842
- Number of simulation being carried out for each parameter value  $N_{prior,sim}$ .  
843

844

**2. Initialization:**

845

```

FOR j=1,...,Ngrid
    θ0(j) = (R0(j), mdi0(j)) ~ U(R0(range) × mdi(range))      % mdi - mean duration of infectiousness
    % in our case R0(range) = [1, 3], mdi(range) = [1, 20], γ(range) = [0, 5].
    S0(j) = N - 1
    I0(j) = 1
    T0(j) = R0(j) = 0
    ν0(j) = (S0(j), I0(j), T0(j), R0(j))
    L(Y0|θ(j)) = 1
END FOR

```

846

847

848

849

850

851

852

853

854

**3. Calibration:**

855

```

FOR i = 1,...,n
    FOR j=1,...,Ngrid
        For k = 1,...,Nprior,sim
            zi-1(j,k) = H(k)(ti-1, ν0; θ(j))          % previous state
            zi(j,k) = H(k)(ti, ν0; θ(j))          % current state
            newlyi(j,k) = zi(j,k,T) + zi(j,k,R) - zi-1(j,k,T) - zi-1(j,k,R)
        END FOR k
        me = mean (newlyi(j,1), ..., newlyi(j,Nprior,sim))      % calculate mean for Poisson distribution
        IF me > 0                                     % Poisson distribution needs parameter greater 0
            L(yi|θ(j)) = PDF(PoissonDistribution(mean), yi)          % if so, evaluate
        ELSE: IF yi = 0 then 1 ELSE α END IF % if not, check whether yi = 0, we use α = 10-100.
        END IF
    END FOR
    L(I.Poi)(Yi|θ(j)) = L(I.Poi)(Yi-1|θ(j)) × L(I.Poi)(yi|θ(j))          % update likelihood
END FOR

```

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

**4. Output:** sample of parameters  $\theta$  with corresponding likelihood values  $L(Y_i|\theta)$ .

871

**S1.10 text**

872

**Detailed pseudo code for Particle Filter for SITR**

873

**1. Input:**

- ODE solution  $x(t, x_0; \theta)$  for an ODE  $x'$  integrated for time  $t$  with initial value  $x_0$  and parameter  $\theta$  such as in equation (5). 875
- Data set  $y_1, \dots, y_n$  877
- Number of ensemble members  $N_{grid}$ . 878
- Total populations size  $N$ . 879

**2. Initialization:**

```

FOR j=1,...,Ngrid 880
    θ0(j) = (R0(j), mdi0(j)) ~ U(R0(range) × mdi(range)) 881
    % mdi - mean duration of infectiousness
    % in our case R0(range) = [1, 3], mdi(range) = [1, 20], γ(range) = [0, 5]. 882
    S0(j) ~ U([̄αN, N]) 883
    % we use ̄α = 0.9
    I0(j) ~ U([0, N - S0(j)]) 884
    T0(j) ~ U([0, N - S0(j) - I0(j)]) 885
    R0(j) ~ U([0, N - S0(j) - I0(j) - T0(j)]) 886
    ̂ν0(j) = (S0(j), I0(j), T0(j), R0(j)) 887
    w0(j) = 1/Ngrid 888
END FOR 889

```

**3. Loop:**

```

FOR i=1,...,n 890
    Wi-1 = ∑j=1Ngrid wi-1(j) 891
    FOR j=1,...,Ngrid: ̂wi-1(j) = wi-1(j)/Wi-1  END FOR 892
    % normalize weights
    neff = 1/ ∑j=1Ngrid (̂wi-1(j))2 893
    IF neff < Ngrid/2  % check whether to re-sample 894
        n1 = 7
        C1 = Πn1/2/Gamma(n1/2 + 1)  % with the gamma function Gamma 895
        A1 = (8/C1(n1 + 4) * (2√Π)n1)-1/(n1+4) 896
        hopt = 2 * A1 * Ngrid-1/(n1+4) 897
        xK = (-1, -0.999, ..., 0.999, 1) 898
        dK = ((n1 + 2)/2/C1) * (1 - xK2)  % kernel 899
        cK = dK / ∑j=1length(dK) dKj 900
        FOR j = 1, ..., length(dK): cKj = cKj-1 + dKj  END FOR 901
        f: function interpolating (xK, cK) 902
        ψi-1(j) drawn from { (θi-1(1), ̂νi-1(1)), ..., (θi-1(Ngrid), ̂νi-1(Ngrid) ) } with weights { ̂wi-1(1), ..., ̂wi-1(Ngrid) } 903
        ̂wi-1(j) = 1/Ngrid, j = 1, ..., Ngrid 904

```

```

 $\sigma_{i-1} = StdDev \left( \psi_{i-1}^{(1)}, \dots, \psi_{i-1}^{(N_{grid})} \right)$  % vector of standard deviations 908

IF length  $\left( \cup_{j=1}^{N_{grid}} \left\{ \psi_{i-1}^{(j)} \right\} \right) < \max(\alpha_1, 0.01N_{grid})$  % we use  $\alpha_1 = 20.$  909
 $\sigma_{i-1} = StdDev \left( \psi_{\max(i-2-1,0)}^{(1)}, \dots, \psi_{\max(i-2-1,0)}^{(N_{grid})} \right)$  910
 $\sigma_{i-1}^{(I)} = \max \left( \sigma_{i-1}^{(I)}, y_i/5 \right)$  911
END IF 912

FOR  $k = 1, \dots, n_1 :$  913
 $z_k = f(\Xi)$  with a  $\Xi \sim UniformDistribution([0, 1]^{N_{grid}})$  914
END FOR 915
 $\psi_{i-1}^{(j)} = \psi_{i-1}^{(j)} + h_{opt} z SD$  % with SD a diagonal matrix with entries  $\sigma_{i-1}$  regularization noise 916

FOR  $j = 1, \dots, N_{grid}$  % check boundaries 917
 $\hat{\nu}_{i-1}^{(j,m)} = \max \left( 0, \min \left( N, \psi_{i-1}^{(j,m)} \right) \right), m \in \{S, I, T, R\}$  918
 $\theta_{i-1}^{(j,R_0)} = \max \left( 1, \min \left( R_0^{(\text{range,up})}, \psi_{i-1}^{(j,R_0)} \right) \right)$  919
 $\theta_{i-1}^{(j,mdi)} = \max \left( 1, \min \left( mdi^{(\text{range,up})}, \psi_{i-1}^{(j,mdi)} \right) \right),$  920
END FOR 921
ELSE 922
nothing % do not change  $\theta$  or  $\hat{\nu}.$  923
END IF 924

% % % Forward Propagation 925
FOR  $j = 1, \dots, N_{grid}$  926
 $\hat{\nu}_i^{(j)} = x(t_i - t_{i-1}, \hat{\nu}_{i-1}^{(j)}; \theta_{i-1}^{(j)})$  % forward propagation 927
 $\theta_i^{(j)} = \theta_{i-1}^{(j)}$  928
END FOR 929

% % % Assign weights 930
FOR  $j = 1, \dots, N_{grid}$  931
 $y_{i,prior}^{(j)} = \hat{\nu}_i^{(j,T)} + \hat{\nu}_i^{(j,R)} - \hat{\nu}_{i-1}^{(j,T)} - \hat{\nu}_{i-1}^{(j,R)}$  % prior observations 932
 $\sigma_{i,obs}^2 = 10000 + \frac{1}{50} \left( \frac{1}{3} \sum_{j=i-3}^{i-1} y_j \right)^2.$  % observation variance 933
 $w_i^{(j)} = \tilde{w}_{i-1}^{(j)} \times PDF \left( \mathcal{N} \left( y_{i,prior}^{(j)}, \sigma_{i,obs}^2 \right), y_i \right)$  934
END FOR 935
END FOR 936

```

4. **Output:** posterior sample of parameters  $\theta_n$  and states  $\hat{\nu}_n.$

937

### S1.11 text

938

#### Detailed pseudo code for the Ensemble Kalman filter for SITR

939

##### 1. Input:

940

- ODE solution  $x(t, x_0; \theta)$  for an ODE  $x'$  integrated until time  $t$  with initial value  $x_0$  and parameter  $\theta$  such as in equation (5). 941
- Data set  $y_1, \dots, y_n$  942
- Number of ensemble members  $N_{grid}$ . 943
- Total populations size  $N$ . 944
- 945

##### 2. Initialization:

946

```

FOR j=1,...,Ngrid
     $\theta_0^{(j)} = (R_0^{(j)}, mdi_0^{(j)}) \sim U(R_0^{(\text{range})} \times mdi^{(\text{range})})$  % mdi - mean duration of infectiousness 947
    % in our case  $R_0^{(\text{range})} = [1, 3]$ ,  $mdi^{(\text{range})} = [1, 20]$ ,  $\gamma^{(\text{range})} = [0, 5]$ . 948
     $S_0^{(j)} \sim U([\bar{\alpha}N, N])$  % we use  $\bar{\alpha} = 0.9$ . 949
     $I_0^{(j)} \sim U([0, N - S_0^{(j)}])$  950
     $T_0^{(j)} \sim U([0, N - S_0^{(j)} - I_0^{(j)}])$  951
     $R_0^{(j)} \sim U([0, N - S_0^{(j)} - I_0^{(j)} - T_0^{(j)}])$  952
     $\hat{\nu}_0^{(j)} = (S_0^{(j)}, I_0^{(j)}, T_0^{(j)}, R_0^{(j)})$  953
     $\psi_0^{(j)} = (\hat{\nu}_0^{(j)}, \theta_0^{(j)}, y_0 = 0)$  954
END FOR 955

```

956

##### 3. Loop:

957

```

FOR i=1,...,n
    % % Forward propagation 958
    FOR j=1,...,Ngrid
         $\nu_{i,prior}^{(j)} = x(t_i - t_{i-1}, \hat{\nu}_{i-1}^{(j)}; \theta_{i-1}^{(j)})$  % propagation 959
    END FOR 960
    FOR j=1,...,Ngrid
         $\nu_{i,prior}^{(j)} = \alpha' \left( \nu_{i,prior}^{(j)} - \text{mean} \left( \nu_{i,prior}^{(1)}, \dots, \nu_{i,prior}^{(N_{grid})} \right) \right) + \nu_{i,prior}^{(j)}$  961
        % Inflation with an inflation parameter  $\alpha' = 1.07$  962
         $\psi_i^{(y,j)} = \nu_{i,prior}^{(j,T)} + \nu_{i,prior}^{(j,R)} - \hat{\nu}_{i-1}^{(j,T)} - \hat{\nu}_{i-1}^{(j,R)}$  % prior observations 963
         $\sigma_{i,obs}^2 = 10000 + \frac{1}{50} \left( \frac{1}{3} \sum_{j=i-3}^{i-1} y_j \right)^2$  % observation variance 964
         $\sigma_{i,prior}^2 = \text{variance} \left( \psi_i^{(1,y)}, \dots, \psi_i^{(N_{grid},y)} \right)$  % prior variance 965
         $\sigma_i^{(m)} = \text{co-variance} \left( \psi_i^y, \nu_{i,prior}^{(m)} \right), m \in \{R_0, mdi, S, I, T, R\}$  % co-variance observation prior 966
         $\tilde{y}_i^{(j)} = \max \left( 0, y_i + \xi_i^{(j)} \right)$  with  $\xi_i^{(j)} \sim \mathcal{N}(0, \sigma_{i,obs}^2)$  % noisy observation 967
    END FOR 968
    % % Update 969
    FOR j=1,...,Ngrid
         $\delta_i^{(j)} = \frac{\sigma_{i,obs}^2}{\sigma_{i,obs}^2 + \sigma_{i,prior}^2} \psi_i^{(y,j)} + \frac{\sigma_{i,prior}^2}{\sigma_{i,obs}^2 + \sigma_{i,prior}^2} \tilde{y}_i^{(j)}$ . 970
    END FOR 971

```

972

973

974

```

 $\theta_i^{(j,m)} = \frac{\sigma_i^{(m)}}{\sigma_{i,prior}^2} \delta_i^{(j)} + \theta_{i-1}^{(j,m)}, m \in \{R_0, mdi\}$  975
 $\hat{\nu}_i^{(j,m)} = \frac{\sigma_i^{(m)}}{\sigma_{i,prior}^2} \delta_i^{(j)} + \hat{\nu}_{i-1}^{(j,m)}, m \in \{S, I, T, R\}$  976
END FOR 977

% % Check boundaries 978
FOR j=1,...,Ngrid 979
     $\hat{\nu}_i^{(j,S)} = \text{IF } \hat{\nu}_i^{(j,S)} > N : N - 1 \text{ ELSE } \hat{\nu}_i^{(j,S)} \text{ END IF}$  980
     $\hat{\nu}_i^{(j,m)} = \text{IF } \hat{\nu}_i^{(j,m)} > N : \text{median}(\hat{\nu}_i^{(j=1,...,N_{grid},m)}) \text{ ELSE } \hat{\nu}_i^{(j,m)}, m \in \{I, T, R\} \text{ END IF}$  981
     $\theta_i^{(j,R_0)} = \text{IF } \theta_i^{(j,R_0)} > R_0^{(\text{range,up})} : \min(R_0^{(\text{range,up})}, \text{median}(\theta_i^{(j=1,...,N_{grid},R_0)})) \text{ ELSE } \theta_i^{(j,R_0)} \text{ END IF}$  982
     $\theta_i^{(j,mdi)} = \text{IF } \theta_i^{(j,mdi)} > mdi^{(\text{range,up})} : \min(mdi^{(\text{range,up})}, \text{median}(\theta_i^{(j=1,...,N_{grid},mdi)})) \text{ ELSE } \theta_i^{(j,mdi)}$  983
END IF 984
     $\hat{\nu}_i^{(j,m)} = \text{IF } \hat{\nu}_i^{(j,m)} < 0 : \max(1, \text{mean}(\hat{\nu}_i^{(j=1,...,N_{grid},m)})) \text{ ELSE } \hat{\nu}_i^{(j,m)}, m \in \{S, I, T, R\} \text{ END IF}$  985
     $\theta_i^{(j,m)} = \text{IF } \theta_i^{(j,m)} < 0 : \max(1, \text{mean}(\theta_i^{(j=1,...,N_{grid},m)})) \text{ ELSE } \theta_i^{(j,m)}, m \in \{R_0, mdi\} \text{ END IF}$  986
END FOR 987
END FOR 988

```

#### 4. Output: a posteriori sample of parameters $\theta_n$ and states $\hat{\nu}_n$ .

## References

1. Lipsitch M, Finelli L, Heffernan RT, Leung GM, Redd; for the 2009 H1N1 Surveillance Group SC. Improving the evidence base for decision making during a pandemic: the example of 2009 influenza A/H1N1. *Biosecurity and bioterrorism: biodefense strategy, practice, and science.* 2011;9(2):89–115. 991  
992  
993  
994
2. Mills CE, Robins JM, Lipsitch M. Transmissibility of 1918 pandemic influenza. *Nature.* 2004;432(7019):904–906. 995  
996
3. Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, et al. Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. *New England Journal of Medicine.* 2009;361(27):2619–2627. 997  
998  
999
4. Cowling BJ, Lau MS, Ho LM, Chuang SK, Tsang T, Liu SH, et al. The effective reproduction number of pandemic influenza: prospective estimation. *Epidemiology.* 2010;21(6):842. 1000  
1001  
1002
5. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology.* 2013;178(9):1505–1512. 1003  
1004  
1005
6. Cauchemez S, Epperson S, Biggerstaff M, Swerdlow D, Finelli L, Ferguson NM. Using routine surveillance data to estimate the epidemic potential of emerging zoonoses: application to the emergence of US swine origin influenza A H3N2v virus. *PLoS Med.* 2013;10(3):e1001399. 1006  
1007  
1008  
1009
7. Cauchemez S, Carrat F, Viboud C, Valleron A, Boelle P. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in medicine.* 2004;23(22):3469–3487. 1010  
1011  
1012
8. Höhle M, Jørgensen E, O'Neill PD. Inference in disease transmission experiments by using stochastic epidemic models. *Journal of the Royal Statistical Society: Series C (Applied Statistics).* 2005;54(2):349–366. 1013  
1014  
1015

9. White LF, Pagano M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Stat Med.* 2008 Jul;27(16):2999–3016. 1016  
1017  
1018
10. O'Neill PD. A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Mathematical biosciences.* 2002;180(1):103–114. 1019  
1020  
1021
11. Obadia T, Haneef R, Boelle PY. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Med Inform Decis Mak.* 2012;12:147. 1022  
1023  
1024
12. Cauchemez S, Boelle PY, Donnelly CA, Ferguson NM, Thomas G, Leung GM, et al. Real-time estimates in early detection of SARS. *Emerging Infect Dis.* 2006 Jan;12(1):110–113. 1025  
1026  
1027
13. Cauchemez S, Boëlle PY, Thomas G, Valleron AJ. Estimating in real time the efficacy of measures to control emerging communicable diseases. *American Journal of Epidemiology.* 2006;164(6):591–597. 1028  
1029  
1030
14. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology.* 2004;160(6):509–516. 1031  
1032  
1033
15. Davoudi B, Miller JC, Meza R, Meyers LA, Earn DJD, Pourbohloul B. Early real-time estimation of the basic reproduction number of emerging infectious diseases. *Physical Review X.* 2012;2. 1034  
1035  
1036
16. Cauchemez S, Ferguson NM. Methods to infer transmission risk factors in complex outbreak data. *J R Soc Interface.* 2012 Mar;9(68):456–469. 1037  
1038
17. Saramaki J, Kaski K. Modelling development of epidemics with dynamic small-world networks. *J Theor Biol.* 2005 Jun;234(3):413–421. 1039  
1040
18. Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press; 1992. 1041  
1042
19. Daley DJ, Gani JM. *Epidemic Modelling: An Introduction.* Cambridge; New York: Cambridge University Press; 1999. 1043  
1044
20. Alkema L, Raftery AE, Clark SJ. Probabilistic projections of HIV prevalence using Bayesian melding. *The Annals of Applied Statistics.* 2007;p. 229–248. 1045  
1046
21. Elderd BD, Dukic VM, Dwyer G. Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. *Proceedings of the National Academy of Sciences.* 2006;103(42):15693–15697. 1047  
1048  
1049
22. Birrell PJ, Ketsetzis G, Gay NJ, Cooper BS, Presanis AM, Harris RJ, et al. Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proceedings of the National Academy of Sciences.* 2011;108(45):18238–18243. 1050  
1051  
1052  
1053
23. Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science.* 2003;300(5627):1961–1966. 1054  
1055  
1056
24. Choi B, Rempala GA. Inference for discretely observed stochastic kinetic networks with applications to epidemic modeling. *Biostatistics.* 2012 Jan;13(1):153–165. 1057  
1058  
1059

25. Ionides EL, Breto C, King AA. Inference for nonlinear dynamical systems. *Proc Natl Acad Sci USA.* 2006 Dec;103(49):18438–18443. 1060  
1061
26. Shaman J, Karspeck A, Yang W, Tamerius J, Lipsitch M. Real-time influenza forecasts during the 2012-2013 season. *Nat Commun.* 2013;4:2837. 1062  
1063
27. Yang W, Karspeck A, Shaman J. Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics. *PLOS Computational Biology.* 2014;10:e1003583. 1064  
1065  
1066
28. Yang W, Cowling BJ, Lau EH, Shaman J. Forecasting Influenza Epidemics in Hong Kong. *PLoS Comput Biol.* 2015 Jul;11(7):e1004383. 1067  
1068
29. Shaman J, Karspeck A. Forecasting seasonal outbreaks of influenza. *Proc Natl Acad Sci USA.* 2012 Dec;109(50):20425–20430. 1069  
1070
30. Ong JB, Chen MI, Cook AR, Lee HC, Lee VJ, Lin RT, et al. Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore. *PLoS ONE.* 2010;5(4):e10036. 1071  
1072  
1073
31. Dukic V, Lopes HF, Polson NG. Tracking Epidemics With Google Flu Trends Data and a State-Space SEIR Model. *Journal of the American Statistical Association.* 2012;107. 1074  
1075  
1076
32. Chretien JP, George D, Shaman J, Chitale RA, McKenzie FE. Influenza forecasting in human populations: a scoping review. *PLoS ONE.* 2014;9(4):e94130. 1077  
1078  
1079
33. Bettencourt LM, Ribeiro RM. Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One.* 2008;3(5):e2185. 1080  
1081
34. Abbey H. An examination of the Reed-Frost theory of epidemics. *Human biology.* 1952;24(3):201–233. 1082  
1083
35. Zimmer C, Sahle S. Deterministic inference for stochastic systems using multiple shooting and a linear noise approximation for the transition probabilities. *IET Systems Biology.* 2015;9:181 – 192. 1084  
1085  
1086
36. Zimmer C. Reconstructing the hidden states in time course data of stochastic models. *Mathematical BioSciences.* 2015;269:117 – 129. 1087  
1088
37. Thomas P, Matuschek H, Grima R. Intrinsic noise analyzer: a software package for the exploration of stochastic biochemical kinetics using the system size expansion. *PloS one.* 2012;7(6):e38518. 1089  
1090  
1091
38. Van Kampen NG. Stochastic processes in physics and chemistry. vol. 1. Elsevier; 1992. 1092  
1093
39. Yaesoubi R, Cohen T. Generalized Markov models of infectious disease spread: A novel framework for developing dynamic health policies. *European Journal of Operational Research.* 2011;215:679–687. 1094  
1095  
1096
40. Gillespie DT. A General Method for Numerically Simulating the Stochastic Time Evolution of coupled Chemical Reactions. *Journal of Computational Physics.* 1976;22 (4):403–434. 1097  
1098  
1099
41. Influenza prediction challenge. Center for disease control and prevention. 2016;<http://www.cdc.gov/flu/news/flu-forecast-website-launched.htm>. 1100  
1101

42. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, et al. COPASI - a COmplex PAthway SIMulator. *Bioinformatics*. 2006;22 (24):3067–3074. 1102  
1103
43. Mathematica, Version 10.4. Wolfram Research, Inc. 2015;Champaign, IL. 1104
44. Pahle J, Challenger JD, Mendes P, McKane AJ. Biochemical fluctuations, 1105  
optimisation and the linear noise approximation. *BMC Systems Biology*. 2012;6. 1106
45. Challenger JD, McKane AJ, Pahle J. Multi-compartment linear noise 1107  
approximation. *Journal of Statistical Mechanics: Theory and Experiment*. 1108  
2012;P11010. 1109
46. Straube R, von Kamp A. LiNA – A Graphical Matlab Tool for Analyzing 1110  
Intrinsic Noise in Biochemical Reaction Networks. online tutorial. 2013;[http://www2.mpi-magdeburg.mpg.de/projects/LiNA/Tutorial\\_LiNA\\_v1.pdf](http://www2.mpi-magdeburg.mpg.de/projects/LiNA/Tutorial_LiNA_v1.pdf). 1111  
1112
47. van Kampen NG. Stochastic processes in physics and chemistry. Elsevier; 2007. 1113
48. Grima R. An effective rate equation approach to reaction kinetics in small 1114  
volumes: Theory and application to biochemical reactions in nonequilibrium 1115  
steady-state conditions. *The Journal of Chemical Physics*. 2010;133:035101. 1116
49. Thomas P, Matuschek H, Grima R. Intrinsic Noise Analyzer: A Software Package 1117  
for the Exploration of Stochastic Biochemical Kinetics Using the System Size 1118  
Expansion. *Plos ONE*. 2012;7:e38518. 1119
50. Komorowski M, Finkenstädt B, Harper CV, Rand DA. Bayesian inference of 1120  
biochemical kinetic parameters using the linear noise approximation. *BMC* 1121  
*Bioinformatics*. 2009;10:343. 1122