Bayesian Sequential Inference with hidden semi-Markov processes

Joint work with Patrick Aschermayr
and Nikos Demiris (MRC, AUEB)

Research Showcase, Department of Statistics, LSE

June 21, 2024

Outline

Outline

Introduction

- Hidden Markov and semi-Markov models find several applications ranging from speech recognition to biostatistics. (see e.g. Muprhy 2013)
- The name can be misleading, as the latent process is not only assumed to be Markov but also to have discrete states.
- In some cases, applications lead to offline (batch) data but we are also interested in sequential data (Chiappa 2014).

Hidden Markov Models (HMMs)

A hidden Markov model (HMM) can be formulated as follows:

$$\begin{aligned} e_t &\sim g_\theta(e_t|s_t), \quad t = 1, 2, \ldots, T \\ s_t|s_{t-1} &\sim f_\theta(s_t|s_{t-1}) \end{aligned}$$

where e_t are the data, $g_\theta(\cdot)$ is known and $f_\theta(s_t|s_{t-1})$ is determined by the transition probabilities P_{ij}

$$P_{ij} = P(s_t = i|s_{t-1} = j), \quad \forall i, j$$

May be also thought of as dynamic change-point model, model based clustering or latent class model for dependent data.

Hidden semi Markov Models (HSMMs)

Let d_t denote the remaining time in the current state of s_t.

An alternative formulation is now given as

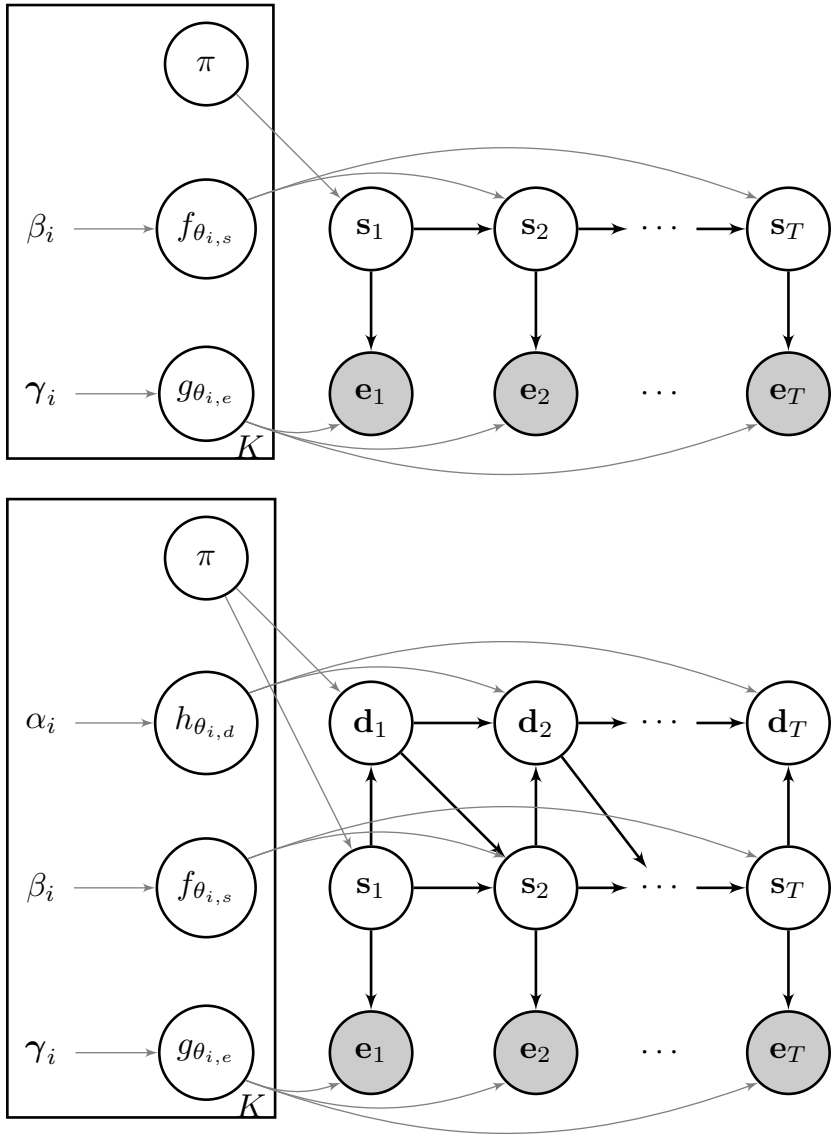
$$s_t|s_{t-1}, d_{t-1} \sim \begin{cases} \delta(s_{t-1}), & \text{if } d_{t-1} > 0 \\ f_\theta(s_t|s_{t-1}, d_{t-1}), & \text{if } d_{t-1} = 0 \end{cases}$$

$$d_t|s_{t-1}, d_{t-1} \sim \begin{cases} \delta(d_{t-1} - 1), & \text{if } d_{t-1} > 0 \\ h_\theta(d_t|s_t, d_{t-1}), & \text{if } d_{t-1} = 0 \end{cases},$$

where $h_\theta(\cdot)$ is the Geometric distribution (often not a good fit).

Hidden semi Markov models (HSMMs) generalise HMMs by allowing for different distributions than the Geometric, e.g. Negative Binomial, Poisson etc.

HMMs vs HSMMs



Outline

Forward-backward algorithm

Denote $e = (e_1, \ldots, e_T)$ and similarly s, d. Then define the augmented likelihood $f(e, s, d|\theta)$ and the integrated likelihood $f(e|\theta)$.

For HMMs it is possible to evaluate the $f(e|\theta)$ directly using the forward-backward algorithm to provide an EM-type algorithm.

An approximate version of the forward algorithm exists for HSMMs but it can get computationally expensive; in some cases it can get to $O(TKd_{max}^2)$, where d_{max} is a maximum duration we can introduce.

Data augmented scheme

In this work, we aim to provide a computational scheme working with the augmented likelihood $f(e, s, d|\theta)$.

Looking for a Markov chain Monte Carlo (MCMC) scheme that samples from the posterior of s, d and θ. Application in a sequential setting is also desired.

Standard MCMC algorithms are challenging. The parameter space is discrete, hence no derivatives, and no natural blocking schemes are available.

Particle filter

Let $x_t = (s_t, d_t)$ and assume that x_0 is known. The particle filter proceeds as follows at each time $t = 1, \ldots, T$, for a fixed θ:

1. Draw n independent x_t samples $\{x_t^{(i)}\}_{i=1}^n$ with equal weights from $\pi(x_t|x_{t-1})$, given $\{x_{t-1}^{(i)}\}_{i=1}^n$. prediction
2. Compute their weights $\{w_t^{(i)}\}_{i=1}^n$. This allows to calculate any expectation wrt $\pi(x_t|e_{1:t})$. filtering
3. To avoid degeneracy, sample with replacement from $\{x_t^{(i)}, w_t^{(i)}\}_{i=1}^n$ to obtain an unweighted set $\{x_t^{(i)}, 1^{(i)}\}_{i=1}^n$

We can use the particle filter to construct the following algorithms

- in itself gives an online algorithm (assuming known θ),
- within particle MCMC allows for offline inference on (x_t, θ),
- within a SMC^2 for sequential inference on (x_t, θ).

Particle filter based MCMC algorithm

Developed MCMC algorithm:

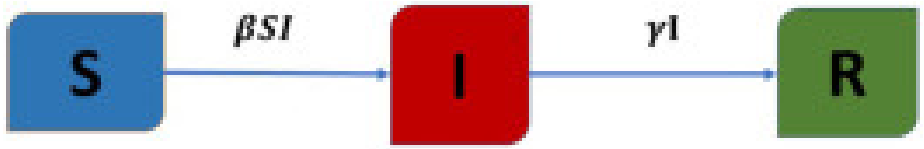
- Sample from the conditional posterior of $\{x_t\}_{t=1}^{T}|\theta$ using a particle filter.
- Sample from the conditional posterior of $\theta|\{x_t\}_{t=1}^{T}$ using Hamiltonian MCMC.

Benefits:

- Allows to sample from the marginal posterior of (s_t, d_t)
- Easy to extend to sequential versions

Outline

SIR type models for epidemics



$$\begin{aligned}
\frac{dS_t}{dt} &= -\beta S_t I_t \\
\frac{dI_t}{dt} &= \beta S_t I_t - \gamma I_t \\
\frac{dR_t}{dt} &= \gamma I_t
\end{aligned}$$

R number:

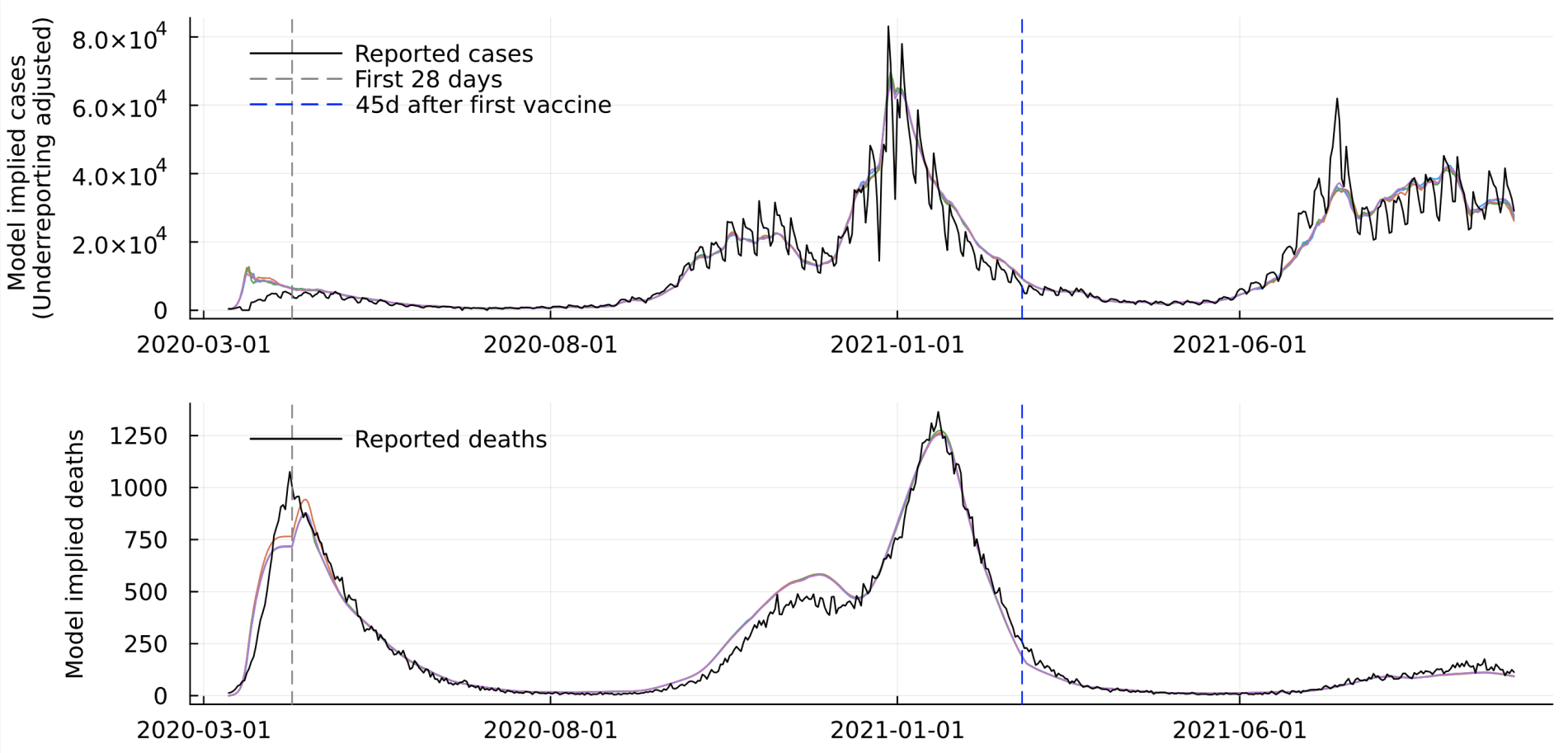
$$R_0 = \frac{\beta S_0}{\gamma}$$

Modelling transmission rate β

- Time varying and stochastic in nature, depends on the virus as well as social and environmental factors.
- Several approaches based on Brownian motion or Gaussian process.
- Several approaches based on change-points.
- We were looking for a middle ground.

Data

Reported cases and deaths in the UK, publicly available from gov.uk (600 points)



Model - ODE transmission

A more elaborate transmission model: with E and I states split into two parts (for better approximation) and a vaccination term.

$$
\begin{aligned}
\frac{\mathrm{d}S_t}{\mathrm{d}t} &= -\beta_t S_t \frac{(I_{1,t}+I_{2,t})}{N} - \rho\nu_{t-U}, \\
\frac{\mathrm{d}E_{1,t}}{\mathrm{d}t} &= \beta_t S_t \frac{(I_{1,t}+I_{2,t})}{N} - \epsilon E_{1,t}, \\
\frac{\mathrm{d}E_{2,t}}{\mathrm{d}t} &= \epsilon E_{1,t} - \epsilon E_{2,t}, \\
\frac{\mathrm{d}I_{1,t}}{\mathrm{d}t} &= \epsilon E_{2,t} - \gamma I_{1,t}, \\
\frac{\mathrm{d}I_{2,t}}{\mathrm{d}t} &= \gamma I_{1,t} - \gamma I_{2,t}, \\
\frac{\mathrm{d}R_t}{\mathrm{d}t} &= \gamma I_{2,t} + \rho\nu_{t-U},
\end{aligned}
$$

Model - ODE quantities

Except for β_t, all the unknown quantities in the ODE of the transmission model, e.g. ϵ, γ, ρ and the initial states were given informative priors based on other studies

β_t was modelled with several HMM and HSMM variants, i.e. Negative Binomial or Poisson durations, with different numbers of states.

Model - reported and model implied cases

The model for the reported cases, c_t^r is defined as

$$c_t^r \sim \text{Negative Binomial}\left(c_t, c_t + \frac{c_t^2}{\phi_c}\right),$$

where c_t are the model implied cases coming from the ODE.

The reported cases were adjusted for under-reporting based on a previous study

The model implied cases c_t are obtained for solving the ODE in the time interval $(t-1, t]$, hence the model has a state space representation.

Model - reported Covid-19 deaths

The model implied deaths d_t are considered a function of the model implied cases c_t over the last 28 days, see (Flaxman et al 2020), in line with the UK definition, as well as available estimates of the infection to fatality ratio (*ifr*)

$$d_t = ifr_t \times \sum_{\tau=max(1,t-28)}^{t-1} c_\tau f_{t-\tau},$$

The reported deaths d_t^r were then modelled as

$$d_t^r \sim \text{Negative Binomial}_{Alternative}\left(d_t^i, d_t^i + \frac{{d^i}_t^2}{\phi_d}\right).$$

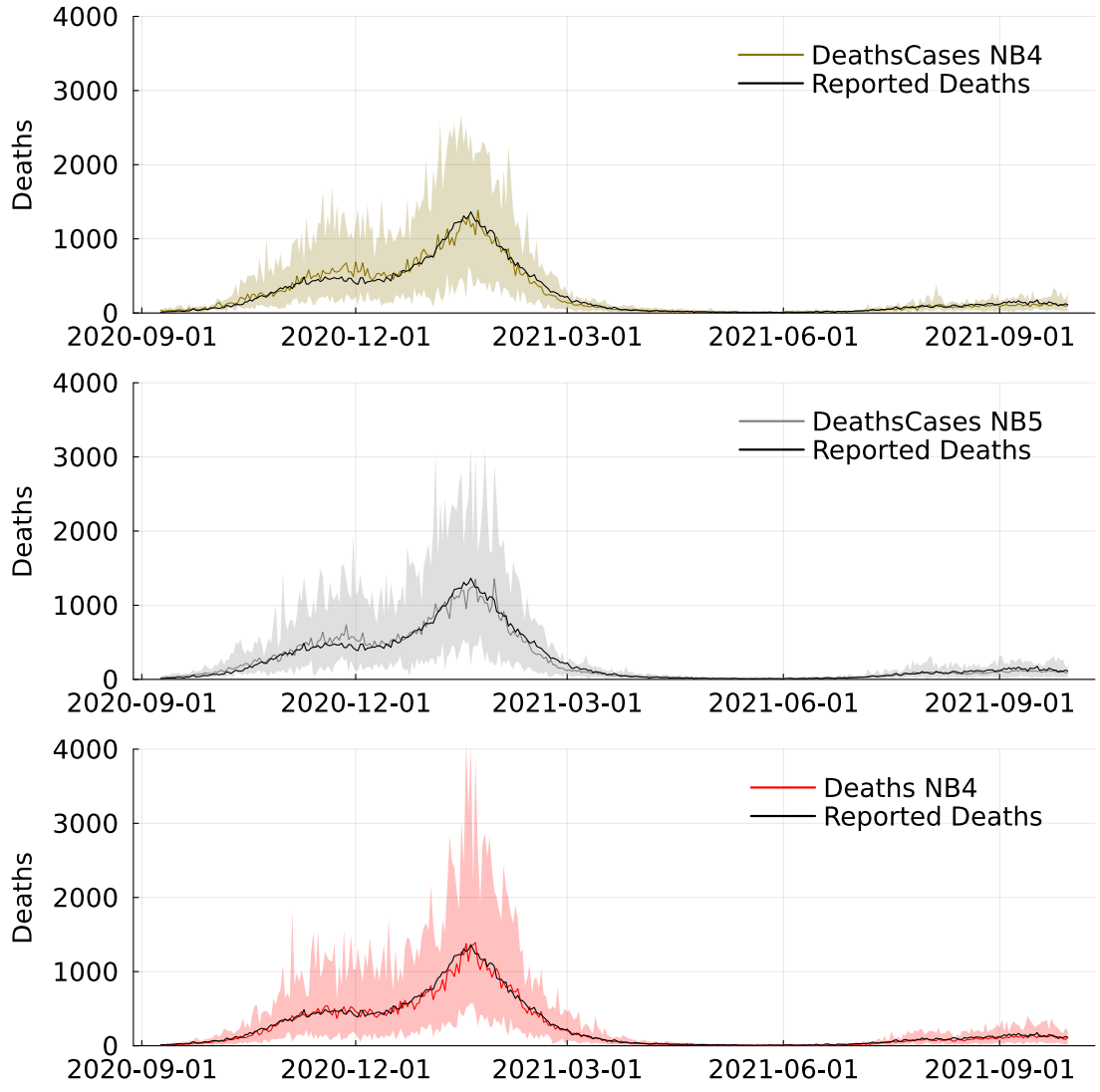
Reported cases, deaths or both?

- One of the substantive questions we wanted to answer is whether one should use the reported cases and/or reported deaths.
- Reported cases are known to be problematic, including under-reporting.
- Reported deaths appear to be more reliable but still have issues (definition, *ifr* estimates over different times etc).
- We considered models with reported deaths only as well as models with reported deaths and cases.

Model choice based on prediction

- We implemented a SMC^2 version of the algorithm to obtain efficiently obtain predictive distributions as data accumulate.
- Focus on predicting deaths since the data are more reliable. The predictive distribution from different models were evaluated based on the log-score.
- As mentioned earlier models with reported deaths only as well reported deaths and cases were considered. Also models with different duration distributions and number of states

Model predictions

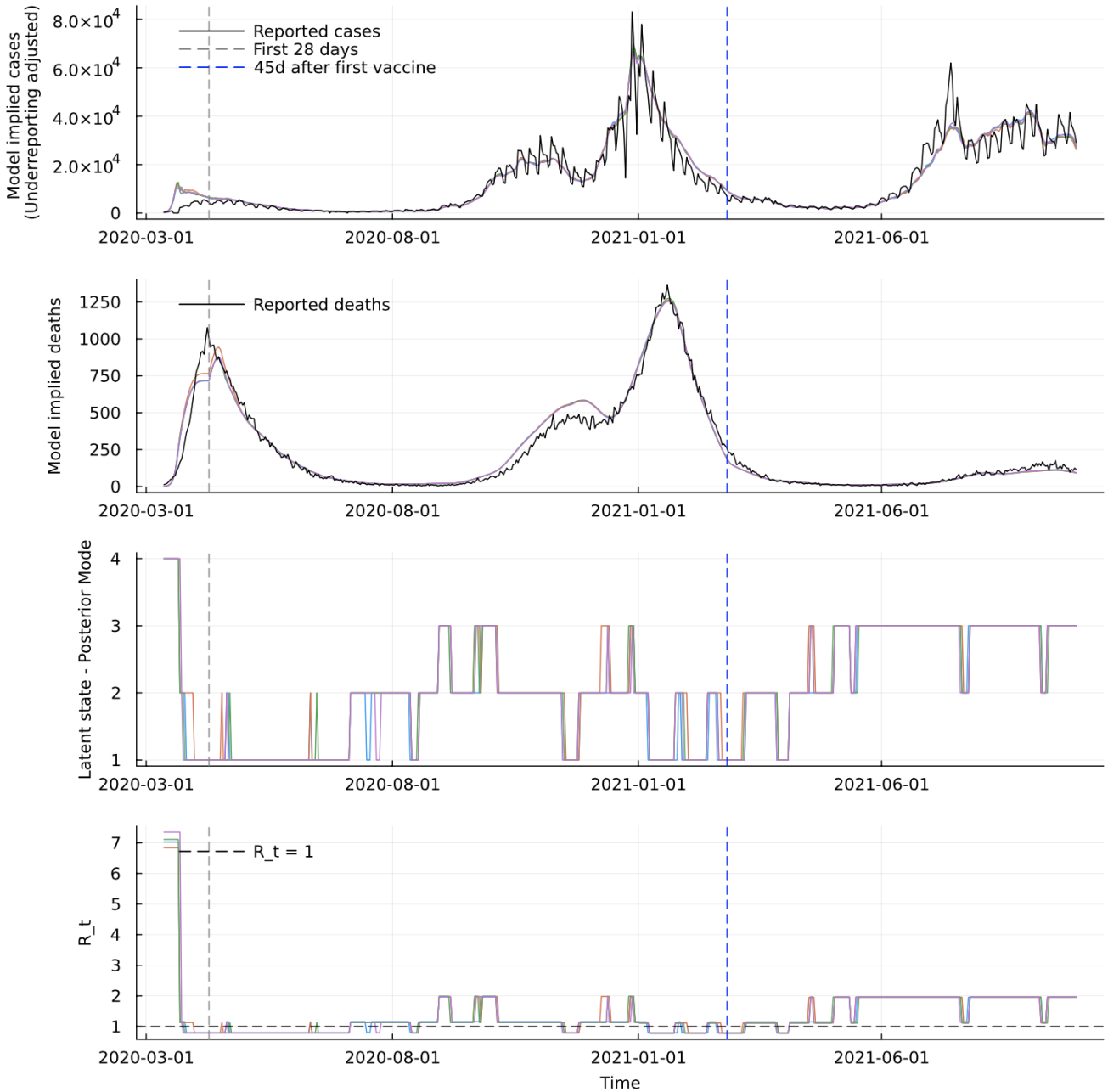


Model choice results

Daily sequential model choice	
Model	**Daily cumulative log PL**
Deaths - 4 states	-26394
Deaths and Cases - 4 states	-16523
Deaths and Cases - 5 states	-16601

Weekly sequential model choice	
Model	**Weekly cumulative log PL**
Deaths - 4 states	-1985
Deaths and Cases - 4 states	-1845
Deaths and Cases - 5 states	-1843

Model Output



NB4 model estimates

θ	Mean	MCSE	SD	Rhat	Q2.5	Q25.0	Q50.0	Q75.0	Q97.5
$\log \beta_1$	-1.72	0.0	0.08	1.01	-1.89	-1.78	-1.72	-1.67	-1.57
$\log \beta_2$	-1.36	0.01	0.08	1.03	-1.5	-1.41	-1.36	-1.31	-1.2
$\log \beta_3$	-0.81	0.0	0.09	1.0	-0.99	-0.87	-0.81	-0.76	-0.63
$\log \beta_4$	0.45	0.01	0.22	1.01	0.02	0.3	0.44	0.59	0.9
γ_1	0.45	0.0	0.04	1.0	0.37	0.42	0.45	0.48	0.54
γ_2	0.46	0.0	0.04	1.0	0.38	0.43	0.45	0.48	0.53
ϵ	0.94	0.0	0.1	1.0	0.76	0.87	0.94	1.0	1.13
p_1	0.87	0.0	0.09	1.01	0.66	0.83	0.89	0.94	0.98
p_2	0.5	0.01	0.14	1.01	0.23	0.38	0.5	0.6	0.77
p_3	0.17	0.01	0.1	1.01	0.03	0.09	0.15	0.23	0.4
$p_{thirdstate,1}$	0.35	0.01	0.15	1.0	0.09	0.23	0.34	0.45	0.67
$p_{thirdstate,2}$	0.35	0.01	0.15	1.0	0.08	0.24	0.34	0.46	0.67
r_1	36.12	0.18	6.53	1.0	24.69	31.63	35.71	39.77	50.58
r_2	24.19	0.15	5.29	1.0	14.72	20.7	23.91	27.49	35.39
r_3	14.19	0.17	4.32	1.0	7.22	11.0	13.72	16.96	23.56
r_4	28.13	0.12	5.33	1.0	19.08	24.22	27.75	31.69	39.0
ψ_1	0.76	0.01	0.07	1.04	0.62	0.72	0.77	0.81	0.88
ψ_2	0.75	0.01	0.07	1.03	0.6	0.71	0.75	0.8	0.87
ψ_3	0.55	0.01	0.09	1.01	0.36	0.48	0.55	0.61	0.73
ψ_4	0.5	0.01	0.15	1.0	0.21	0.4	0.5	0.61	0.81
ϕ_{cases}	4.91	0.0	0.11	1.0	4.68	4.83	4.91	4.99	5.13
ϕ_{deaths}	5.25	0.0	0.12	1.0	5.02	5.17	5.26	5.33	5.49

Table 3. Posterior output statistics of four PMCMC chains on a HSMM-EM with SEIR style ODE, 4 states and Negative Binomial duration distribution for applications in Section 5. 1200 iterations have been used with burnin set to 700, resulting in 2000 total samples. Real data can be seen in Figure 6, and is described in more detail in Section 5.1. Initial parameter have been sampled from the prior distributions.

Discussion - Future directions

- Flexible modelling framework for SIR-type HSMMs.
- Feasible computational toolkit on a challenging MCMC problem.
- Model extensions, e.g. covariate dependent durations.
- Computational issues, e.g. multimodality and label switching, especially in over-parametrised models.