Journal of the Royal Statistical Society Series A: Statistics in Society, 2023, **186**, 601–615
https://doi.org/10.1093/jrsssa/qnad030
Advance access publication 28 February 2023
Discussion Paper

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Semi-mechanistic Bayesian modelling of COVID-19 with renewal processes

Samir Bhatt[1], Neil Ferguson[1], Seth Flaxman[2], Axel Gandy[2], Swapnil Mishra[1] and James A. Scott[2]

[1]MRC Centre for Global Infectious Disease Analysis, Imperial College, London, UK
[2]Department of Mathematics, Imperial College, London, UK

Address for correspondence: Axel Gandy, Department of Mathematics, Imperial College London, London SW7 2AZ, UK.
Email: a.gandy@imperial.ac.uk

Read before The Royal Statistical Society in London at the second meeting on 'Statistical aspects of the Covid-19 Pandemic' organized by the Discussion Meetings Committee on Thursday, 16 June 2022, Dr Shirley Coleman in the Chair.

Abstract

We propose a general Bayesian approach to modelling epidemics such as COVID-19. The approach grew out of specific analyses conducted during the pandemic, in particular, an analysis concerning the effects of non-pharmaceutical interventions (NPIs) in reducing COVID-19 transmission in 11 European countries. The model parameterises the time-varying reproduction number R_t through a multilevel regression framework in which covariates can be governmental interventions, changes in mobility patterns, or other behavioural measures. Bayesian multilevel modelling allows a joint fit across regions, with partial pooling to share strength. This innovation was critical to our timely estimates of the impact of lockdown and other NPIs in the European epidemics: estimates from countries at later stages in their epidemics informed those of countries at earlier stages. Originally released as Imperial College Reports, the validity of this approach was borne out by the subsequent course of the epidemic. Our framework provides a fully generative model for latent infections and derived observations, including deaths, cases, hospitalizations, ICU admissions, and seroprevalence surveys. In this article, we additionally explore the confounded nature of NPIs and mobility. Versions of our model were used by New York State, Tennessee, and Scotland to estimate the current epidemic situation and make policy decisions.

Keywords: infectious disease modelling, renewal equations, COVID-19, mechanistic modelling

1 Introduction

This article presents a general framework for semimechanistic Bayesian modelling of infectious diseases using renewal processes. The term semimechanistic relates to statistical estimation within some constrained mechanism. Variants of this general model have been used in specific analyses of COVID-19 (Flaxman, Mishra, Gandy, Unwin, Mellan, et al., 2020; Mellan et al., 2020; Mishra et al., 2020; NYS Press Office, 2020; Olney et al., 2020; The Scottish Government, 2020; H. Unwin et al., 2020; Vollmer et al., 2020) and continue to be used in ongoing work to make policy decisions. The present article motivates and discusses the key statistical and epidemiological features of this framework, starting from a counting process setup. Various extensions of the basic model are considered, including a latent infection process. We discuss the limitations and applications of the modelling framework to stimulate further research.

The model uses a flexible regression-based framework for parameterising transmission and ascertainment rates. This allows the fitting of multilevel models (Gelman & Hill, 2006; Hox et al., 2010; Kreft & de Leeuw, 2011) for several regions simultaneously. Such partial pooling of parameters has specific advantages in the context of infectious diseases. Suppose we wish to estimate the

Received: December 13, 2020. **Revised:** March 24, 2022. **Accepted:** March 31, 2022

effect of non-pharmaceutical interventions (NPIs) (Cowling et al., 2020; Flaxman, Mishra, Gandy, Unwin, Mellan, et al., 2020; Islam et al., 2020) or mobility (Badr et al., 2020; Miller et al., 2020) on transmission rates. Estimating separate models for each region could lead to a set of noisy estimates for at least two reasons. There is typically little high-quality data at the early stages of an epidemic, and such data are generally correlated, reducing the information content that can be used to infer such an effect. In addition, NPIs often occur in quick succession and their effects are confounded (Hung, 2003; WHO, 2003). This is exacerbated by the random times between infections (the generation distribution) and between infections and observations, which smooths the observed data, making it more difficult to attribute changes in transmission rates to a particular NPI. Alternatively, one could fit a single model, pooling the effect across all regions. This ignores region-level variation and can lead to poor predictive performance, in particular underestimating variance for previously unmodeled regions. One could augment such a model with group-level indicators, but this results in a large number of parameters, which are difficult to estimate and leads to overfitting with classical estimation techniques. The Bayesian multilevel regression approach of partial pooling provides a natural solution to these problems.

Sometimes the inferential goal is not to assess the effect of a covariate on outcomes, but rather to infer *latent* transmission rates and their effect on outcomes. Previous studies have focused on estimating reproduction numbers from case data (Bettencourt & Ribeiro, 2008; Cori et al., 2013; Ferguson et al., 2001; Fraser et al., 2009; Kelly et al., 2010; Riley et al., 2003), sometimes directly substituting observed case counts for the unknown number of infected individuals (Wallinga & Teunis, 2004). However, the emergence of SARS-CoV-2 has highlighted the shortcomings of methods that rely on just case data. Limited testing capacity at the early stages of the pandemic led to only a small proportion of infections being detected and reported (Li et al., 2020). Those tested were typically more likely to have been hospitalised or were at higher risk of infection or death. In particular, this proportion, referred to as the infection ascertainment rate (IAR), is country-specific and likely to have changed over time due to changes in testing policies and capacity. If unaccounted for, it will lead to biases in the inferred transmission rates.

The problem of varying case ascertainment highlights the need for more flexible observation models to rigorously incorporate various types of data, from hospitalizations to seroprevalence surveys. Daily death data have been used in (Flaxman, Mishra, Gandy, Unwin, Mellan, et al., 2020) to recover reproduction numbers in the early stages of the SARS-CoV-2 pandemic and has been seen as more reliable than case data. However, there have been clear variations in case and mortality definitions as well as reporting across time and countries. It is, therefore, important to appropriately model noise within the observational models. Our framework allows for multiple types of data including deaths, cases, hospitalizations, ICU admissions, and the results of seroprevalence surveys. This improves robustness of inferred parameters to biases in any one type of data.

Our model uses discrete renewal processes to propagate infections within modelled populations. These have been used in a number of previous studies (Cauchemez et al., 2008; Cori et al., 2013; Fraser, 2007; Nouvellet et al., 2018) and are linked to other popular approaches to infectious disease modelling. Champredon et al. (2018) show that the renewal equation leads to identical dynamics as Erlang-distributed susceptible–exposed–infected–recovered (SEIR) compartmental models, when a particular form is used for the generation distribution. A special case of this is the standard susceptible–infected–recovered (SIR) model (Kermack & McKendrick, 1927). The approach is also connected to counting processes such as the general branching processes (Bellman & Harris, 1948; Crump & Mode, 1969; Jagers, 1969; Kimmel, 1983; Pakkanen et al., 2022). Self- exciting Hawkes processes are also related to renewal processes, with the expectation of the Hawkes intensity function resulting in a renewal equation (Rizoiu et al., 2017).

We describe the general model in detail and start by considering the simplest version in Section 2. The motivation for the model lies in continuous-time counting processes, and this connection is discussed in Section 3. Sections 4 and 5 present the infection and observation processes in more detail and consider important extensions of the basic model. Section 6 considers how to use the framework for multilevel modelling. Section 7 compares our approach to standard time series models and outlines the key challenges involved in modelling with our framework. Section 8 considers the specific aspect of confounding and causality when estimating the effects of variables on transmission rates. Section 8.4 considers a simple simulation analysis and a discussion of when our approach is expected to fail, and Section 9 has a brief discussion.

2 Model overview

We now formulate a basic version of the model for one homogeneous population. The same model can be used for multiple regions or groups jointly. In what follows, we will consider discrete time, $t \in \mathbb{Z}$, but continuous analogues are similar (Pakkanen et al., 2022). Let $R_t > 0$ be the time-varying reproduction number at time $t > 0$, determining the average number of secondary infections caused by an infected person. The number of seeded infections $i_\nu, \ldots, i_{-1}, i_0$ for some integer-valued timepoint $\nu \leq 0$ are given a prior distribution. For $t > 0$, we let new infections i_t be defined by

$$i_t = R_t \sum_{s<t} i_s g_{t-s}, \tag{1}$$

where the generation time, the lag between infections, is given through a probability mass function g, i.e., $g_t \geq 0$ and $\sum_{t=1}^{\infty} g_t = 1$.

We never observe the exact time at which a person becomes infected, rather we observe events reported by health care systems, introducing bias due to under/over ascertainment and reporting lags. Recorded observations occur at certain times $t > 0$. In general, there may be multiple types: cases, hospitalisations, and death counts, for example. Each such type is driven by its own time-varying ascertainment rate $\alpha_t > 0$. The predictions of the recorded observation at time t are linked to past infections by

$$\hat{y}_t = \alpha_t \sum_{s \leq t} i_s \pi_{t-s}, \tag{2}$$

where π is a distribution for the lag between an infection and when it gives rise to a recorded observation. The sampling distribution of the observations with these means is typically non-negative and discrete, and may depend on auxiliary parameters. When multiple types are observed, we can superscript the quantities as $\hat{y}_t^{(l)}$, $\alpha_t^{(l)}$ and $\pi^{(l)}$ and assign independent sampling distributions for each type. We can connect $\hat{y}$ to a likelihood for true observations y with, e.g., Poisson or negative binomial likelihoods.

Transmission rates R_t and ascertainment rates α_t can be modelled flexibly using Bayesian regression models. Multilevel modelling allows us to share parameters between regions, borrowing strength from regions with advanced epidemics to inform estimates in regions with earlier epidemics. One can, for example, model transmission rates as depending on a binary covariate for an NPI, say full lockdown. The coefficient for this can be *partially pooled* between these groups. The effect is to share information between groups, while still permitting between-group variation.

3 Motivation from continuous time

Our model can be motivated from a continuous time perspective as follows. Infections give rise to additional infections in the future, referred to as offspring. Letting $N^I(t)$ denote the number of infections occurring up to time t, defined by its intensity

$$\lambda(t) = R(t) \textstyle\int_{s<t} g(t-s) N^I(Ds), \quad t > 0, \tag{3}$$

where g is the density of a probability distribution on $\mathbb{R}^+$ defining the time between infections, and where $\{R(t) : t > 0\}$ is a non-negative stochastic process. The process can be initialised by assuming values for $N^I(t)$ for t in the seeding period $[\nu, 0]$.

Equation (3) is similar to the Hawkes intensity, however, the *memory kernel* g is scaled by a time-specific factor $R(t)$. The integrand g allows the intensity to increase due to previous infection events, while $R(t)$ tempers the intensity for other time-specific considerations. Under this assumption, since g integrates to unity, the expected number of offspring is simply $R(t)$, and so this is the *instantaneous reproduction number* or alternatively the *branching factor* of the Hawkes process. The generation time, defined as the time from an infection to a secondary infection, is distributed according to g and so g is the *generation distribution*.

Recorded observations are caused by infections that occurred in the past—that is, a given infection may lead to observation events (cases or deaths) in the future. Letting $N^Y(t)$ be the count of some observation type over time defined by the intensity

$$\lambda_y(t) = \alpha(t) \int_{s<t} \pi(t-s) N^I(\mathrm{d}s), \tag{4}$$

for $t > 0$, where $\pi : \mathbb{R}^+ \to \mathbb{R}^+$ is a function and $\{\alpha(t) : t \geq 0\}$ is a non-negative stochastic process. This is similar to Equation (3), however, the intensity increases due to past infections, rather than past observations.

Consider the special case where π is a probability density and where $\alpha(t') = \alpha(t)$ for all t'. The average number of observation events attributable to a single infection is then $\alpha(t)$, and so this is an *instantaneous ascertainment rate*. π is then interpreted as the distribution for the time from an infection to an observation, and, therefore, we call it the *infection-to-observation* distribution.

4 Infection process

Starting from the continuous model, we now describe a discrete model, which results in the formulation of Section 2. This discrete model is more amenable to inference. Let I_t be the number of new infections at time t; this is the equivalent of $N^I_t - N^I_{t-1}$ in the continuous model. As a basic modelling block, we use the following discrete version of (3):

$$\mathrm{E}[I_t \mid R_{1:t}, I_{v:t-1}] = R_t L_t, \tag{5}$$

where $L_t := \sum_{s<t} I_s g_{t-s}$ is the *case load* or *total infectiousness* by time $t > 0$. Moreover, letting $i_t := \mathrm{E}[I_t \mid R_{1:t}, I_{v:0}]$ and taking the conditional expectation given reproduction numbers $R_{1:t}$ and seeded infections $I_{v:0}$ on both sides of (5) gives

$$i_t = R_t \mathrm{E}[L_t \mid R_{1:t}, I_{v:0}] = R_t \sum_{s<t} \mathrm{E}[I_s \mid R_{1:s}, I_{v:0}] g_{t-s} = R_t \sum_{s<t} i_s g_{t-s},$$

which is Equation (1). This is a discrete renewal equation, which can alternatively be interpreted as an AR(t)-process with known coefficients g_k. From this point of view, the basic model in Section 2 uses i_t as synonymous with actual infections. Since infections are simply a deterministic function of other parameters, there is no need to treat them as unknown latent parameters to sample. This can lead to lower sampling times and faster convergence when performing Bayesian inference.

4.1 Modelling latent infections

The model of Section 2 can be extended by replacing each i_t with the actual infections from the counting process I_t, and then assigning a prior to I_t. Although sampling can be slower, this has certain advantages. When past infection counts are low, significant variance in the offspring distribution can imply that the number of new infections I_t has high variance. This is not explicitly accounted for in the basic model. In addition, this approach cleanly separates infections and observations; the latter being modelled *conditional* on actual infections. The sampling distribution can then focus on idiosyncrasies relating to the observation process.

We assign a prior to I_t conditional on previous infections and current transmission R_t. The expected value for this is given by Equation (5). Appendix A.1 shows that assuming the variance of the prior to be a constant proportion d of this mean is equivalent to letting d be the *coefficient of dispersion* for the offspring distribution. $d > 1$ implies overdispersion, and can be used to account for superspreading events, which has been shown to be an important aspect for modelling COVID-19 (Lloyd-Smith et al., 2005). The parameter d can be assigned a prior.

Any two-parameter family can be used to match these first two moments. Letting this be continuous rather than discrete allows inference to proceed using Hamiltonian Monte Carlo, whereby new values for I_t are proposed simultaneously with all other parameters. Possible candidates include log-normal, Gamma, and the Weibull distributions. If an explicit distribution for the

offspring distribution is desired, one can show that assuming a Gamma distribution with rate λ for this results in a Gamma distribution for I_t with rate λ. The coefficient of dispersion is then simply $d = \lambda^{-1}$.

4.2 Population adjustments

If R_t remains above unity over time, infections grow exponentially without limit (in branching processes literature this is referred to as supercritical). In practice, infections should be bounded from above by S_0, the initial susceptible population. All else being equal, transmission rates are expected to fall as the susceptible population is depleted.

Consider first the model using I_t, which was described in Section 4.1. Equation (5) can be replaced with

$$\mathrm{E}[I_t \mid R_{1:t}, I_{v:t-1}] = (S_0 - I_{t-1})\left(1 - \exp\left(-\frac{R_{u,t}L_t}{S_0}\right)\right), \tag{6}$$

where $R_{u,t}$ is an *unadjusted* reproduction number, which does not account for the susceptible population. This satisfies intuitive properties. As the *unadjusted* expected infections $R_{u,t}L_t$ approaches infinity, the *adjusted* expected value approaches the remaining susceptible population. The motivation for and derivation of Equation (6) is provided in Appendix A.2. In short, this is the solution to a continuous time model whose intensity is a simplification of Equation (3). We must also ensure that the distribution of I_t cannot put positive mass above $S_0 - I_{t-1}$. A simple solution is to use truncated distributions. Of course, this adjusts the mean value from Equation (6), however, this is unlikely to be significant unless the susceptible population is close to zero.

In the basic model, one can apply the adjustment to i_t by replacing L_t in Equation (6) with

$$\mathrm{E}(L_t \mid R_{1:t}, I_{v:0}) = \sum_{s<t} i_s g_{t-s}. \tag{7}$$

5 Observations

Observations are modelled in discrete time, analogous to how we treated infections in Section 4. Letting $\pi : \mathbb{N} \to \mathbb{R}^+$ and $Y_t := N_t^Y - N_{t-1}^Y$, the discrete analogue to Equation (4) is

$$\mathrm{E}[Y_t \mid \alpha_t, I_{v:t}] = \alpha_t \sum_{s \le t} I_s \pi_{t-s}. \tag{8}$$

Taking the expected value of the above-given seeded infections, transmission rates and the current ascertainment rate gives

$$\mathrm{E}[Y_t \mid \alpha_t, R_{1:t}, I_{v:0}] = \alpha_t \sum_{s \le t} i_s \pi_{t-s}, \tag{9}$$

which is recognisable as Equation (2). Thus we have two possible expressions for the mean of Y_t, one given actual infections, and the other given expected infections i_t. The basic model of Section 2 uses the latter, while the extension in Section 4.1 uses the former.

We assume that $Y_t \sim \mathcal{F}(y_t, \phi)$, where $\mathcal{F}$ is a non-negative discrete family parameterised by its mean y_t and potentially an auxiliary parameter ϕ. This could be a Poisson distribution, where there is no auxiliary parameter. Using a quasi-Poisson or negative binomial instead allows for overdispersion. This can be useful to capture, for example, day-to-day variation in ascertainment rates when infection counts are low. The mean y_t can be taken to be either (8) or (9), the latter being used in the basic version of the model. Hidden in this formulation is the assumption that the Y_t's are conditionally independent given y_t. Using multiple observation series $Y_t^{(l)}$ can help to improve the model inferences and identifiability of certain parameters. We simply assume that each such series is conditionally independent given the underlying infection process.

6 Multilevel models

Transmission rates can be modelled quite generally within the framework. If the aim is simply to estimate transmission in a single region over time, one approach could be to let $R_t = \psi^{-1}(\gamma_t)$, where ψ is a link function and γ_t is some autocorrelation process, for example, a random walk. If the goal is to estimate the effect of NPIs in M regions on transmission, we can let $R_t^{(m)}$ denote the time-varying reproduction number in region m at time t, specifying

$$R_t^{(m)} = \psi^{-1}\left(\beta_0^{(m)} + \sum_{k=1}^{p} x_t^{(m)}\beta_k^{(m)}\right), \tag{10}$$

where $x_t^{(m)}$ are binary encodings of NPIs, and $\beta_0^{(m)}$ and $\beta_k^{(m)}$ are region-specific intercepts and effects respectively. The intercepts are used to allow regions to have their own baseline transmission rates (and can be interpreted as R_0). Collecting these group-specific parameters into $\beta^{(m)}$, we can partially pool them by letting $\beta^{(m)} \sim \mathcal{N}(0, \Sigma)$, for each group m, and then assigning a prior to the covariance matrix Σ. This could be an inverse-Wishart prior, or alternatively, Σ can be decomposed into variances and a correlation matrix, which are each given separate priors (Tokuda et al., 2011).

One possible option for ψ is the log-link. This provides easily interpretable effect sizes; a one-unit change in a covariate multiplies transmission by a constant factor. However, this can lead to prior mass on unreasonably high transmission rates. With this in mind, an alternative is to use a generalisation of the logit link for which

$$\psi^{-1}(x) = \frac{K}{1 + e^{-x}}, \tag{11}$$

and where K is the maximum possible value for transmission rates. This serves a similar purpose to the carrying capacity in a logistic growth model.

The ascertainment rate α_t can also be modelled with similar considerations to the above. This flexibility is useful, particularly because these quantities are likely to change as an epidemic progresses. This has been clearly seen during the COVID-19 epidemic, where the infection ascertainment rate changed over time due to increased testing capacity and improved contact tracing systems. Multilevel modelling approaches are equally applicable to the specification of α_t.

7 Forecasting, epidemiological constants, and seeding

A key benefit of using a semimechanistic approach is that forecasts are constrained by plausible epidemiological mechanisms. For example, in the absence of any further interventions or behavioural changes, and looking at a medium-term forecast of just incidence (daily new cases/infections), a traditional time series forecasting approach may predict a constant function based on observing broadly constant incidence, but a semimechanistic approach would expect a monotonic decrease based on a constant rate of transmission and the depletion of the susceptible population. The performance of epidemiologically constrained models is generally good (Carias et al., 2019); this is perhaps not surprising as examining the discrete renewal equation shows that these models correspond to autoregressive filters with a convex combination of coefficients specified by the generation distribution. However, similar to financial forecasting, the predictive capability of epidemic models are likely to be better interpreted as scenarios rather than actual predictions due rapidly changing policies and the unpredictable behavioural responses of human populations.

A second benefit of epidemic models is to provide a plausible mechanism to explain (noncausally) the changes observed in noisy data. For example, in estimating the effect of an intervention on observed death data, we need to consider what that intervention affects, i.e., the time-varying reproduction number R_t. As we have described above, we link the reproduction number to the number of latent infections to an observed quantity (cases, hospitalisations, or deaths) with an epidemiologically motivated mechanism. While we can statistically estimate parameters for how the intervention affects R_t, certain important parameters will be entirely unidentifiable and need to be fixed as constants or with very tight priors. For example, to reliably estimate the number of

infections, an infection fatality ratio needs to be chosen. A failure to choose an appropriate infection fatality ratio can result in a bimodal posterior where changes can either be attributed to the rapid depletion of the susceptible population (so-called "herd immunity") or to the effect of interventions. From a statistical perspective, it is difficult to disentangle which mode of the posterior best represents reality. When properly interpreted, this can be an informative finding, but obtaining epidemiologically plausible estimates from the semimechanistic model requires fixing the infection fatality ratio using estimates obtained from the literature. A second example is the fact that COVID-19 deaths occur, on average 3 weeks after infection. Omitting the infection-to-symptom and symptom-to-death distributions will bias effect estimates. This point was proved over and over when rising cases at the beginning of an epidemic wave were dismissed by the claim that hospitalisations and deaths were not rising, forgetting about the inherent lag in these measures.

Infection seeding is another fundamentally challenging aspect of epidemic modelling. Estimating the initial effect of seeding is crucial to understanding a baseline reproduction number (R_0) which is modified by behaviour, interventions, and population depletion. This seeding is heavily confounded by importation and underascertainment. Both these factors can influence estimates of the initial growth rates, and this in turn can affect the impact of changes in transmission as time progresses. We have proposed heuristic approaches to mitigate issues with early seeding, but principled statistical approaches need to be developed. In particular, Bayesian pair plots show a strong correlation between seeding parameters and R_0, which can potentially lead again to a bimodal posterior where initial growth dynamics can be explained through R_0 or via initial seed infections.

8 Confounding and causality: estimating the effect of interventions

Section 6 showed that changes in the reproduction number over time can be explained by parameterising it in terms of covariates, such as NPIs or mobility. Clarifying the effects of interventions on disease transmission is important to guide policy because NPIs have large economic and human consequences. In practice, effect sizes may not be identifiable for various reasons. NPIs often occurred in quick succession or simultaneously, leading to collinearity. They could be confounded with unobserved behavioural changes. Finally, the random distribution for the time between an infection and its recording as a case or death adds a large amount of variance to the observed data—one should not expect to see sharp or immediate changes, especially in reported deaths.

Flaxman, Mishra, Gandy, Unwin, Mellan, et al. (2020) estimated the effectiveness of NPIs across 11 European countries and used partial pooling of effect sizes to address the identification problem. At that time, little data existed other than information on deaths and the timing of interventions. NPIs, which were coded as a binary set of mandatory government measures (e.g., school closures, ban on public events, lockdown), could not fully explain the patterns seen in some countries (e.g., Sweden), and especially at the subnational level. Mobility data became available in April and was used to model the epidemic in Italy, Brazil and the USA (Mellan et al., 2020; H. J. T. Unwin et al., 2020; Vollmer et al., 2020). Such data are useful as they may help account for behavioural changes that confound the effects of NPIs. However, since mobility affects transmission and is linked to the introduction of NPIs and potentially also to voluntary behavioural measures, we expect it to be a confounder. Sharma et al. (2021) further allow for a residual stochastic process (a random walk) to be included alongside the fixed NPI effects and perform estimation at a subnational level using a randomised study design.

Section 8.2 extends the model in Flaxman, Mishra, Gandy, Unwin, Mellan, et al. (2020) to further investigate this issue of confounding, and models both NPIs and mobility jointly. This is in keeping with standard practice in regression/ANOVA: expanding a model to take into account more explanatory variables. Nonetheless, NPIs may partially affect transmission *via a pathway through mobility*. A joint model of mobility and NPIs does not account for this. Therefore, in Section 8.3 we take a first and basic step in assessing causal considerations through simple mediation analysis. We begin, however, by exploring the relationship between interventions and mobility.

8.1 Interventions and mobility

Here, we study the first epidemic wave in 2020, the same period considered by Flaxman, Mishra, Gandy, Unwin, Mellan, et al. (2020). We consider the simple case of regressing average mobility

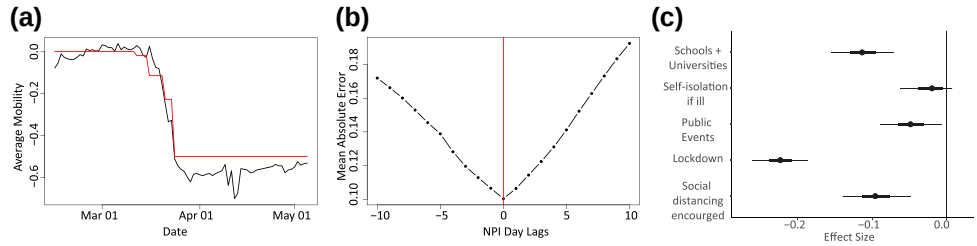

Figure 1. Simple regression of mobility against NPIs. (a) Regression prediction for the UK with mobility in black (daily time series) and the fit for NPIs in red (step function). (b) The effect on the mean absolute error from lagging the NPIs 10 days forward and backwards. (c) The coefficient effect sizes from the regression, with NPIs on the *y*-axis and regression effect sizes on the *x*-axis.

on the NPIs defined in Flaxman, Mishra, Gandy, Unwin, Mellan, et al. (2020), asking whether the changes in mobility can be explained by the timing of NPIs? Regressing average mobility on NPIs in a Bayesian linear model (no intercept or partial pooling), we find a correlation (Pearson's) of over 85% with a mean absolute error of 0.1%. Given mobility generally ranges from −1 to 1, this is a good overall fit. Figure 1a shows that visually, these fits correspond well with changes in average mobility. One could conjecture that mobility and NPIs are lagged, but lagging NPI dates either forwards or backwards in time does not result in a better fitting model, see Figure 1b. Indeed, Figure 1b supports the hypothesis that the timing of NPIs and changes in mobility are coupled. The coefficient sizes from this regression are consistent with the (Flaxman, Mishra, Gandy, Unwin, Mellan, et al., 2020) finding that the NPI with the largest effect size is a lockdown, see Figure 1c. We note that the definition of lockdown encompasses a number of specific interventions including closing workplaces and stores, banning gatherings of various sizes, stay-at-home orders, and more. A more thorough analysis would include fine-grained intervention definitions as in Sharma et al. (2021). While this regression analysis does not model transmission or the trajectory of the epidemic, it provides evidence for the consistency of mobility and NPIs, and the potential for confounding. For regularisation, we used a hierarchical shrinkage prior (Piironen & Vehtari, 2017) that performs both shrinkage and variable selection simultaneously.

8.2 Controlling for mobility

Section 8.1 found a correlation between interventions and mobility, demonstrating that mobility is a potential confounder. Here, we control for this by jointly modelling NPIs and mobility. This is done using the same 11 European countries, sets of NPIs and death data as used in Flaxman, Mishra, Gandy, Unwin, Mellan, et al. (2020).

A two-stage approach (Haug et al., 2020) is used, whereby latent R_t is first estimated through a nonparametric daily random walk, i.e., independent of NPIs. We first nonparametrically estimate R_t to account for the various lags and biases in the observed data. In the second stage, R_t is regressed on NPIs and mobility. The random walk can in theory select any arbitrary function for R_t that best describes the data without any prior information on which interventions happened, when, or how well they worked. Given these estimates of R_t for all 11 European countries, we run a simple partial pooling model to see if interventions and/or mobility can reproduce the trends in R_t. The model used is a linear regression with country-level intercepts (to account for variation in R_0), and both joint and country-specific effect sizes for interventions/mobility. As with the earlier analysis, we use a hierarchical shrinkage prior to the coefficients (Piironen & Vehtari, 2017).

Three variations of the model are considered: NPIs only, mobility only, and NPIs and mobility together. Markov chain Monte Carlo (MCMC) convergence diagnostics in all cases did not indicate fitting problems. We found the best fitting models [via PSIS-LOO (Vehtari et al., 2017)] to be NPIs alone or NPIs and mobility together. Relative to the NPIs and mobility together model the expected log posterior difference (± standard error) in WAIC of the model with only NPIs is -5.2 ± 4 (not significant) and -565.6 ± 49.2 (significant) with only mobility. Therefore, in fits to the estimated R_t, the model with mobility alone is substantially worse than the models with NPIs. Controlling for mobility does not appear to significantly change the relative ranking of

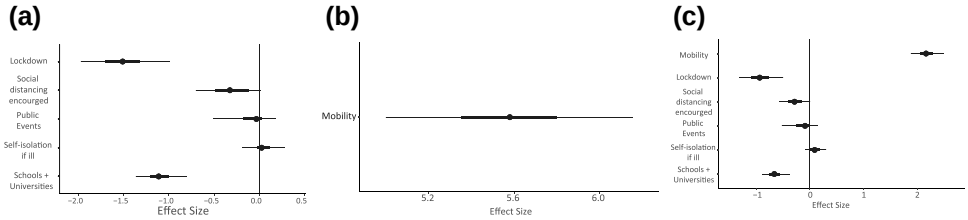

Figure 2. Regression of NPIs and/or mobility against nonparametric R_t. (a) NPIs only. (b) Mobility only. (c) NPIs and mobility together. Mobility only was not *significantly* preferred by WAIC. *Y* axis are covariates and *X* axis the regression effect sizes.

the estimated NPI effects. As in Flaxman, Mishra, Gandy, Unwin, Mellan, et al. (2020), the largest effect size is attributed to lockdown, as seen in Figure 2. This is true with and without the inclusion of the mobility variable.

An advantage of the two-stage approach is that it is scaleable to a large number of regions (e.g., Laydon et al., 2021; Nouvellet et al., 2021). R_t can be estimated in each region in parallel using separate models. Partial pooling can still be leveraged to estimate effects in the second stage. Once R_t has been estimated, any number of interesting statistical analyses can be conducted. Nonetheless, the estimated R_t is not entirely nonparametric; it is clearly influenced by the choice of using a first-order random walk in the first stage. This analysis could be extended by considering a range of alternative priors for R_t, such as Gaussian processes. More importantly, however, this approach has not considered causal relationships between NPIs and mobility. This is the focus of the next example.

8.3 Causal mediation

We would expect that part of the effect of NPIs on transmission occurs indirectly through its effect on mobility. If we a priori hypothesise that changes in mobility are both an effect of NPIs and a cause of reductions in transmission, causal mediation analysis (Pearl, 2009) provides a simple means to disentangle the total effect of a variable into a direct and indirect effect. The indirect effect occurs via some mediator, which in this case is hypothesised to be mobility.

Here, we consider lockdown on its own because performing causal mediation with all NPIs is challenging and lockdown is consistently the NPI with the largest effect size as shown above and in Flaxman, Mishra, Gandy, Unwin, Mellan, et al. (2020). As previously mentioned, the definition of lockdown represents an aggregate of policies, varying between countries. This analysis is, therefore, simply illustrative rather than being fully exhaustive. Briefly, to perform causal mediation we consider two transmission models

$$R_t^{(m)} = \tilde{R}_m^1 \exp\left(\left(\beta_1^1 + \beta_{1,m}^1\right) L_{t,m} + \epsilon_{t,m}^1\right), \tag{12}$$

$$R_t^{(m)} = \tilde{R}_m^2 \exp\left(\left(\beta_1^2 + \beta_{1,m}^2\right) L_{t,m} + \left(\beta_2^2 + \beta_{2,m}^2\right) M_{t,m} + \epsilon_{t,m}^2\right), \tag{13}$$

where $L_{t,m}$ is a binary indicator for lockdown and $M_{t,m}$ is mobility in country m respectively. $\tilde{R}_m^i$ and $\epsilon_{t,m}^i$ are country-specific parameters modelling baseline transmission and a weekly random walk respectively. All other aspects of both models are the same as in Flaxman, Mishra, Gandy, Unwin, Mellan, et al. (2020). Model (12) includes effects for lockdown, while (13) additionally considers mobility. β_1^1 is the total effect for lockdown, while β_1^2 is the partial effect when controlling for mobility. The mediated effect is, therefore, $\beta_1^1 - \beta_1^2$. This quantifies the effect of lockdown *via the pathway through mobility*. We find this mediated effect reduces R_t by 18.3% with a 95% credible interval of [12.2%, 44.4%] (see Figure 3). Individual coefficients are shown in Figure 3.

These mediation results suggest a causal link between a lockdown policy and mobility that eventually leads to reduced transmission rates. They also suggest that the mediated effect is far less than

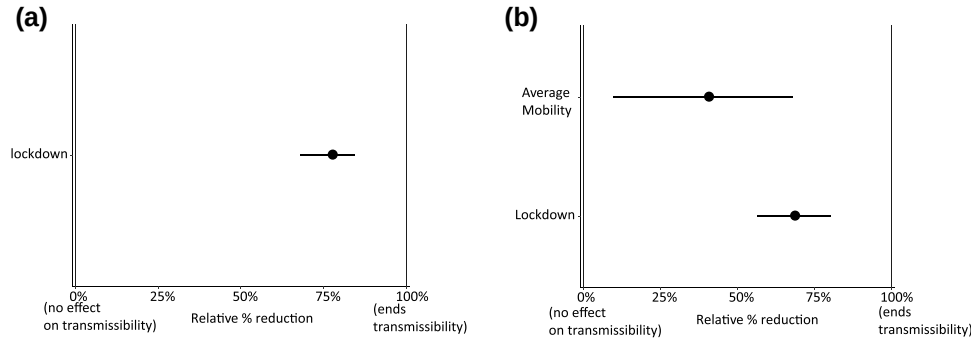

Figure 3. Mediation analysis. (a) The effect of lockdown from Model (12). (b) The effect of lockdown and mobility from Model (13).

the total effect of lockdown, suggesting lockdown will have other causal pathways. Of course, mobility is also mediated through other pathways, and a more extensive causal analysis is beyond the scope of this article. The exclusion of other NPIs may introduce omitted variable bias.

8.4 Simulation

We perform a simple simulation analysis to demonstrate the utility of our modelling framework. In these simulations, we create artificial data with properties resembling the real data (e.g., intervention timings) but with hypothetical NPI effect sizes. We fitted our modelling framework to this simulated data to see if we could recover the hypothetical NPI effect sizes. We perform three simulation experiments where, to ensure consistency, the NPI dates are selected as those that actually happened in reality (Flaxman, Mishra, Gandy, Unwin, Mellan, et al., 2020). In selecting NPI timings from real data, we could create a more plausible representation of reality with the same orderings and collinearity. For all three simulations, R_0 and the initial epidemic seeding were set as those previously used (Flaxman, Mishra, Gandy, Unwin, Mellan, et al., 2020). The model used for estimating the effect sizes for all three simulations is the partial pooling model as described in (Flaxman, Mishra, Scott, et al., 2020). The three simulation scenarios were as follows: (a) All five NPIs effect sizes were set to 20%—corresponding to the case where all interventions work equally well. (b) All five effect sizes were set to 18%, except there was an *unobserved* NPI which had an effect size of 10% and was applied at a random time at least 7 days before the last NPI occurred. This scenario corresponds to the case where there is an NPI that we did not account for in our model but which has an effect on transmission. (c) A single NPI was highly effective with an effect size of 70% and the remaining 30% were uniformly distributed among the remaining four NPIs. This scenario corresponds to a single important NPI that has the main effect in reducing transmission.

In Figure 4, we show that our model can recover the true effect sizes in all scenarios and, to a degree, motivates the appropriateness of our framework for modelling the effect of NPIs. However, our simulation experiments do not mean there are no significant shortcomings in our approach. A number of well-known problems with statistical estimation apply to our model. Residual confounding from unobserved covariates (e.g., passive NPIs, behavioural changes) as in scenario (b) can be mitigated, but in general, we recommend the inclusion of a stochastic process to model residual variation not captured by fixed effects (Sharma et al., 2021). Effect size estimates can be biased, even when residual variation is modelled. We recommend using debiasing techniques popular in statistics and machine learning (Chernozhukov et al., 2017; van der Laan & Rubin, 2006). Generalisation to different countries and times is not expected to be guaranteed (Sharma et al., 2021) and care should be taken when applying effect sizes to other countries and different time periods. Our modelling framework will inevitably be sensitive to decisions about fundamental epidemiological parameters, and these need to be regularly updated based on the best current evidence.

In summary, these simulation results show our framework is empirically motivated and is a useful approach to estimating the effect of interventions on infectious disease epidemics. However,

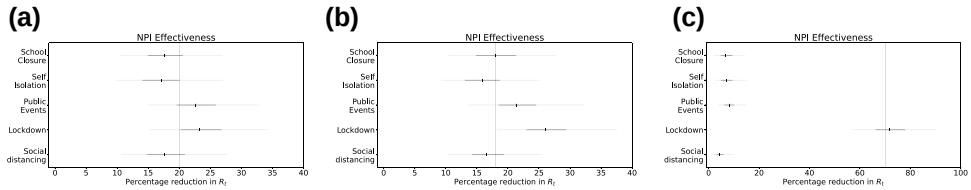

Figure 4. Simulation scenarios. (a) All NPIs have an effect size of 20%. (b) All NPIs have an effect size of 18% but there is an unobserved NPI with an effect size of 10%. (c) One NPI has an effect size of 70% and the remaining 30% is distributed to the four other NPIs.

our approach does not surmount the common statistical problems affecting regression models in general. Care must be taken with critical decisions regarding data and modelling. We believe our approach serves as a basic framework from which further development is needed (Sharma et al., 2021).

8.5 Code and software

All analysis was run using the programming language R and the software package Stan (Carpenter et al., 2017). Code for transmission modelling is available from previously published studies (Flaxman, Mishra, Gandy, Unwin, Mellan, et al., 2020; Scott et al., 2020) and an online repository at https://github.com/ImperialCollegeLondon/semi_mechanistic_renewal_processes.

9 Discussion

This article has discussed a class of Bayesian semimechanistic statistical models for epidemics such as COVID-19, which are able to capture key epidemiological mechanisms. The model has appeared in various forms for specific analyses during the COVID-19 crisis and, at the time of writing, continues to be used to inform public policy. By presenting it in a general form and discussing key modelling difficulties we hope to stimulate discussion around it. As is constantly the predominant factor in empirical statistical approaches, the model is limited by data quality and availability. In this analysis, we use a coarse definition of interventions that are likely to miss important details driving transmission. A key recommendation for future pandemic preparedness is to establish data pipelines that can quickly facilitate statistical modelling of the type outlined in this paper.

One key difficulty within the framework is dealing with confounded variables, particularly those used to explain changes in transmission during the early stages of an epidemic. The analyses in Section 8 make a first step in dealing with these. A number of model enhancements have not been included here and are an important area for further research. These include explicitly accounting for importations, allowing for uncertainty in the generation and infection-to-observation distributions, a more expressive causal model, and the inclusion of residual effects using stochastic processes. The presented model can readily be fit using probabilistic programming languages such as Stan (Stan Development Team, 2018), though we note that the adaptive Hamiltonian Monte Carlo algorithm can at times face convergence problems when latent infections are modelled directly, or when multiple regions are jointly modelled. We conjecture that convergence may be improved by carefully choosing initial parameters for the sampler. Future research could explore whether alternative samplers can be developed to fit these models more pragmatically.

Funding

S.B., S.M., and N.F. acknowledge support from the MRC Centre for Global Infectious Disease Analysis (MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and also part of the EDCTP2 programme supported by the European Union. S.B. and N.F. are funded by the National Institute for Health Research (NIHR) Health Protection Research Unit in Modelling and Health Economics, a partnership between the UK Health

Security Agency, Imperial College London and LSHTM (grant code NIHR200908). Disclaimer: "The views expressed are those of the author(s) and not necessarily those of the NIHR, UK Health Security Agency or the Department of Health and Social Care." S.B. acknowledges support from the Novo Nordisk Foundation via The Novo Nordisk Young Investigator Award (NNF20OC0059309), which also funds S.M. S.B. acknowledges support from the Danish National Research Foundation via a chair grant. S.B. acknowledges support from The Eric and Wendy Schmidt Fund For Strategic Innovation via the Schmidt Polymath Award (G-22-63345).

Supplementary material

Supplementary material are available at *Journal of the Royal Statistical Society: Series A* online.

Conflict of interest: None declared.

Appendix

A.1 Offspring dispersion

Define the offspring distribution of any given infection to be the distribution of the random number of offspring attributable to that infection. We show that assuming the variance of these distributions are a constant proportion of the mean implies, under suitable independence assumptions, the same result for new infections I_t for all time points.

Assume some ordering over infections at each period, and let $O_t^{(i)}$ denote the number of offspring of the ith infection at time t. This can be decomposed as

$$O_t^{(i)} = \sum_{s=t+1}^{\infty} O_{ts}^{(i)}, \tag{A1}$$

where $O_{ts}^{(i)}$ are the number of offspring of i birthed at time s. The branching process behind Equation (5) implies that $O_{ts}^{(i)}$ has mean $R_s g_{s-t}$. Assume that $\{O_{ts}^{(i)} : s \geq t\}$ are mutually independent and have variance which is a fixed proportion d of the mean. By Equation (A1), this implies the same variance relationship for $O_t^{(i)}$. In particular, if $R_s = R_t$ for $s > t$ then $O_t^{(i)}$ has mean R_t and variance dR_t. New infections at time t can be expressed as

$$I_t = \sum_{s=1}^{t-1} \sum_{i=1}^{I_s} O_{st}^{(i)}. \tag{A2}$$

Assume that all $O_{st}^{(i)}$ appearing in Equation (A2) are mutually independent conditional on everything occurring up to time $t-1$, the result clearly follows by taking the variance of both sides of Equation (A2) given R_t and $I_{v:t-1}$.

A.2 Population adjustment

Here, we motivate Equation (6), which is used to adjust transmission rates for the size of the infectable population. The most obvious starting point for such an adjustment would be to let

$$\mathrm{E}[I_t \mid R_t, I_{v:t-1}] = \left(\frac{S_0 - I_{t-1}}{S_0}\right) R_{u,t} L_t, \tag{A3}$$

where R_{u_t} is defined as in Section 4.2. This is similar in form to a *discrete logistic growth model*. Such models are well known as examples of simple models that exhibit chaotic dynamics (May, 1976). In particular, it is possible that the expected value on the left hand side exceeds the remaining susceptible population. Intuitively, this issue occurs because multiple infections can occur simultaneously in the discrete model. We, therefore, propose solving this by using a population adjustment motivated by the solution to a continuous time model whose intensity is a simplification of Equation (3).

Suppose we observe $I_{v:t-1}$ and current transmission R_t. We evolve infections from time $t-1$ to t continuously, and hence avoid overshooting. Define a continuous time counting $\tilde{I}(s)$ process starting at time $t-1$ by the intensity

$$\tilde{\lambda}(s) = \left(\frac{S_0 - \tilde{I}(s)}{S_0}\right) R_{u,t} L_t, \tag{A4}$$

for $s \geq t-1$, and with initial condition $\tilde{I}(t-1) = I_{t-1}$. In the online supplementary material, we show that

$$\mathrm{E}[\tilde{I}(t)] = I_{t-1} + (S_0 - I_{t-1})\left(1 - \exp\left(-\frac{R_{u,t} L_t}{S_0}\right)\right), \tag{A5}$$

which is the motivation for Equation (6).

References

Badr H. S., Du H., Marshall M., Dong E., Squire M. M., & Gardner L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases*, *20*(11), 1247–1254. https://doi.org/10.1016/S1473-3099(20)30553-3

Bellman R., & Harris T. (1948). On the theory of age-dependent stochastic branching processes. *Proceedings of the National Academy of Sciences*, *34*(12), 601–604. https://doi.org/10.1073/pnas.34.12.601

Bettencourt L. M., & Ribeiro R. M. (2008). Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS ONE*, *3*(5), e2185. https://doi.org/10.1371/journal.pone.0002185

Carias C., O'Hagan J. J., Gambhir M., Kahn E. B., Swerdlow D. L., & Meltzer M. I. (2019). Forecasting the 2014 West African Ebola Outbreak. *Epidemiologic Reviews*, *41*(1), 34–50. https://doi.org/10.1093/epirev/mxz013

Carpenter B., Gelman A., Hoffman M. D., Lee D., Goodrich B., Betancourt M., Brubaker M., Guo J., Li P., & Riddell A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Cauchemez S., Valleron A. J., Boëlle P. Y., Flahault A., & Ferguson N. M. (2008). Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*, *452*(7188), 750–754. https://doi.org/10.1038/nature06732

Champredon D., Dushoff J., & Earn D. J. (2018). Equivalence of the Erlang-distributed SEIR epidemic model and the renewal equation. *SIAM Journal on Applied Mathematics*, *78*(6), 3258–3278. https://doi.org/10.1137/18M1186411

Chernozhukov V., Chetverikov D., Demirer M., Duflo E., Hansen C., & Newey W. (2017). Double/debiased/Neyman machine learning of treatment effects. *American Economic Review*, *107*(5), 261–265. https://doi.org/10.1257/aer.p20171038

Cori A., Ferguson N. M., Fraser C., & Cauchemez S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, *178*(9), 1505–1512. https://doi.org/10.1093/aje/kwt133

Cowling B. J., Ali S. T., Ng T. W., Tsang T. K., Li J. C., Fong M. W., Liao Q., Kwan M. Y., Lee S. L., Chiu S. S., Wu J. T., Wu P., & Leung G. M. (2020). Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong; An observational study. *The Lancet Public Health*, *5*(5), e279–e288. https://doi.org/10.1016/S2468-2667(20)30090-6

Crump K., & Mode C. J. (1969). A general age-dependent branching process. II. *Journal of Mathematical Analysis and Applications*, *25*(1), 8–17. https://doi.org/10.1016/0022-247X(69)90210-8

Ferguson N. M., Donnelly C. A., & Anderson R. M. (2001). Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature*, *413*(6855), 542–548. https://doi.org/10.1038/35097116

Flaxman S., Mishra S., Gandy A., Unwin H. J. T., Mellan T. A., Coupland H., Whittaker C., Zhu H., Berah T., Eaton J. W., Monod M., Perez-Guzman P. N., Schmit N., Cilloni L., Ainslie K. E. C., Baguelin M., Boonyasiri A., Boyd O., Cattarino L., & Team I. C. C. R. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Flaxman S., Mishra S., Scott J., Ferguson N., Gandy A., & Bhatt S. (2020). Reply to: The effect of interventions on COVID-19. *Nature*, *588*(7839), E29–E32. https://doi.org/10.1038/s41586-020-3026-x

Fraser C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE*, *2*(8), e758. https://doi.org/10.1371/journal.pone.0000758

Fraser C., Donnelly C. A., Cauchemez S., Hanage W. P., Van Kerkhove M. D., Hollingsworth T. D., Griffin J., Baggaley R. F., Jenkins H. E., Lyons E. J., Jombart T., Hinsley W. R., Grassly N. C., Balloux F., Ghani A. C., Ferguson N. M., Rambaut A., Pybus O. G., Lopez-Gatell H., & Roth C. (2009). Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science*, *324*(5934), 1557–6161. https://doi.org/10.1126/science.1176062

Gelman A., & Hill J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Haug N., Geyrhofer L., Londei A., Dervic E., Desvars-Larrive A., Loreto V., Pinior B., Thurner S., & Klimek P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour*, *4*, 1303–1312. https://doi.org/10.1038/s41562-020-01009-0

Hox J. J., Moerbeek M., & de Schoot R. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

Hung, L. S. (2003). The SARS epidemic in Hong Kong: What lessons have we learned? *Journal of the Royal Society of Medicine*, *96*(8), 374–378. https://doi.org/10.1177/014107680309600803

Islam N., Sharp S. J., Chowell G., Shabnam S., Kawachi I., Lacey B., Massaro J. M., D'Agostino Sr. R. B., & White M. (2020). Physical distancing interventions and incidence of coronavirus disease 2019: Natural experiment in 149 countries. *BMJ*, *370*, m2743. https://doi.org/10.1136/bmj.m2743

Jagers P. (1969). A general stochastic model for population development. *Scandinavian Actuarial Journal*, *1969*(1–2), 84–103. https://doi.org/10.1080/03461238.1969.10405220

Kelly H. A., Mercer G. N., Fielding J. E., Dowse G. K., Glass K., Carcione D., Grant K. A., Effler P. V., & Lester R. A. (2010). Pandemic (H1N1) 2009 influenza community transmission was established in one Australian state when the virus was first identified in North America. *PLoS ONE*, *5*(6), e11341. https://doi.org/10.1371/journal.pone.0011341

Kermack W. O., & McKendrick A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, *175*(772), 700–721. https://doi.org/10.1098/rspa.1927.0118

Kimmel M. (1983). The point-process approach to age- and time-dependent branching processes. *Advances in Applied Probability*, *15*(1), 1–20. https://doi.org/10.2307/1426979

Kreft I., & de Leeuw J. (2011). *Introducing multilevel modeling*. SAGE Publications.

Laydon D. J., Mishra S., Hinsley W. R., Samartsidis P., Flaxman S., Gandy A., Ferguson N. M., & Bhatt S. (2021). Modelling the impact of the tier system on SARS-CoV-2 transmission in the UK between the first and second national lockdowns. *BMJ Open*, *11*(4), e050346. https://doi.org/10.1136/bmjopen-2021-050346

Li R., Pei S., Chen B., Song Y., Zhang T., Yang W., & Shaman J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, *368*(6490), 489–493. https://doi.org/10.1126/science.abb3221

Lloyd-Smith J. O., Schreiber S. J., Kopp P. E., & Getz W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, *438*(7066), 355–359. https://doi.org/10.1038/nature04153

May R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, *261*, 459–467. https://doi.org/10.1038/261459a0

Mellan T. A., Hoeltgebaum H. H., Mishra S., Whittaker C., Schnekenberg R. P., Gandy A., Unwin H. J. T., Vollmer M. A. C., Coupland H., Hawryluk I., Faria N. R., Vesga J., Zhu H., Hutchinson M., Ratmann O., Monod M., Ainslie K., Baguelin M., Bhatia S., … Bhatt S. (2020). *Report 21—estimating COVID-19 cases and reproduction number in Brazil* (Imperial College London Technical Report). https://doi.org/10.25561/78872

Miller A. C., Foti N. J., Lewnard J. A., Jewell N. P., Guestrin C., & Fox E. B. (2020). 'Mobility trends provide a leading indicator of changes in SARS-CoV-2 transmission,' medRxiv. https://www.medrxiv.org/content/early/2020/05/11/2020.05.07.20094441, preprint: not peer reviewed.

Mishra S., Scott J., Zhu H., Ferguson N. M., Bhatt S., Flaxman S., & Gandy A. (2020). 'A COVID-19 model for local authorities of the United Kingdom,' medRxiv. https://www.medrxiv.org/content/early/2020/11/27/2020.11.24.20236661, preprint: not peer reviewed.

Nouvellet P., Bhatia S., Cori A., Ainslie K. E. C., Baguelin M., Bhatt S., Boonyasiri A., Brazeau N. F., Cattarino L., Cooper L. V., Coupland H., Cucunuba Z. M., Cuomo-Dannenburg G., Dighe A., Djaafara B. A., Dorigatti I., Eales O. D., van Elsland S. L., Nascimento F. F., & Donnelly C. A. (2021). Reduction in mobility and COVID-19 transmission. *Nature Communications*, *12*(1), 1090. https://doi.org/10.1038/s41467-021-21358-2

Nouvellet P., Cori A., Garske T., Blake I. M., Dorigatti I., Hinsley W., Jombart T., Mills H. L., Nedjati-Gilani G., Van Kerkhove M. D., Fraser C., Donnelly C. A., Ferguson N. M., & Riley S. (2018). A simple approach to measure transmissibility and forecast incidence. *Epidemics*, *22*, 29–35. https://doi.org/10.1016/j.epidem.2017.02.012

NYS Press Office (2020). *Amid ongoing COVID-19 pandemic, governor cuomo announces state is bringing in international experts to help advise the state's reopening plan*. https://www.governor.ny.gov/news/amid-ongoing-covid-19-pandemic-governor-cuomo-announces-state-bringing-international-experts

Olney A. M., Smith J., Sen S., Thomas F., & Unwin H. J. T. (2020). Estimating the effect of social distancing interventions on COVID-19 in the United States. *American Journal of Epidemiology*, *190*(8), 1504–1509. https://doi.org/10.1093/aje/kwaa293

Pakkanen M. S., Miscouridou X., Berah T., Mishra S., Mellan T. A., & Bhatt S. (2022). 'Unifying incidence and prevalence under a time-varying general branching process', arXiv, arXiv:2107.05579, preprint: not peer reviewed.

Pearl J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, *3*, 96–146. https://doi.org/10.1214/09-SS057

Piironen J., & Vehtari A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, *11*(2), 5018–5051. https://doi.org/10.1214/17-EJS1337SI

Riley S., Fraser C., Donnelly C. A., Ghani A. C., Abu-Raddad L. J., Hedley A. J., Leung G. M., Ho L. M., Lam T. H., Thach T. Q., Chau P., Chan K. P., Lo S. V., Leung P. Y., Tsang T., Ho W., Lee K. H., Lau E. M., Ferguson N. M., & Anderson R. M. (2003). Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*, *300*(5627), 1961–1966. https://doi.org/10.1126/science.1086478

Rizoiu M.-A., Xie L., Sanner S., Cebrian M., Yu H., & Van Hentenryck P. (2017). Expecting to be hip: Hawkes intensity processes for social media popularity. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17 (pp. 735–744). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3038912.3052650

Scott J. A., Gandy A., Mishra S., Unwin J., Flaxman S., & Bhatt S. (2020). *epidemia: Modeling of epidemics using hierarchical Bayesian models*. https://imperialcollegelondon.github.io/epidemia/

Sharma M., Mindermann S., Rogers-Smith C., Leech G., Snodin B., Ahuja J., Sandbrink J. B., Monrad J. T., Altman G., Dhaliwal G., Finnveden L., Norman A. J., Oehm S. B., Sandkühler J. F., Aitchison L., Gavenčiak T., Mellan T., Kulveit J., Chindelevitch L., & Brauner J. M. (2021). Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. *Nature Communications*, *12*(1), 5820. https://doi.org/10.1038/s41467-021-26013-4

Stan Development Team (2018). The {Stan} Core Library. http://mc-stan.org/

The Scottish Government (2020). Coronavirus (COVID-19): Modelling the epidemic. https://www.gov.scot/collections/coronavirus-covid-19-modelling-the-epidemic/

Tokuda T., Goodrich B., Van Mechelen I., Gelman A., & Tuerlinckx F. (2011). *Visualizing distributions of covariance matrices* (Technical Report No. 18–18). Columbia University.

Unwin H. J. T., Mishra S., Bradley V. C., Gandy A., Mellan T. A., Coupland H., Ish-Horowicz J., Vollmer M. A. C., Whittaker C., Filippi S. L., Xi X., Monod M., Ratmann O., Hutchinson M., Valka F., Zhu H., Hawryluk I., Milton P., Ainslie K. E. C., … Flaxman S. (2020). *Report 23: State-level tracking of COVID-19 in the United States* (Imperial College London Technical Report).

Unwin H. J. T., Mishra S., Bradley V. C., Gandy A., Mellan T. A., Coupland H., Ish-Horowicz J., Vollmer M. A. C., Whittaker C., Filippi S. L., Xi X., Monod M., Ratmann O., Hutchinson M., Harrison Zhu F. V., Hawryluk I., Milton P., Ainslie K. E. C., Baguelin M., & Flaxman S. (2020). State-level tracking of COVID-19 in the United States. *Nature Communications*, *11*(1), 6189. https://doi.org/10.1038/s41467-020-19652-6

van der Laan M. J., & Rubin D. (2006). Targeted maximum likelihood learning. *International Journal of Biostatistics*, *2*(1), Article 11. https://doi.org/10.2202/1557-4679.1043

Vehtari A., Gelman A., & Gabry J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vollmer M. A. C., Mishra S., Unwin H. J. T., Gandy A., Mellan T. A., Bradley V., Zhu H., Coupland H., Hawryluk I., Hutchinson M., Ratmann O., Monod M., Walker P., Whittaker C., Cattarino L., Ciavarella C., Cilloni L., Ainslie K., & Bhatt S. (2020). A sub-national analysis of the rate of transmission of COVID-19 in Italy. medRxiv 2020.05.05.20089359.

Wallinga J., & Teunis P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, *160*(6), 509–516. https://doi.org/10.1093/aje/kwh255

WHO (2003). *SARS: Chronology of a serial killer* (Technical Report 95). WHO.

Journal of the Royal Statistical Society Series A: Statistics in Society, 2023, **186**, 616–632
https://doi.org/10.1093/jrsssa/qnad043
Advance access publication 27 April 2023
Discussion Paper

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

A sequential Monte Carlo approach to estimate a time-varying reproduction number in infectious disease models: the Covid-19 case*

Geir Storvik[1,2], Alfonso Diz-Lois Palomares[3], Solveig Engebretsen[2], Gunnar Øyvind Isaksson Rø[3], Kenth Engø-Monsen[4], Anja Bråthen Kristoffersen[3], Birgitte Freiesleben de Blasio[3,5] and Arnoldo Frigessi[5,6]

[1]Department of Mathematics, University of Oslo, Oslo, Norway
[2]Norwegian Computing Center, Oslo, Norway
[3]Department of Method Development and Analytics, Norwegian Institute of Public Health, Oslo, Norway
[4]Telenor Research, Fornebu, Norway
[5]Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway
[6]Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

Address for correspondence: Arnoldo Frigessi, Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, P.O. Box 1122 Blindern, 0317 Oslo, Norway. Email: arnoldo.frigessi@medisin.uio.no

*Read before The Royal Statistical Society in London at the second meeting on 'Statistical aspects of the Covid-19 Pandemic' organized by the Discussion Meetings Committee on Thursday, 16 June 2022, Dr Shirley Coleman in the Chair.

Abstract

The Covid-19 pandemic has required most countries to implement complex sequences of non-pharmaceutical interventions, with the aim of controlling the transmission of the virus in the population. To be able to take rapid decisions, a detailed understanding of the current situation is necessary. Estimates of time-varying, instantaneous reproduction numbers represent a way to quantify the viral transmission in real time. They are often defined through a mathematical compartmental model of the epidemic, like a stochastic SEIR model, whose parameters must be estimated from multiple time series of epidemiological data. Because of very high dimensional parameter spaces (partly due to the stochasticity in the spread models) and incomplete and delayed data, inference is very challenging. We propose a state-space formalization of the model and a sequential Monte Carlo approach which allow to estimate a daily-varying reproduction number for the Covid-19 epidemic in Norway with sufficient precision, on the basis of daily hospitalization and positive test incidences. The method was in regular use in Norway during the pandemics and appears to be a powerful instrument for epidemic monitoring and management.

Keywords: epidemic modelling, multiple data sources, real time situation awareness, reproduction number, SEIR model, sequential Monte Carlo

1 Introduction

We propose a dynamic approach for the estimation of a time-varying or instantaneous reproduction number for a mathematical infectious disease spread model. We apply our method to the Covid-19 pandemic in Norway. Like in most other countries, the pandemic has been tackled with a combination of non-pharmaceutical interventions, from social distancing to partial lock-down, imposed or advised

Received: April 20, 2023. **Accepted:** April 24, 2023

at various time points. Various viral variants with different characteristics have been competing in the population. Vaccination has also been gradually introduced. As a consequence of these changes, and of varying human behaviour, which can emerge both abruptly and smoothly, the reproduction number varies. Instantaneous estimates of reproduction numbers are useful for situational awareness. Being able to estimate such changes rapidly is important in guiding decision makers in future policy planning.

A reproduction number is precisely defined within a mathematical model of transmission. A large class of models, which has been shown to be very appropriate, is the so-called Susceptible-Exposed-Infectious-Removed (SEIR) model (S = susceptible, E = exposed, I = infected, R = recovered). In particular, *stochastic* compartmental models are preferable when there are relatively low number of infections (Keeling & Rohani, 2011, chapter 6). SEIR models are parametrized so that a meaningful estimate of the reproduction number (basic and effective) can be derived. Such compartmental models are based on many latent variables, in particular the number of individuals in each compartment within each region at every time point, and depend on many epidemiological parameters, including the transmission strengths which are key components of reproduction numbers. All unknowns must be estimated from data, which in a realistic situation are scarce and incomplete. For this reason, inference is in general very difficult, because of a high dimensional parameter space, rather flat likelihoods or posterior distributions and, often, weak identifiability (De Angelis et al., 2015).

Moreover, data which carry information about the transmissibility of the virus appear with an inevitable delay. In this paper, we use two data sources, the daily number of patients hospitalized because of Covid-19 and the daily number of positive laboratory-confirmed RNA test cases. Both these data sources carry information about the transmission of the virus in the society, but with a random time delay from transmission. Therefore, the uncertainty of the estimates of a daily reproduction number increases in the last period of data. All this is particularly challenging during an emerging epidemic. For these reasons, instantaneous reproduction numbers are rarely assumed in SEIR models. In this paper, we propose and test a sequential Monte Carlo (SMC) approach to inference which allows the efficient estimation of instantaneous reproduction numbers for the Covid-19 pandemic from real data.

While our SMC approach is generic, we implement it on top of a stochastic SEIR model which we developed for the Covid-19 epidemic (Engebretsen et al., 2021) and which was in regular use by the Norwegian health authorities. This model assumes a spatial scale resolved on county level, and uses mobile phone mobility data for the geographical spread of the virus in temporal steps of 6 hr. In Engebretsen et al. (2021), the transmissibility parameter, which represents the probability of transmission upon a contact times the contact rate in the population, is assumed to be constant in time, and is only changed at designed change-points. Inference in Engebretsen et al. (2021) is performed by a version of approximate Bayesian computation (ABC). Markov chain Monte Carlo (MCMC) convergence was essentially impossible to reach, because of the difficulty to design parameter perturbations which would lead, through the stochastic SEIR, to minor and controlled changes of the posterior distribution. It is very difficult to use ABC when the number of parameters is large, as is the case when including daily-varying reproduction numbers.

In this paper, we propose to perform Bayesian inference in combination with SMC. Static parameters related to the dynamic process for the reproduction number are estimated sequentially through SMC using sufficient statistics (Fearnhead, 2002; Storvik, 2002). Application of SMC methods is challenging because the latent processes are of high dimensions, the SEIR model is only available as a computer algorithm, and data are very limited. A further very important practical aspect when working in real time during a pandemic comes from the continuous need to improve the model, to change epidemiological assumptions, to use different data registries, and to improve the computational efficiency of algorithms. For these reasons, it has been very important to develop an SMC which is very flexible, so that changes in model specifications are easy to implement. This paper shows how a careful design of combination of model and algorithm reaches these aims and leads to a good fit to both data sources. We produce an analysis of the Norwegian pandemic, quantifying how interventions impacted the transmissibility and showing how our estimated instantaneous reproduction numbers were capturing changes rapidly enough. Based on the dynamic model, we perform future predictions of the situation for the next three weeks. We discuss the quantification of uncertainty in forecasts, which is of paramount importance for decision making.

There is an important literature on time-varying reproduction numbers applied in various epidemics, see for example Cauchemez et al. (2006) and Viboud et al. (2018). A Bayesian framework

for estimating time-varying reproduction numbers was proposed by Cauchemez et al. (2006), and applied to the SARS epidemic. In Cori et al. (2013), a time-varying reproduction number is defined as the ratio of the number $C(t)$ of infected in day t over $\sum_k w(k)C(T-k)$, where $w(\cdot)$ is the distribution of the generation time of Covid-19, which is often set to be distributed with a mean and a standard deviation estimated from specific studies (Ferretti et al., 2020). There is a very successful R-package implementing this Bayesian method, called `EpiEstim` (Cori et al., 2021). We include a comparison between `EpiEstim` and our approach based on only one of the two data sources.

A series of interesting papers has been recently presented in the Royal Statistical Society Special Topic Meeting on Covid-19 transmission (RSS, 2021). Pellis et al. (2022) give a thorough discussion of the reproduction number, its model-dependent definition, its estimation and the challenges in communicating it to the public. A model-agnostic alternative to the reproduction number is the instantaneous epidemic growth rate, and the two measures are compared in Parag et al. (2022). Mishra et al. (2022) extend in various ways the renewal model in Flaxman et al. (2020) and assume a random walk model for a weekly constant reproduction number. In order to study the effect of school closures, Bekker-Nielsen Dunbar et al. (2022) fit a multivariate time-series model to case incidence data and perform an interesting counterfactual analysis. Although in the present paper we concentrate inference about the reproduction numbers, we emphasize that several other measures can be extracted from our model, as shown in the weekly reports published by the Norwegian Institute of Public Health during the pandemic (NIPH, 2022).

Several papers have applied MCMC algorithms to estimate parameters in compartmental models (e.g Gibson & Renshaw, 1998; O'Neill et al., 2000; O'Neill & Roberts, 1999) as well as alternative models (Teh et al., 2021). The recent paper (Birrell et al., 2020) studies various SMC approaches in a SEIR model for influenza. It demonstrates the superiority of the SMC methodology compared to MCMC for such dynamical models. The paper is also useful as a general reference to SMC in epidemic modelling. Our work shares many similarities to this approach, including the use of several sources of data. A difference is our use of a dynamic model for the reproduction numbers. Also, a stochastic delay between infection and observation time is included in our setting. Another general inference framework is implemented in the R-package `pomp` (King et al., 2016). The package contains multiple different implementations of estimation procedures, including SMC, for inference for partially observed Markov process models. There are several examples of applications of the package to epidemic inference, see for example King et al. (2016) Stocks et al. (2020), and Blackwood et al. (2013).

The outline of the rest of the paper is as follows. In Section 2, the context and the data are described. In Section 3, we specify the full model, formulated as a state space model. In Section 4, we discuss how SMC algorithms can be applied in inference, including estimation of several static parameters. A simulation study and experimental results are reported in Section 5 specifically for the Norwegian Covid-19 pandemic. Additional results, including sensitivity analyses, are collected in the supplementary material. We conclude the paper with a summary and discussion. In the supplementary material details about experimental settings and algorithmic specifications are given. Data and codes are available on our GitHub repository.

2 Context and data

We start by setting the scene of the inferential task. The core is an existing model of the epidemic which has as input a set of parameters and variables, including daily reproduction numbers, and as output a series of time series of infection incidence. In our case, the model is a stochastic compartmental SEIR-type model that produces numbers of susceptible, exposed, pre-symptomatic, symptomatic and asymptomatic infectious and recovered at every time point. We also keep track of the disease incidence. We use two data time series to inform the SEIR model: the daily number of new hospital admissions of Covid-19 patients, and the daily number of laboratory-confirmed positive Polymerase chain reaction (PCR) tested cases. In order to exploit these data, we furthermore model the process of hospitalization and testing of the SEIR output, in particular of the daily incidence of infected. The inferential task is to estimate the input parameters.

The hospitalization data contain admission to all hospitals in Norway of all patients who are admitted with Covid-19 as the main cause. Admission on a certain day informs us of a transmission event that has occurred some days before. This time gap can differ between individuals. We make

several assumptions on various time lags, as specified in supplementary material. For example, the number of days between symptom onset and admission to hospital is estimated to be negative binomial distributed with parameters estimated in a separate study of the Norwegian Covid-19 registry (Whittaker et al., 2021). On average, for a patient being hospitalized, the time gap between infection and hospitalization is estimated to be approximately 14 days. For concreteness, in this paper we assumed the distributions of the various time lags to be given.

The second data set is the time series of daily number of positive PCR tests. Again, there is a time gap between onset of symptoms and testing, which we estimate to be randomly distributed with mean about 4 days. It is important for inference that the two data sets are as coherent as possible.

We use two additional data sets, which enter the SEIR model as input variables: the daily number of positive cases who have been tested in Norway but infected outside of Norway, so-called imported cases; and the total number of PCR tests made in Norway, as a surrogate of the effort made to detect positive cases.

We start with the population of Norway, distributed according to the national census in the 11 counties (see www.ssb.no/en/statbank/table/12871/). Like in Engebretsen et al. (2021), we seed our model continuously with positive cases imported from abroad on the day of recorded symptom onset or, if not available, when detected by testing. Imported cases that are hospitalized are not counted in the time series of hospitalizations, because they do not inform the model about the transmissibility of the virus in Norway. Because not all imported cases are likely to be discovered, we assume that each imported case stands for an unknown number of further undetected imported cases. We model this latent import with an additional Poisson distributed number of cases per observed imported case, with Poisson mean estimated from the data during calibration. We call this mean the amplification factor.

A final aspect, which is not central in this paper but that we mention for completeness, is that we use a geographical SEIR model on county (regional) level, so that the various compartments are geographically defined. Individuals are moved at random between the 11 counties of Norway using a mobility matrix, which is obtained every 6 hr from the movements of mobile phones in Norway, as explained in Engebretsen et al. (2021). In the present paper, all parameters in the model are however shared between counties. Even if hospitalization and test data are available at county level, in this paper we use only nationally aggregated data, because of the heterogeneity in the population size among the regions. A very important further aspect is the need to obtain inferential results as rapidly as possible, in at most a few hours, so to be able to publish results quickly just after release of the data update.

3 Model

Let $\boldsymbol{y}_t$ be the vector of hospitalization and test data on day t. Let $\boldsymbol{s}_t$ be the output vector of compartmental variables in the SEIR model at time t, for example the number of individuals in each county who are infected and symptomatic. Here, we consider the model generically as an algorithm which outputs the compartmental variables $\mathbf{s_t}$ at each time point t. R_t is the unknown reproduction number at time t. We consider the following state-space model:

$$R_t \sim P_R(R_{t-1}; \mathrm{d}R_t); \quad \text{for the reproduction number,} \tag{1a}$$

$$\boldsymbol{s}_t \sim P_s(R_t, \mathbf{s}_{t-1}; \mathrm{d}\boldsymbol{s}_t); \quad \text{representing the SEIR process,} \tag{1b}$$

$$\boldsymbol{y}_t \sim H_{y|s}(\boldsymbol{s}_{1:t}; \mathrm{d}\boldsymbol{y}_t); \quad \text{for the hospital and test data.} \tag{1c}$$

To simplify notation, we do not include the dependence of the models on the set of static parameters $\boldsymbol{\theta}$. The distribution P_R needs to be available analytically and easy to sample from. The distribution P_s is assumed to be only available through a computer algorithm and we are only able to simulate from this distribution. In certain situations, this distribution can be available as a huge and complex Markov process. However, this is not often the case, for example because of the complexity of the code or because of the lack of availability of sensitive data, like the mobility matrices in our case. The dimension of $\boldsymbol{s}_t$ is large, while R_t is low-dimensional. In this work, we consider a common scalar R_t for all counties. Note that the data $\boldsymbol{y}_t$ depend on the whole history $\boldsymbol{s}_{1:t}$ making $\boldsymbol{y}_t$

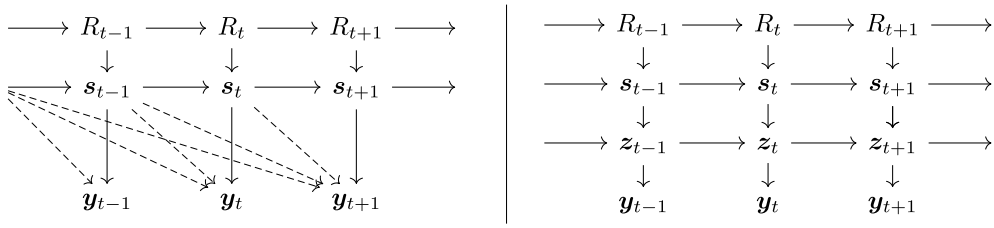

Figure 1. Graphical representation of model (1) (left) and its reformulation (right). The hyper-parameters $\boldsymbol{\theta}$ are not included and can influence all conditional distributions. The dashed arrows illustrate the dependence due to the random delays between infections and testing and possibly hospitalization.

only weakly informative about (R_t, s_t). This is due to the fact that there is a random delay from transmission to being tested and possibly hospitalized. A graphical representation of the model is given in the left panel of Figure 1. Our aim is to construct an efficient SMC method for the computation of $p(R_t, s_t \mid \boldsymbol{y}_{1:T})$ and we are interested in estimating the current status ($t = T$), in smoothing ($t < T$), and in forecasting $t > T$.

The stochastic process $\{R_t\}$ is assumed to be Markov. We suggest three alternative prior models for P_R. Let $\varepsilon_t \sim N(0, \sigma_R^2)$ be independent from all other variables. The first model is a random walk (RW) on log scale, the second model extends this to an AR(1) structure. In the third model, R_t is assumed to be piece-wise constant, with jumps occurring at random.

$$
\begin{aligned}
&R_t = \mathrm{e}^{\log(R_{t-1})+\varepsilon_t} && \text{RW, model 1}\\
&R_t = \mathrm{e}^{\mu+a(\log(R_{t-1})-\mu)+\varepsilon_t} && \text{AR, model 2}\\
&R_t = \begin{cases} R_{t-1} & \text{with probability } 1-\phi_c \\ R_{t-1}\mathrm{e}^{\varepsilon_t} & \text{with probability } \phi_c. \end{cases} && \text{PC, model 3}
\end{aligned}
$$

For all models, we assume a constant start: Until a given time point D, R_t is constant and equal to R_0. The date D is set to the day when the social distancing implementation started. Before D, the reproduction number was not likely to change significantly. In this paper, we set D equal to 8 March 2020, when teleworking started in many companies and universities in Norway. The Norwegian government announced the first package of interventions on 12 March 2020. The models above describe the dynamics after D. All the static parameters in the three models must be estimated.

The proposed method is not dependent on a specific model (1b), and would work with any epidemic model. In this work, we use the particular SEIR model described in Engebretsen et al. (2020, 2019), and whose algorithm is available though the `spread` package (Engebretsen et al., 2020).

In particular, this SEIR model has six compartments in each region (county) i: susceptible (S), exposed and not infectious (E_1), pre-symptomatic and infectious (E_2), infectious symptomatic (I), infectious asymptomatic (I_a), and recovered (R). The dynamics is described by the following equations:

$$
\begin{aligned}
S^i_{t+\delta_t} &= S^i_t - X^i_{1,t} && X^i_{1,t} \sim \text{Binom}(S^i_t, \beta_t(I^i_t + r_{I_a} I^i_{a,t} + r_{E_2} E^i_{2,t})\delta_t)\\
E^i_{1,t+\delta_t} &= E^i_{1,t} + X^i_{1,t} - X^i_{2,t} && X^i_{2,t} \sim \text{Binom}(E^i_{1,t}, \lambda_1\delta_t)/N^i_t)\\
E^i_{2,t+\delta_t} &= E^i_{2,t} + X^i_{3,t} - X^i_{4,t} && X^i_{3,t} \sim \text{Binom}(X^i_{2,t}, (1-p_a)\delta_t)\\
& && X^i_{4,t} \sim \text{Binom}(E^i_{2,t}, \lambda_2\delta_t)\\
I^i_{t+\delta_t} &= I^i_t + X^i_{4,t} - X^i_{5,t} && X^i_{5,t} \sim \text{Binom}(I^i_t, \gamma\delta_t)\\
I^i_{a,t+\delta_t} &= I^i_{a,t} + X^i_{2,t} - X^i_{3,t} - X^i_{6,t} && X^i_{6,t} \sim \text{Binom}(I^i_{a,t}, \gamma\delta_t),
\end{aligned}
\tag{2}
$$

where δ_t is 6 hr. We assume random mixing in each county in each 6 hr period, and move individuals between counties at the end of each such period according to a mobility matrix. In Engebretsen et al. (2021), mobile phone data are used to estimate such mobility matrices. These matrices report the number of individuals moving from county A to county B during each period, which we pick at random among those currently present in A, but favour the residents of B to

return to B, to capture commuting. This rule makes the computational complexity of the SEIR model quadratic in the number of counties, due to the need for storage of both current visited location and residence of individuals. The reproduction number R_t is related to β_t through the equation

$$R_t = \beta_t \left(\frac{1 - p_a + p_a r_{I_a}}{\gamma} + \frac{(1 - p_a) r_{E_2}}{\lambda_2} \right).$$

Finally, we describe the likelihood model for the data (1c). A main difficulty is the link between the latent process $\{\boldsymbol{s}_t\}$ and the observation process $\{\boldsymbol{y}_t\}$ because of the unknown stochastic delays between infection and observation time, making the computation of $H_{y|s}(\boldsymbol{s}_{1:t}; \mathrm{d}\boldsymbol{y}_t)$ hard. We introduce an auxiliary process $\boldsymbol{z}_t = (z^H_{t,0:L^H}, z^T_{t,0:L^T})$ with two components, one dedicated to the hospitalization and the other to the test data. The auxiliary variable $z^H_{t,v}$ is defined as the number of individuals who are infected at time t and hospitalized v days later. The time lag v is assumed to vary in $\{0, \ldots, L^H\}$, for some appropriate L^H. Similarly $z^T_{t,v}$ is the number of individuals who are infected at time t and tested positive v days later. We rewrite (1c) as

$$\boldsymbol{z}_t \sim G(\boldsymbol{z}_{t-1}, \boldsymbol{s}_t; \mathrm{d}\boldsymbol{z}_t), \tag{3a}$$

$$\boldsymbol{y}_t \sim H_{y|z}(\boldsymbol{z}_{1:t}; \mathrm{d}\boldsymbol{y}_t), \tag{3b}$$

where $G(\boldsymbol{z}_{t-1}, \boldsymbol{s}_t; \mathrm{d}\boldsymbol{z}_t)$ is a Markov transition distribution assumed to be easy to simulate from. Now $H_{y|z}(\boldsymbol{z}_{1:t}; \mathrm{d}\boldsymbol{y}_t)$ is easy to compute. In more details, define y^H_t to be the number of daily Covid-19 admissions to hospital. An individual who is infected at time t is hospitalized with probability p^H_u at time $t + u$. The time lag u from infection until hospitalization is assumed to follow a discrete distribution on the integers $\{0, 1, \ldots, L^H\}$. Let ρ^H_u be the probability of delay u. To structure the model further, we attach to each infected individual its potential time lag until hospitalization. So $z^H_{t,u}$ is the number of individuals infected at time t and who could possibly be hospitalized u time-units later. Let I_t be the total number of infected individuals at time t, which is available through the SEIR model as a component of $\boldsymbol{s}_t$. We can formulate the model as

$$\begin{aligned} z^H_{t,0} | I_t &\sim \text{Binomial}(I_t, \rho^H_0); \\ z^H_{t,u} | I_t, z^H_{t,0:u-1} &\sim \text{Binomial}\left(I_t - \sum_{v=0}^{u-1} z^H_{t,v}, \frac{\rho^H_u}{1 - \sum_{v=0}^{u-1} \rho^H_v} \right) \quad u = 1, \ldots, L^H; \\ y^H_t &\sim \text{Beta- Binomial}\left(\sum_{u=0}^{L} z^H_{t-u,u}, \alpha^H, \beta^H_t \right), \end{aligned} \tag{4}$$

where we here consider a Beta-Binomial distribution for a patient being hospitalized to take into account variability in hospitalization between regions. The β^H_t parameter is specified indirectly through a time-varying probability p^H_t such that $\beta^H_t = \alpha(1 - p^H_t)/p^H_t$ where p^H_t is predefined using the age–structure of the individuals having tested positive. By storing and sequentially updating the quantities $\sum_{u=0}^{L} z^H_{t-u,u}$ as well, we obtain a first-order Markovian state-space structure as illustrated in the right panel of Figure 1.

Regarding the test data, as in Engebretsen et al. (2021) we consider the probability that an infected case is tested by means of a PCR test. We ignore that tests can lead to false positive responses. The logit of this probability is assumed to be linear in the total number of daily tests N^T_t in day t in addition to a time independent intercept. We write the detection probability ρ^T_t at time t as

$$\rho^T_t = \frac{\exp(\pi_0 + \pi_1 N^T_t)}{1 + \exp(\pi_0 + \pi_1 N^T_t)}. \tag{5}$$

The time lag between infection and testing is assumed to follow a discrete distribution on $\{0, 1, \ldots, L^T\}$ for an appropriate L^T. The approach for handling this delay is exactly as for the hospital incidence, with p_t^T now playing the role of p_t^H. We introduce a new set of auxiliary variables for the test data $\{z_{t,u}^T\}$, similarly to the ones introduced for the hospitalization data. Defining $\boldsymbol{z}_t = (\boldsymbol{z}_{t,0:L^H}^H, \boldsymbol{z}_{t,0:L^T}^T)$, we are within the model formulation (3).

4 Sequential Monte Carlo

Let $\boldsymbol{x}_t = (R_t, \boldsymbol{s}_t, \boldsymbol{z}_t)$. Our aim is to perform inference on the whole set of latent variables $\boldsymbol{x}_{1:t} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ as well as on static hyper-parameters $\boldsymbol{\theta}$ at each time point t, by means of the posterior distribution $p(\boldsymbol{x}_{1:t}, \boldsymbol{\theta}|\boldsymbol{y}_{1:t})$. A description of the SMC algorithm with resampling at each step is given in Algorithm 1. See also Chopin and Papaspiliopoulos (2020), which contains more general algorithms. Although in this paper we focus on inference for R_t, also $\boldsymbol{s}_t$ will be of interest. In our setting, the main computational burden is the sampling from $P_s(R_t, \boldsymbol{s}_{t-1}; \mathrm{d}\boldsymbol{s}_t)$ which has been parallelized in our implementation. For resampling, residual resampling (Liu & Chen, 1998) has been applied. However, the resampling step is both hard to parallelize and requires message passing, resulting in that a too high number of cores can decrease performance.

Algorithm 1 Auxiliary SMC with resampling at each time step. Operations involving index b must be performed for $b = 1, \ldots, B$. Here, P_t^x denotes the transition distribution for $\boldsymbol{x}_t$, while Q_t is the proposal distribution for $\boldsymbol{x}_t$. The indices $A_t^{1:B}$ defines the ancestral particles at time t after resampling.

1: $\boldsymbol{x}_0^b \sim Q_0(\mathrm{d}\boldsymbol{x}_0)$ ▷ Proposal at time 0

2: $w_0^b = \frac{P_0(\mathrm{d}\boldsymbol{x}_0^b)H(\boldsymbol{y}_t|\boldsymbol{x}_0^b)}{Q_0(\mathrm{d}\boldsymbol{x}_0^b)}, \quad W_0^b = w_0^b / \sum_{m=1}^B w_0^m$ ▷ Calculating weights

3: $\ell_0^B = \frac{1}{B}\sum_{b=1}^B w_0^b$ ▷ Estimate of $p(\boldsymbol{y}_0)$

4: **for** $t = 1$ to T **do**

5: $\quad A_t^{1:B} = \text{resample}(W_{t-1}^{1:B})$ ▷ Resampling

6: $\quad \boldsymbol{x}_t^b \sim Q_t(\boldsymbol{x}_{t-1}^{A_t^b}, \mathrm{d}\boldsymbol{x}_t)$ ▷ Proposal at time t

7: $\quad w_t^b = w_{t-1}^b \frac{P_t(\boldsymbol{x}_{t-1}^{A_t^b}, \mathrm{d}\boldsymbol{x}_t^b)H(\boldsymbol{y}_t|\boldsymbol{x}_t^b)}{Q_t(\boldsymbol{x}_{t-1}^{A_t^b}, \mathrm{d}\boldsymbol{x}_t^b)}, \quad W_t^b = w_t^b / \sum_{m=1}^B w_t^m$ ▷ Calculating weights

8: $\quad \ell_t^B = \frac{1}{B}\sum_{b=1}^B w_t^b$ ▷ Estimate of $p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1})$

9: **end for**

The SMC algorithm is by design sequential so that by storing values of $\boldsymbol{x}_t$ obtained at the previous day, updates can easily be performed as new data arrive. A main challenge here is that the state $\boldsymbol{x}_t$ at time t heavily depends on *future* observations $\boldsymbol{y}_{t+h}$ because of the delay in hospitalization. Although the reformulated model reduces immediate dependence, there are still strong correlations backwards, as illustrated in Figure 1. There are clearly possibilities to develop more efficient proposal distributions, despite the availability of the SEIR model only as a computer algorithm. Because the main purpose of this paper is to investigate the utility of SMC methods for the estimation of daily-varying reproduction numbers, we use only simple bootstrapping proposals. We do however take into account the delay aspect by using fixed lag smoothing, using data ahead of current time, that is

$$\widehat{R}_{t|t+l_t} = E[R_t|\boldsymbol{y}_{1:t+l_t}] \approx \frac{1}{B}\sum_{b=1}^B R_t^b$$

where $\{R_t^b\}$ are simulations of R_t based on data up to time point $t + l_t$. Fixed lag smoothing with a lag of $l_t = 24$ days has been used in the Covid-19 runs in this paper. At the end of the time series, $l_t = \min\{T - t, 24\}$ is used. The estimates on the last days will be more uncertain.

4.1 Parameter estimation

Algorithm 1 assumes that the parameters $\boldsymbol{\theta}$ are known. Now we describe Bayesian inference for some of the static parameters. We denote by $\boldsymbol{\theta}_R$ the set of parameters in the model for $\{R_t\}$, $\boldsymbol{\theta}_s$ the ones in the $\{\boldsymbol{s}_t\}$ process, and $\boldsymbol{\theta}_y$ the parameters appearing in the data model. As mentioned, some of these parameters are fixed based on other data sources, and here for simplicity we do not propagate their uncertainty. For other parameters, sequential updates of their estimates are desirable. In principle, all parameters $\boldsymbol{\theta}$ could be included as part of the state vector $\boldsymbol{x}_t$ where the propagation of these static components just keeps them fixed. However, repeated resampling will quickly give degenerate samples for these parameters.

A review of parameter estimation in SMC is given in Kantas et al. (2015). Off-line methods such as particle MCMC (PMCMC, Andrieu et al., 2010) have proven to be very effective in many applications, but require repeated runs of the SMC routine. Although much smaller number of particles can be applied in such settings, some experiments with our models indicate that at least 250 particles are necessary, in which case one run uses about 10 min using 4 cores on Linux server and more cores did not help much in this case. Some experiments with an implementation of the Particle Metropolis-Hastings algorithm, based on the pseudo-marginal method by Andrieu and Roberts (2009), is reported in the online supplementary Section D. We will however focus on online methods for parameter estimation here.

In cases where sufficient statistics $\boldsymbol{v}_t(\boldsymbol{x}_{1:t})$ for the parameters are available, the SMC algorithm can be easily updated to target $p(\boldsymbol{x}_{1:t}, \boldsymbol{v}_t(\boldsymbol{x}_{1:t})|\boldsymbol{y}_{1:t})$ instead (Fearnhead, 2002; Storvik, 2002). This is the case for the $\boldsymbol{\theta}_R$ parameters. Simulations of $\boldsymbol{\theta}_R$ at each time point can then be obtained from $p(\boldsymbol{\theta}_R|\boldsymbol{v}_t(\boldsymbol{x}_{1:t}))$ which then again can be used to obtain new samples of $\boldsymbol{x}_{t+1}$ (step 6 in Algorithm 1). A crucial step is that $\boldsymbol{v}_t(\boldsymbol{x}_{1:t})$ can be recursively updated. The online supplementary Section A gives details on how $\boldsymbol{\theta}_R$ can be updated by this approach. Note that these methods can suffer from the degeneracy problem (Andrieu et al., 2005). In the online supplementary Section D, we validate the parameter estimates obtained by this procedure both through comparisons between different runs and by using the samples obtained by the Particle Metropolis-Hastings algorithm.

5 Results

5.1 Norwegian Covid-19 data

For the analysis of the Norwegian Covid-19 case, we used the 11 counties as our spatial scale. Mobility data and imported cases were used at regional scale, while hospital incidence data and test data were used at national scale. Also the $\{R_t\}$ prior process was assumed to be common for all regions. The hospital incidence data are from the Norwegian national Beredt-C19 registry and the Norwegian Intensive Care and Pandemic Registry, and the test data are from the Norwegian Surveillance System for Communicable Diseases (MSIS registry). The reproduction number is assumed constant until 7 March 2020. The results are based on data up to 1 July 2021 with test data included only from 1 August 2020, when testing capacity in Norway was scaled as needed, and after which testing criteria had become rather stable. The number of parameters is large: the dimension of $\boldsymbol{s}_t$ is 3,157. In addition, we have the reproduction numbers R_t and the auxiliary variables z_ts. The prior distributions assumed for the parameters involved in the dynamics of R_t are given in the first column of Table 1.

When running the model, we did not estimate the SEIR parameters in equation (2), nor the parameters π_0 and π_1 in equation (5), because they were estimated separately as described in Engebretsen et al. (2021). Also the parameters in equation (4) were pre-estimated through other data sources. All further details on the model are given in the online supplementary Section B. Each run is based on $B = 20{,}000$ particles. One run of the 500 days considered here, using a linux server with 128 cores, took approximately 5 hr, which is appropriate for practical real-time purposes. Figure 2 shows our estimates of R_t, with a 7-precedent-days moving-average smoothing (daily estimates are given in the online supplementary Figure S3), using the three considered models for the reproduction number. Posterior medians and symmetric 95% credibility intervals of the estimated parameters are reported in the second column of Table 1. The autoregressive (AR) model for R_t was simplified by fixing $\mu = 0$: when we estimated μ, we obtained an estimate very close to zero, but we also experienced some difficulties in the estimation of σ^2. In addition, the marginal log-likelihood (MLIK) was slightly higher. We therefore opted for the simpler AR model.

Table 1. Posterior medians and 95% credibility intervals (in parentheses) for hyper-parameters of the three models of the R_t dynamics

	Model		
Prior	RW	AR	CP
$a \sim \text{N}(0.5, 0.25)$		0.563 (0.430,0.659)	
$\frac{1}{\sigma^2} \sim \text{Gamma}(2.4, 0.28)$	0.176 (0.157,0.202)	0.340 (0.300,0.477)	0.574 (0.485,0.678)
$\phi \sim \text{Beta}(1, 9)$			0.158 (0.119, 0.198)
MLIK	−3028.20	−3008.68	−3020.69
MLIK-no test	−1252.15	−1221.09	−1230.71

Note. Although the prior for σ is defined through the precision, the numbers are for σ itself. The two last rows of the table give the value of the marginal log-likelihoods based on both data sources (MLIK) and on hospitalization data only (MLIK-no test).

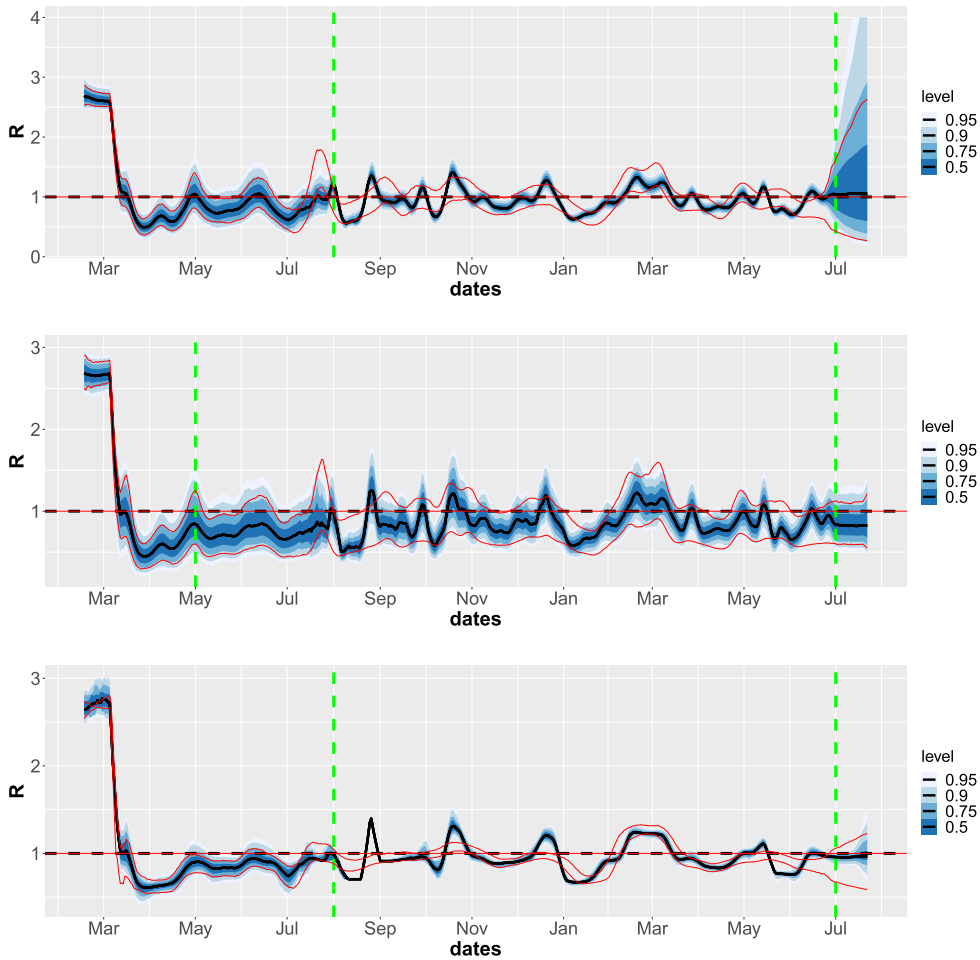

Figure 2. Estimates of the weekly averaged instantaneous reproduction number $\frac{1}{7}\sum_{u=0}^{6} R_{t-u}$ based on the random walk (RW, upper), autoregressive (AR, middle) and piece-wise constant model (PC, lower). We used hospital data from 8 March 2020 to 1 July 2021, test data from 1 August 2020 to 1 July 2021. We used $B = 20{,}000$ particles. Corresponding not-averaged daily estimates are given in the online supplementary Figure C.2. The first vertical stippled green line corresponds to the date at which test data are included in the analysis, the second stippled green vertical line corresponds to the last date with observations, after that Bayesian prediction is performed. Blue shadow bands indicate posterior uncertainty quantiles. The red curves correspond to 50% median-centred credibility intervals only based on hospitalization data.

In Figure 2, we see that the different models lead to rather similar estimates for the time dynamics of $\{R_t\}$. The AR model (central panel) seems to have more uncertainty. On the other hand, this model is to be the preferred in terms of MLIK (Table 1, last rows). Note that all models capture quickly the dramatic reduction in transmission in mid-March 2020 when Norway had the first lockdown, which was implemented between 9 and 14 March 2020. In the second half of April 2020, society was re-opened, including schools and kindergarten, and the 2-m distance rule was reduced to 1 m: $\{R_t\}$ appears to increase to around 1. A new peak appears in the end of July and beginning of August 2020, in correspondence with the end of the Norwegian vacations, the returning of Norwegians from abroad and the arrival in Norway of tourists. In particular, there have been several isolated clusters, for example in two cruise ships and the return of students to university campuses. These clusters were well controlled by contract tracing, and the reproduction number dropped rapidly below 1. The autumn 2020 was also characterized by a series of larger local outbreaks, which were rapidly controlled. The growth of R_t starting in October 2020 was also due to local outbreaks, and the incoming winter season which affects the viral transmission. The number of cases was so large that contact tracing became less effective. The Norwegian government imposed a second national intervention, first in the end of October 2020 and then again in early November 2020. This reduced the reproduction number again to below 1, in two pushes, where we can see that the second strengthening of the intervention was in fact needed for this purpose. We can see that the autumn 2020 interventions allowed the reproduction number to fall from approximately 1.5 to below 1 in about three weeks. Interestingly, we see a peak of R_t just around Christmas 2020, in connection with vacation travel and intensive shopping. January 2021 marks the arrival of the alpha variant of the virus, which was more transmissible and which increased the risk of hospitalization. The alpha variant was predominant in Norway at the end of March 2021, and we see R_t just below 1.5 again. This increase happened despite at the end of January 2021 Norway introduced the strictest lockdown rules during the whole pandemic so far, including essential closure of all borders, and vaccination of the elderly started. During March and April 2021, R_t started to decrease again, to remain below or around 1 until the middle of May 2021. Governmental interventions were reduced from May 2021, and the reproduction number stabilized around 1. At the end June 2021, approximately 50% of the adult Norwegian population had been vaccinated at least once, and approximately 30% twice. The effective reproduction number R_t reflects the effect of vaccination which is included in the SEIR model. It is remarkable how the estimated reproduction number quantify the history of the epidemic so precisely. Another feature is the cyclic behaviour of R_t, with a drop following an increase. The AR prior model on log scale also attracts towards 1. In Norway, this is expected because of the rapid intervention strategy of the government (named 'control') whenever R_t was growing rapidly above 1, and the rapid reopening when the epidemic appeared under control. Local outbreaks were frequent, also visible in the raw data, but they were rapidly controlled by appropriate contact tracing and other successful local interventions.

The red lines in Figure 2 give 50% centred credibility intervals based only on hospitalization data. We investigate in this way the value of the test data as a second source of information. Because we used test data only from August 1 2020, the estimates are essentially identical until then. After 1 August 2020, the estimates based only on hospital data are smoother, indicating that the test data contain information about transmission (and its change) that is not transferred to the hospitalization. One reason for this is that the younger generations have been infected in the autumn more than the elderly ones, who are most at risk for hospitalization. The test data also contribute to a more precise estimate of the daily prevalence of infected in Norway. We also observe some misalignment in time between the two estimates, probably because the time delays were not stationary, while we assumed them constant during the whole epidemic.

The last green horizontal line in Figure 2 corresponds to the last day with data, 1 July 2021, after which we perform a three weeks prediction, by simply running the model forward in time. The predictions for the three models have similar means but differ in uncertainty. Both the RW and the piece-wise constant model are non-stationary, which explains the high increase in uncertainty after the last observation point. On the other hand, the AR model shows a more stable prediction performance, as expected. Figure 3 gives predicted number of new daily infected cases. Before the last green dashed line, these credibility intervals give predictions based on the observed data, after this line, the predictions are obtained by running the SEIR model forward three weeks in time

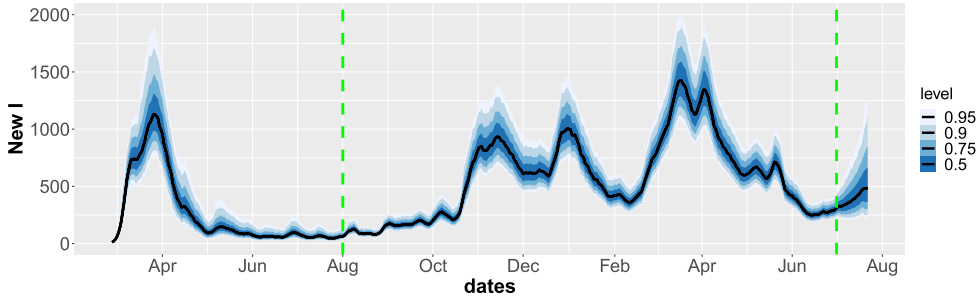

Figure 3. Predicted number of newly infected based on the autoregressive model in Figure 2. The estimates after the last dashed green line are predictions 3 weeks ahead.

using the predicted reproduction numbers. When predicting forward in time, mobility matrices, imported cases and total number of tests are needed as input. We here re-use the values in the previous 21 days in the forecasts. We see the three waves which hit Norway in March/April 2020, from November 2020 to January 2021 and in March 2021. The number of cases is estimated around 1,000 per day in the peaks (1,300 in the third wave). The number of cases would in fact grow during July 2021, as here correctly predicted.

Table 1 provides the estimates of the parameters in the three dynamic models of $\{R_t\}$. How these estimates are learned over time is shown in the online supplementary Figures C.3–C.5. The plots also allow a comparison of the prior (which is the first time point to the left) and the final posterior estimate (last time point on the right). The estimates stabilize nicely. We report about some limited validation of the parameter estimates in the online supplementary Section D. Note that the variance of the RW dynamics is estimated to be smaller than the one of the AR model, as is also clear in Figure 2. On the other hand, the variance related to the piece-wise constant model is considerably smaller even if the variability seems to be smaller in Figure 2. This is due the fact that for most time points there are no discontinuities, while when changes occur there may be large discontinuities.

In Figure 4, we use the estimated parameters, including the instantaneous reproduction numbers, to simulate the daily hospitalization incidence and the daily number of positive tests. We propagate uncertainty and produce probabilistic estimates, which we compare with the actual data. These plots show that we are able to fit both data sources well. We note the weekly structure in the test data. These plots also show three weeks ahead forecasts. The super-imposition of the actual data of these three weeks, which were not used in the analysis, show that predictions were good.

5.2 Sensitivity analysis

There are several parameters that are fixed in the current implementation:

- Hyper-parameters in the prior model for $\{R_t\}$, see Table 1.
- Parameters related to the observations: the age-dependent probabilities of being hospitalized and detected positive by testing, and the parameters describing the distribution for delay from infection to testing or hospitalization.
- Several parameters inside the SEIR model.
- The number of particles used in the SMC runs.

We did not perform a systematic sensitivity analysis, but focus on the SMC design parameters. The online supplementary Figure C.1 shows posterior median estimates of the instantaneous reproduction numbers for several different prior settings for the AR model. The figure indicates that the estimates of $\{R_t\}$ are not sensitive to the tested prior settings (credibility intervals also show similar results).

Next, we studied the importance of the number of particles. The online supplementary Figure C.7 compares results for the AR model based on $B = 20{,}000$ vs. $B = 2{,}000$ particles. The difference in the

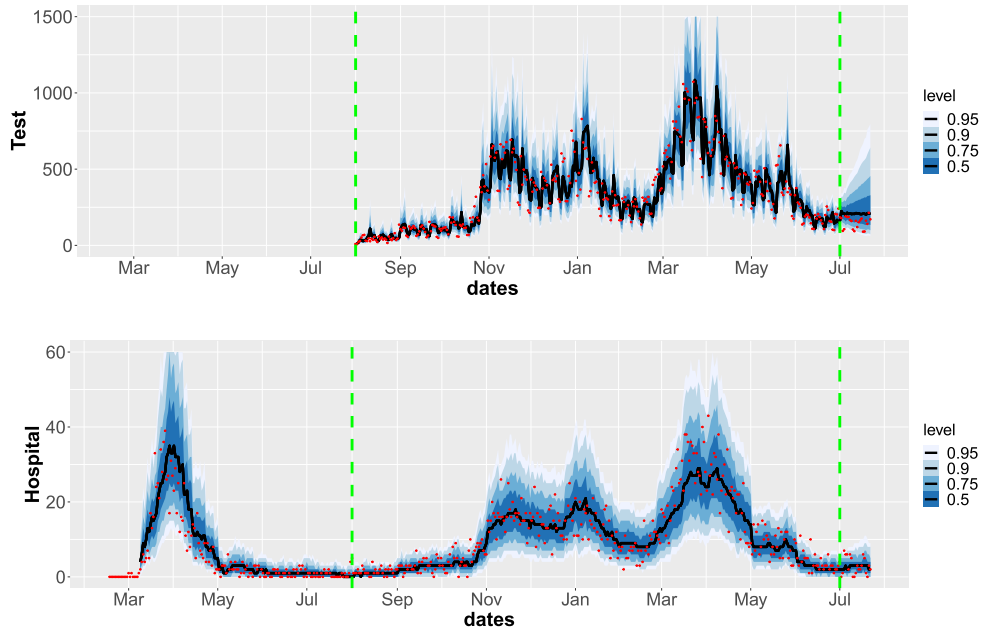

Figure 4. Predicted (blue shadows representing the quantiles of the posterior predictive distribution) and observed daily hospitalization incidence and laboratory-confirmed positive tests incidence (red dots) based on the autoregressive model.

estimated marginal likelihoods is small compared to the differences between models (about 5.0). This figure (and other similar ones, not included) shows that the results are quite stable and that $B = 2{,}000$ might have been sufficient for estimation of $\{R_t\}$. However, if interest is also in the latent structure s_t or marginal likelihood values, more particles have shown to be necessary.

5.3 Comparison with EpiEstim

EpiEstim (Cori et al., 2013) is a popular method for estimation of the reproduction number R_t based on case incidence as well as imported cases. It does however not allow for the incorporation of multiple data, so that hospital data are not taken into account, nor mobility. In Figure 5, we compare estimates of R_t obtained by our SMC approach, though only using test data and this time from the start of the epidemic, with the results obtained by using the `EpiEstim` package in R (Cori et al., 2021, version 2.2.4). For `EpiEstim` we assumed a serial interval (the time between the onset of symptoms of a primary case and the onset of symptoms of secondary cases) with mean 7.5 days and a standard deviation of 3 days. We see a very good agreement between the two estimated curves in general. However, the 95% confidence bands are unrealistically narrower for `EpiEstim`. There are some differences between the estimates during March 2020 and during the summer 2020 and November 2020. There are also some small lag differences between the estimates obtained by our approach compared to the results obtained by `EpiEstim`. This is probably because the assumed models for the delay between infection and testing times are quite different. Furthermore, the effect of the importations of cases, which is not accounted for by `EpiEstim`, may be very relevant to explain the disagreement between the models in certain periods such as in January 2021 (Christmas) and in the end of March 2021.

5.4 Simulated data

Based on an estimate $\{R_t^*\}$ obtained by the AR model from the real data, we simulated (s_t, z_t, y_t) from the model for all t. Twenty independent data sets were generated and for each case the simulated set $\{y_t\}$ was used for estimating $\{R_t\}$. Figure 6 shows the results from one of these simulated

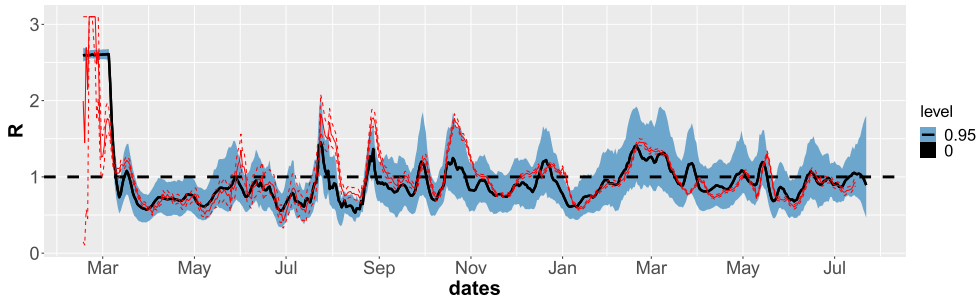

Figure 5. Estimates of R_t using the SEIR model combined with the autoregressive model for $\{R_t\}$ using SMC based only on test data (blue band, 95% confidence bands) and estimates obtained from `EpiEstim` (red lines, also 95% confidence bands).

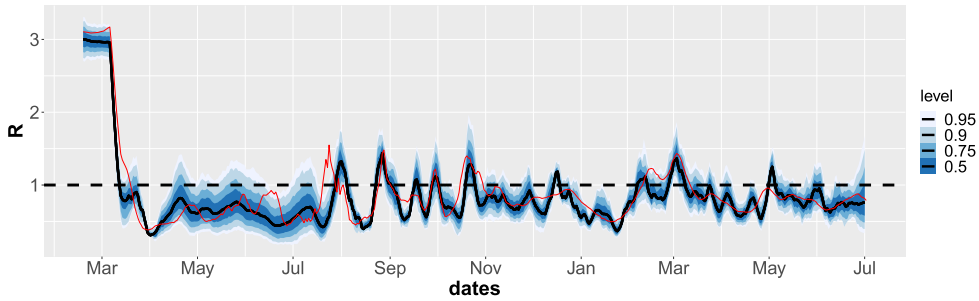

Figure 6. Simulation experiment. Estimates of $\frac{1}{7}\sum_{u=0}^{6} R_{t-u}$ using $B = 20{,}000$ (blue) based on one simulated data set. Same set-up as Figure 2, hospitalization data from 12 March 2020 to 1 July 2021, test data from 1 August 2020 to 1 July 2021. The red curve corresponds to the assumed true $\{R_t^*\}$ process.

data sets, demonstrating that the method is able to capture the main structure quite well, although in periods where the true R_t^* is considerably unstable (July/August 2020) estimates are worse. The online supplementary Figure C.6 shows results based on all 20 data sets, demonstrating the strong information content in such type of data about the $\{R_t\}$ process.

6 Discussion

A time-varying transmissibility allows to quantify the effects of interventions and changes in the behaviour of people, in real-time. This is of key importance to policy makers, as the interventions often have immense societal and health costs. Understanding, while an epidemic develops, whether the implemented interventions are sufficient or not, or if interventions could be lifted, is essential. The Norwegian government's strategy was to control the epidemic, and this was achieved by multiple national and local non-pharmaceutical interventions, which are reflected in the temporal variations of the reproduction number. To our knowledge, our approach is the first which allows to monitor a daily-varying reproduction number when using a complex compartmental model informed by multiple streams of data. The fact that our estimates of R_t react rapidly to changes in the test data, means that the situation is captured only with a delay given by the generation time of the disease under study and the time gap between transmission and testing. For Covid-19, this amounts to about a week, because of a generation time of about 5 days and a delay between transmission and test of about 2 days. Picking up an exponential growth ($R_t > 1$) before the epidemic grows out of control is essential for surveillance. The possibility of our method to validate the efficacy of contact tracing, to lead back R_t to below 1, or not, is also very important.

We have shown how daily reproduction numbers and the latent compartment-wise populations in an SEIR model can be put into a state space model, so that an SMC technique for inference can be used. Obtaining unbiased estimates of the marginal likelihoods also makes it possible to do parameter estimation within a PMCMC framework, although more work is needed here to make this computationally efficient. Compared to a parallel effort using ABC (Engebretsen et al., 2021), the SMC approach is much faster and also easier to modify with respect to model changes, confirming the findings in Birrell et al. (2020). Our implementation is modular, does not depend on the specific epidemic model (here SEIR), so that alternatives can easily be tested.

So far only simple bootstrap filters have been applied. This can be improved by utilizing more efficient proposals, alternative algorithms such as the resample-move algorithm (Gilks & Berzuini, 2001) which was used in Birrell et al. (2020) or the recent promising ideas of iterated auxiliary filters and twisted models (Guarniero et al., 2017; Heng et al., 2020). We expect these approaches to be very useful when we expand the SEIR model to have different reproduction numbers in each of the 11 Norwegian counties.

We compared three simple dynamics for R_t and found that the AR model was slightly better than the others. More work needs to be done here to compare the models in terms of prediction. In our approach, it is easy to predict the future hospitalization incidence and the number of positive cases tested. Note however that such simple dynamic models for $\{R_t\}$ are mostly suitable for now-casting and for short term forecasting, because of the lack of stationary that interventions and feedbacks imply. There are then interesting questions on how to use our probabilistic prediction of the time-varying transmission strength to propagate uncertainty, in the context of variable planned interventions.

Estimation of (static) parameters is a challenging task in SMC. Several parameters related to the SEIR model, as well as parameters related to the observation processes, were pre-estimated based on external data sources. In Engebretsen et al. (2021), we used a version of ABC. Parameters in the $\{R_t\}$ dynamic process were estimated online based on the procedure of sufficient statistics, a method that can lead to degeneracy. We have also tested out the particle Metropolis-Hastings algorithm by Andrieu et al. (2010). However, convergence was slow and challenging, because of the computational cost of running the SMC algorithm even for a small number of particles. We validated our estimates in the supplementary material, and find that our online estimation procedure worked reasonably well for the given models and data. However, some experiments with the AR model when also including estimation of the parameter μ, did cause degeneracy problems as have been reported in in Andrieu et al. (2005). A more efficient SMC algorithm might be better in utilizing the potential of such algorithms to estimate static parameters. However, our estimates of $\{R_t\}$ appear to be relatively robust with respect to changes in these parameters. Other parameters related to the SEIR model and the observation processes can be more important.

Communicating uncertainty of estimates and the effect of stochastic and uncertain time lags from data back to infections, is a major challenge. Our current strategy has been to report estimates of the reproduction numbers one week back in time as the most reliable estimates, and this needs to be studied further.

As pointed out by one of the our reviewers, it is quite easy to extend the modelling approach (and the algorithm) to include dummy variables describing interventions made by the government. In this case, one would need to estimate a time delay between the time point in which the interventions are decided and when they effect viral transmission. This delay might change in time and be intervention specific. In our model, interventions appear in the data after a delay, and are then reflected in a change of the reproduction number. It is possible to interpret changes in the estimated reproduction numbers in the light of the interventions set in place.

Finally, it is interesting to compare the estimates of the instantaneous reproduction number produced by this SMC model, with the estimates obtained by models which keep the reproduction number constant over longer time intervals, for example over four weeks. The SMC-based R_t is able to capture changes at shorter time scales significantly better, but possibly with larger uncertainty than estimates of reproduction numbers assumed to be constant over longer time periods, if the transmission has been stable during such periods. Comparing prediction power is a further aspects that can be examined. Results from our SMC model were used in the weekly reports of the Norwegian Institute of Public Health at least until April 2022, see www.fhi.no/en/publ/2020/weekly-reports-for-coronavirus-og-covid-19/.

Acknowledgments

We acknowledge the work of many colleagues at the Norwegian Institute of Public Health for the collection and preparation of the data. We thank the editor, the associate editor, and two anonymous reviewers for many valuable comments which have greatly improved the paper.

Conflict of interest: None declared.

Funding

Funding for this project was provided by the centre BigInsight (NFR project 237718), the project 'A real-time analytical pipeline for preparedness, planning and response during the Covid-19 pandemic in Norway' (NFR project 312721), and the Nordic project 'Data streams and mathematical modelling pipelines to support preparedness and decision making for Covid-19 and future pandemics' (Nordforsk NordicMathCovid project).

Data availability

The data sets analysed in this paper come from the national emergency preparedness registry for Covid-19, owned by the Norwegian Institute of Public Health. The preparedness registry is temporary and comprises data from a variety of central health registries, national clinical registries, and other national administrative registries. Further information on the registry, including access to data, is available at www.fhi.no/en/id/infectious-diseases/coronavirus/emergency-preparedness-register-for-covid-19/. See also https://nordicmathcovid.cs.aalto.fi/data/. An R-package called `smc.covid` is available at github.com/geirstorvik/smc.covid. It contain scripts for running the SMC algorithm on some data examples. Because data are sensitive, not all data sets are public. The number of individuals in hospital and the number of positive cases per day are available, but the number of imported cases and the mobility matrices are confidential. For the two latter data sets we have therefore provided some simulated synthetic data sets which resemble the true ones.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

References

Andrieu C., Doucet A., & Holenstein R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(3), 269–342. https://doi.org/10.1111/j.1467-9868.2009.00736.x

Andrieu C., Doucet A., & Tadic V. B. (2005). On-line parameter estimation in general state-space models. In *Proceedings of the 44th IEEE Conference on Decision and Control* (pp. 332–337). IEEE.

Andrieu C., & Roberts G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, *37*(2), 697–725. https://doi.org/10.1214/07-AOS574

Bekker-Nielsen Dunbar M., Hofmann F., Held L, & SUSPend Modelling Consortium. (2022). Assessing the effect of school closures on the spread of Covid-19 in zurich. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(Supplement_1), S131–S142.

Birrell P. J., Wernisch L., Tom B. D., Held L., Roberts G. O., Pebody R. G., & De Angelis D. (2020). Efficient real-time monitoring of an emerging influenza pandemic: How feasible?. *Annals of Applied Statistics*, *14*(1), 74–93. https://doi.org/10.1214/19-AOAS1278

Blackwood J. C., Cummings D. A., Broutin H., Iamsirithaworn S., & Rohani P. (2013). Deciphering the impacts of vaccination and immunity on pertussis epidemiology in thailand. *Proceedings of the National Academy of Sciences*, *110*(23), 9595–9600. https://doi.org/10.1073/pnas.1220908110

Cauchemez S., Boëlle P.-Y., Thomas G., & Valleron A.-J. (2006). Estimating in real time the efficacy of measures to control emerging communicable diseases. *American Journal of Epidemiology*, *164*(6), 591–597. https://doi.org/10.1093/aje/kwj274

Chopin N., & Papaspiliopoulos O. (2020). *An introduction to sequential Monte Carlo*. Springer Series in Statistics, Springer.

Cori A., Ferguson N. M., Fraser C., & Cauchemez S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, *178*(9), 1505–1512. https://doi.org/10.1093/aje/kwt133

Cori A., Kamvar Z., Stockwin J., Jombart T., Dahlqwist E., FitzJohn R., & Thompson R. (2021). *EpiEstim v2.2-3: A tool to estimate time varying instantaneous reproduction number during epidemics*. https://github.com/mrc-ide/EpiEstim

De Angelis D., Presanis A. M., Birrell P. J., Tomba G. S., & House T. (2015). Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics*, *10*, 83–87. https://doi.org/10.1016/j.epidem.2014.09.004

Engebretsen S., Engø-Monsen K., Aleem M. A., Gurley E. S., Frigessi A., & de Blasio B. F. (2020). Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: The case of Bangladesh. *Journal of the Royal Society Interface*, *17*(167), 1–14. https://doi.org/10.1098/rsif.2019.0809

Engebretsen S., Engø-Monsen K., Frigessi A., & de Blasio B. F. (2019). A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions. *PLoS Computational Biology*, *15*(3), e1006879. https://doi.org/10.1371/journal.pcbi.1006879

Engebretsen S., Osnes A. N., Rø G., & White R. (2020). *Spread: Infectious disease spread models (R package version 2020.9.16)*. Norwegian Institute of Public Health.

Engebretsen S., Rø G., Palomares A. D.-L., Engø-Monsen K., Kristoffersen A. B., Ruscio F. Di, Frigessi A., & de Blasio B. F. (2021). *Spatial modelling of the early-phase of the COVID-19 epidemic in Norway*. medRxiv, Cold Spring Harbor Laboratory Press. https://www.medrxiv.org/content/10.1101/2021.10.25.21265166v1

Fearnhead P. (2002). Markov chain Monte Carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics*, *11*(4), 848–862. https://doi.org/10.1198/106186002835

Ferretti L., Wymant C., Kendall M., Zhao L., Nurtay A., Abeler-Dörner L., Parker M., Bonsall D., & Fraser C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, *368*(6491), eabb6936. https://doi.org/10.1126/science.abb6936

Flaxman S., Mishra S., Gandy A., Unwin H. J. T., Mellan T. A., Coupland H., Whittaker C., Zhu H., Berah T., Eaton J. W., & Monod M. (2020). Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Gibson G. J., & Renshaw E. (1998). Estimating parameters in stochastic compartmental models using Markov chain methods. *Mathematical Medicine and Biology: A Journal of the IMA*, *15*(1), 19–40. https://doi.org/10.1093/imammb/15.1.19

Gilks W. R., & Berzuini C. (2001). Following a moving target—Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(1), 127–146. https://doi.org/10.1111/1467-9868.00280

Guarniero P., Johansen A. M., & Lee A. (2017). The iterated auxiliary particle filter. *Journal of the American Statistical Association*, *112*(520), 1636–1647. https://doi.org/10.1080/01621459.2016.1222291

Heng J., Bishop A. N., Deligiannidis G., & Doucet A. (2020). Controlled sequential Monte Carlo. *Annals of Statistics*, *48*(5), 2904–2929. https://doi.org/10.1214/19-AOS1914

Kantas N., Doucet A., Singh S. S., Maciejowski J., & Chopin N. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, *30*(3), 328–351. https://doi.org/10.1214/14-STS511

Keeling M. J., & Rohani P. (2011). *Modeling infectious diseases in humans and animals*. Princeton University Press.

King A. A., Nguyen D., & Ionides E. L. (2016). Statistical inference for partially observed markov processes via the R package pomp. *Journal of Statistical Software*, *69*(12), 1–43. https://doi.org/10.18637/jss.v069.i12

Liu J. S., & Chen R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, *93*(443), 1032–1044. https://doi.org/10.1080/01621459.1998.10473765

Mishra S., Scott J. A., Laydon D. J., Zhu H., Ferguson N. M., Bhatt S., Flaxman S., & Gandy A. (2022). A COVID-19 model for local authorities of the United Kingdom. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(Supplement_1), S86–S95. https://doi.org/10.1111/rssa.12988

NIPH (2022). *Weekly reports for coronavirus and COVID-19*. Norwegian Institute of Public Health. https://www.fhi.no/en/publ/2020/weekly-reports-for-coronavirus-og-covid-19/.

O'Neill P. D., Balding D. J., Becker N. G., Eerola M., & Mollison D. (2000). Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *49*(4), 517–542. https://doi.org/10.1111/1467-9876.00210

O'Neill P. D., & Roberts G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *162*(1), 121–129. https://doi.org/10.1111/1467-985X.00125

Parag K. V., Thompson R. N., & Donnelly C. A. (2022). Are epidemic growth rates more informative than reproduction numbers?. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(Supplement_1), S5–S15. https://doi.org/10.1111/rssa.12867

Pellis L., Birrell P. J., Blake J., Overton C. E., Scarabel F., Stage H. B., & Brooks-Pollock E. (2022). Estimation of reproduction numbers in real time: Conceptual and statistical challenges. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(Supplement_1), S112–S130. https://doi.org/10.1111/rssa.12955

RSS (2021). *Special topic meeting on Covid-19 transmission*. Royal Statistical Society https://rss.org.uk/news-publication/publications/journals/special-topic-meeting-on-r/

Stocks T., Britton T., & Höhle M. (2020). Model selection and parameter estimation for dynamic epidemic models via iterated filtering: Application to rotavirus in germany. *Biostatistics*, *21*(3), 400–416. https://doi.org/10.1093/biostatistics/kxy057

Storvik G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, *50*(2), 281–289. https://doi.org/10.1109/78.978383

Teh Y. W., Bhoopchand A., Diggle P., Elesedy B., He B., Hutchinson M., Paquet U., Read J., Tomasev N., & Zaidi S. (2021). *Efficient Bayesian inference of instantaneous re-production numbers at fine spatial scales, with an application to mapping and nowcasting the Covid-19 Epidemic in British local authorities*. https://rss. org.uk/RSS/media/File-library/News/2021/WhyeBhoopchand.pdfhttps://localcovid.info/2

Viboud C., Sun K., Gaffey R., Ajelli M., Fumanelli L., Merler S., Zhang Q., Chowell G., Simonsen L., & Vespignani A., the RAPIDD Ebola Forecasting Challenge group (2018). The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, *22*, 13–21. https://doi.org/10.1016/j.epidem.2017.08.002

Whittaker R., Kristofferson A. B., Seppälä E., Salamanca B. V., Veneti L., Storm M. L., Bøås H., Aasand N., Naseer U., Bragstad K., & Kvåle R. (2021). Trajectories of hospitalisation for patients infected with SARS-CoV-2 variant B. 1.1. 7 in Norway, December 2020–April 2021. *Journal of Infection*, *83*(4), e14–e17. https://doi.org/10.1016/j.jinf.2021.07.025

Journal of the Royal Statistical Society Series A: Statistics in Society, 2023, **186**, 633–658
Discussion Paper Contribution

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Proposer of the vote of thanks and contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Sylvia Richardson

Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge CB2 0SR, UK

Address for correspondence: Sylvia Richardson, East Forvie Building, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK. Email: sylvia.richardson@mrc-bsu.cam.ac.uk

It is a pleasure to discuss these two papers which shine light on how, in response to urgent public health needs and policy questions about effectiveness of interventions, epidemic modelling coupled with Bayesian inference and computation were judiciously used to provide decision-makers with informative evidence in near real time.

Throughout the pandemic, a range of modelling and simulation-based approaches were implemented by analysts to quantify the state of the epidemic and produce short-term forecasts, see e.g., the ensemble of models considered in the UK by the expert group scientific pandemic influenza group on modelling, ensemble which mirrored the main modelling approaches used internationally. The papers presented today represent two of the main modelling approaches and share common inference principles, but otherwise differ in several aspects. Before commenting on each paper, I thought it would be useful to highlight areas of commonality, differences, and complementarities (see Table 1).

Starting from the common root of semi-mechanistic stochastic models, as opposed, e.g., to agent-based models, the two papers are also aligned in using Bayesian inference and stochastic algorithms though of a different nature. What the Table highlights is that the focus, style, data sources, and level of technical details of the two papers are quite different.

Bhatt et al. has to be read as part of a series of papers by the Imperial team, providing a high-level discussion and outlining model extensions, while Storvik et al. is a standalone paper modelling transmission in Norway, which presents in detail the steps of the model building, computational issues, and comparative model evaluations. Aligned with their different purpose, the two papers adopt different formulations for the latent epidemic process, renewal equation versus susceptible exposed infectious recovered (SEIR) compartmental model. However, with regards to the transmission process dynamics and instantaneous reproduction number (R_t), it has been shown that these two formulations are equivalent under additional distributional assumptions (Champredon et al., 2018).

In Storvik et al., an informative comparison between the inference provided by the SEIR compartmental model and EpiEstim, which is based on the renewal equation, is presented (Figure 5). It shows a good overall correspondence between the time pattern of the respective R_t time series estimated on identical data, but a notable difference in variability estimation. That EpiEstim overconfident in its assessment of uncertainty had already been pointed out in Teh et al. (2022) (Figure 3). Demonstrating a good calibration of uncertainty as done in Storvik et al. is of paramount importance for decision-makers, something that deserves to be given additional prominence by the statistical community.

Received: April 3, 2023. **Accepted:** April 3, 2023

Table 1. Areas of commonality, differences and complementarity between Bhatt et al and Storvik et al

	Bhatt et al.	Storvik et al.
Modelling approach	Semi-mechanistic stochastic model	
Purpose	Part of a body of work, focus on developing hierarchical framework for renewal processes, evaluation of NPIs and mobility	Standalone model for transmission in Norway, short term forecasts for Public Health surveillance
Latent epidemic process	Discrete renewal equation (focus on R_t)	Regional SEIR compartmental model
Literature review	Mostly to previous papers of the Imperial modelling group	Discussion of other modelling approaches and of SMC literature
Data	Deaths in 11 countries, first wave (up to 4th May)	Hospitalisation and positive counts in Norway Between region mobility using phone data 18 months up to 01 July 2021
Observation process	Generic discussion	Beta Binomial
Estimation framework	Bayesian inference: (a) two-stage and (b) extended renewal model with log linear component	Joint Bayesian inference of SEIR compartment and transmission, large number of parameters
Computations	MCMC implemented using Stan	SMC for posterior inference produced daily, requiring tuning
Model evaluation/ comparison	Limited to model comparison of different two-stage models	Sensitivity analyses, comparison of observed and predicted, comparison with EpiEstim

From Sections 2 to 6, Bhatt et al. discuss possible extensions of the renewal approach needed: (a) to account for characteristics of the data collection and (b) to accommodate additional data sources.

(i) The observation equation, equation (2) and its alternative versions (4), (8), or (9), has two crucial elements: the ascertainment process captured through α_t, and the distribution π of the lag between infection and recorded observations. Allowing for the possibility of ascertainment bias is indeed important for data such as counts of Covid-19 positive tests, one of the statistics routinely reported worldwide. Self-selection of the population going forward to be tested increases the probability of being infected, leading to ascertainment bias. What additional information would Bhatt et al. propose to use to identify the ascertainment process parameter α_t? This is far from straightforward. Work carried out in our Turing Royal Statistical Society Health Data Lab showed that it was essential to combine data on infection prevalence from randomised surveillance with observational data to get a handle on the ascertainment bias (Nicholson et al., 2022). However, randomised surveillance data is not commonly available in many countries, and such a data synthesis exercise necessitated a carefully constructed approach to account for different granularity and quality of the data sources to be combined.

(ii) Bhatt et al. rightly encourage the simultaneous use of different types of data to inform the latent epidemic and suggest indexing all unknown quantities by data types. Planning to use simultaneously several types of data (e.g., death, hospitalisation, test counts) is natural, but making Bayesian data synthesis work concretely is more often than not challenging! In particular, delicate issues of conflict between the data sources may arise.

As discussed by Bhatt et al., delay distributions between infection and recorded events need to be estimated on external data. The quantitative estimates of these distributions will be influential on downward inference on transmission. When several data sources are available, misspecification of the corresponding delay distributions might lead to conflicting information on the state of the epidemic. It would be good to hear from Storvik et al. if they encountered any issue of data conflict and how these were resolved.

Overall, substantial progress has been made on the inferential side for semi-mechanistic models. In the future, it would be useful to pay renewed attention to the observation process and the design of data collection contributing to anchor the different parameters in semi-mechanistic models.

In Section 8.2, Bhatt et al. present a two-stage approach, where a time series of R_t for the first wave is first extracted for each of 11 countries included in the previous Flaxman et al. (2020) paper. Non pharmaceutical interventions (NPIs) and mobility data are then used as regressors with random slopes in a hierarchical linear model indexed by countries, with a view to assess the influence of mobility on the variability of transmission, independently of NPIs. It would have been helpful to include details of the source of the mobility data and its potential normalisation.

Using a two-step hierarchical model (referred here as partial pooling) for synthesising information from different geographical units (e.g., countries, regions or cities) has a long history in environmental health (see e.g., Dominici et al., 2000). Typically, the variability of the extracted summaries, here the time series of estimated Rt, is introduced in the first level of the hierarchical model. This does not seem to have been done here despite analysing a period of rapid changes in R_t, likely accompanied by larger uncertainties. Moreover, unmodelled residual autocorrelation in the errors of the linear model might affect the displayed credibility intervals. Finally, the use of shrinkage prior, whilst useful for avoiding singularity, could potentially distort the relative contribution of the regressors as they are highly correlated. All these points warrant further consideration.

In Section 8.3, the model in Flaxman et al., 2020 is usefully extended to include log linear modelling of R_t. In equations (12) and (13), besides lockdown, mobility data as well as country specific random slopes and a random walk for the residuals have been added. The causal mediation analysis interpretation of equations (12) and (13) relies on strong assumptions. In particular, clear directionality in the effect pathway (NPIs reducing mobility, in turn reducing transmission) and the assumption that there are no other confounding variables that could influence both changes in

mobility and transmission. As social behaviour influences both mobility and transmission and can be in turn be affected by current epidemic level, one must remain cautious in going down the path of a causal mediation interpretation in such a complex situation. In fine, the time pattern in the covariates reduces essentially to a 'one degree of freedom' step down. It is not clear if assuming a constant variance for the random walk modelling, the error processes ϵ^1_{tm} and ϵ^2_{tm} in (12) and (13) in such a non-stationary period might have distorted the estimation of fixed effects of lockdown in (12) and (13).

The use of compartmental models in Storvick et al. provides additional information besides estimation of R_t, such as current estimates of the number of people in the compartments and short-term predictions, which are directly useful for public health and hospital authorities planning. Stratification by region and mobile phone data was employed to refine mass action assumptions, leading to over 3,000 parameters to estimate! Hence, Storvik et al. had to dig deep into the sequential Monte Carlo (SMC) machinery to be able to provide a workable real-time implementation, which was used daily by the Norwegian public health authorities. To build their SMC proposal, they project data 24 days ahead of the current time, a computationally expensive step. Is there a trade-off to be sought between the computational burden of using such smoothing at each time step, versus reusing the same proposal for a set number of iteration steps, thus reducing the computational burden but also potentially the acceptance rates? What generic conclusion could they draw from their tailored implementation which could help future large-scale SMC implementation for compartmental models?

Storvik et al. should be congratulated for their careful evaluation and sensitivity analyses. On Figure 4, 3 weeks ahead forecasts are compared with actual data, showing good fit. It is important to display such comparisons in real time during the course of the epidemic, so that understanding and trust are built between modelling teams, public health policy makers, and the public at large and that assumptions, limitations, and good use of a modelling framework are appreciated by non-specialists.

I very much enjoyed reading both papers and would like to congratulate both teams for the important insights they have given us into the implementation of Bayesian inference for epidemic models to help public health policy makers.

Conflicts of interest: none declared.

References

Champredon, D., Dushoff, J., & Earn, D. J. (2018). Equivalence of the Erlang-distributed SEIR epidemic model and the renewal equation. SIAM Journal on Applied Mathematics, *78*(6), 3258–3278.

Dominici, F., Samet, J. M., & Zeger, S. L. (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: A hierarchical modelling strategy. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *163*(3), 263–302. https://doi.org/10.1111/1467-985X.00170

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Imperial College COVID-19 Response Team, Ghani, A. C., Donnelly, C. A., Riley, S., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C., & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature, *584*, 257–261. https://doi.org/10.1038/s41586-020-2405-7

Nicholson, G., Lehmann, B., Padellini, T., Pouwels, K. B., Jersakova, R., Lomax, J., King, R. E., Mallon, A. M., Diggle, P. J., Richardson, S., Blangiardo, M., & Holmes, C. (2022). Improving local prevalence estimates of SARS-CoV-2 infections using a causal debiasing framework. *Nature Microbiology*, *7*(1), 97–107. https://doi.org/10.1038/s41564-021-01029-0

Teh, Y. W., Bhoopchand, A., Diggle, P., Elesedy, B., He, B., Hutchinson, M., Paquet, U., Read, J., Tomasev, N., & Zaidi, S. (2022). Efficient Bayesian inference of instantaneous re-production numbers at fine spatial scales, with an application to mapping and nowcasting the covid-19 epidemic in British local authorities. https://rss.org.uk/RSS/media/File-library/News/2021/WhyeBhoopchand.pdf.

https://doi.org/10.1093/jrsssa/qnad045
Advance access publication 9 June 2023

Seconder of the vote of thanks and contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Peter J. Diggle

CHICAS, Lancaster University Medical School, Lancaster, UK

Address for correspondence: Peter J. Diggle, Lancaster University Medical School, Lancaster LA1 4YF, UK.
Email: p.diggle@lancaster.ac.uk

Both of tonight's papers describe important work that has contributed and, with sufficient political will, can and should continue to contribute to public health policy-setting and decision-making. The term 'semi-mechanistic' could be applied to the modelling approaches used in both, the models being informed by a combination of data and context-specific scientific knowledge. I find it hard to disagree with this as a general philosophy but striking the right balance is a challenge in any specific application. At root, an epidemic is a realisation of a spatio-temporal point process, but is rarely modelled as such; for an exception, see Neal and Roberts (2004). More often, context-specific knowledge feeds the temporal formulation of the model as a set of discrete-time stochastic processes, with the role of space reduced to the indexing of a small number of discrete units (Norwegian counties in Storvik et al., unspecified in Bhatt et al.). Discretising time is usually innocuous, both because of its simple geometry and the availability of data at a time-resolution fine enough to capture the dynamics of the process. Discretising space seems to me more questionable, certainly if the inferential goals include estimation of parameters that relate to the effects of spatially varying risk-factors.

Taking the purely spatial case for illustration, let $Y_i : i = 1, \ldots, n$ be case-counts in a set of n areal units A_i that partition the study region, let $S(x)$ be the (unobserved) disease risk at location x, and write $S_i = \int_{Ai} S(x)dx$. A very widely used class of models for mapping the spatial variation in risk (Besag et al., 1991) specifies the Y_i as conditionally independent Poisson variates with means

$$M_i = O_i S_i = O_i \exp\{d_i'\beta + W_i\}, \tag{1}$$

where the W_i form a Gaussian Markov random field, d_i is a vector of covariates and the offset, and O_i is the number of people at risk in A_i. Arguably a more natural model is a log-Gaussian Cox process, in which case the Y_i are again conditionally independent Poisson variates, but now with means

$$M_i = \int_{Ai} o(x)S(x)dx = \int_{Ai} o(x) \exp\{d(x)'\beta + W(x)\}, \tag{2}$$

where $o(x)$ is population density and $W(x)$ is a spatially continuous Gaussian process. Model (2) delivers the joint predictive distribution of the complete surface $S(x)$ (in practice, represented by a suitably fine pixel grid), from which the joint predictive distribution of the S_i follows immediately if that is what is required (Diggle et al., 2013). Inferences drawn from (1) and (2) can differ materially if important covariates are available at finer spatial scales than the A_i as a consequence of the long-studied effects of ecological bias (Greenland & Morgenstern, 1990). For example, the current Covid-19 epidemic in England has typically been mapped at the level of England's 331 lower tier local authorities (e.g., Riley et al., 2021) but potentially important covariates, including measures of deprivation and ethnicity, are available for each of the country's 34,753 lower super output areas; and they can vary substantially between lower super output areas within the same lower tier local authority.

Reproduction numbers, however defined, might similarly vary over much smaller spatial scales than the ones to which epidemic models are typically fitted, which leads me to wonder exactly what property of an epidemic a spatially agnostic R-number is estimating. Contrarily, I am surprised that the time-trajectories of estimated R-numbers in Figure 2 of Storvik et al. show so much high-frequency variation, even allowing for the widths of their credible intervals. Do the

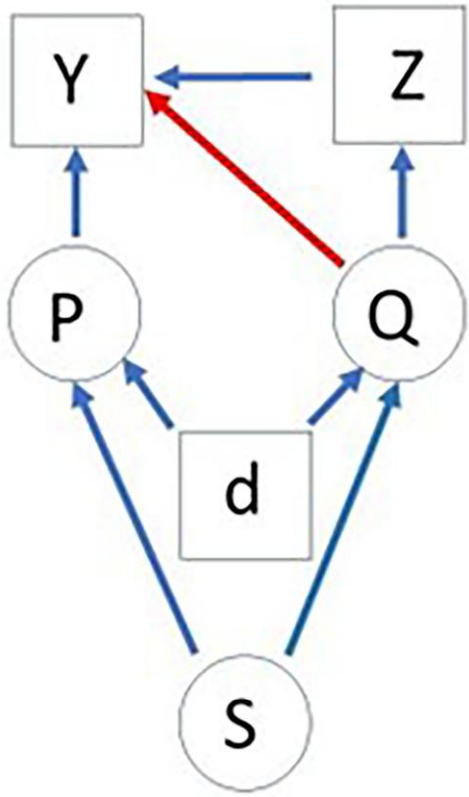

Figure 1. Causal diagram for informative non-response. Squares and circles represent observed and unobserved quantities, respectively. Observed quantities are covariates, *d*, number of responders, *Z*, and number of cases among responders, *Y*. Unobserved quantities are the response rate, *Q*, true prevalence, *P*, and a latent process, *S*, representing unobserved factors that potentially affect both *Q* and *P*. Non-response is uninformative if the red diagonal arrow is absent.

dynamics of the epidemic really change so quickly? And if they do, how useful is an estimate of the current *R*-number as a guide to real-time public health decision-making?

Finally, I would like to comment on an issue that is perhaps beyond the scope of either paper, namely the reliability of the case-numbers that inform the models. A recent WHO report gives world-wide, country-level estimates of the very substantial under-counts in reported COVID-19 deaths (Knutson et al., 2022), and the problem can only be worse for case-counts of a disease with, thankfully, relatively low mortality. Nicholson et al. (2022) show how to de-bias case-count data by jointly modelling routinely reported case-counts together with data from a randomised (hence unbiased) prevalence study, the real-time assessment of community transmission (REACT) study (Riley et al., 2021). REACT was an England-wide repeated cross-sectional study of the spatio-temporal variation in Covid-19 prevalence over the course of the epidemic, based on Covid-19 positive/negative test results from samples, each of approximately 100,000 individuals drawn, on each of 19 rounds over almost two years, as geographically stratified random samples of the national health service-registered population. The response rate was of the order of 20–25%, which the REACT analysis accommodated by adjusting for imbalance in the responders with respect to potential risk-factors. Nevertheless, this leaves a niggling doubt that in studies of this kind, responders within a risk-group may not be representative of the whole group. In the setting of longitudinal studies, this corresponds to the problem of informative missingness, on which topic many papers and at least one book (Daniels & Hogan, 2008) have been written. The prevalence survey counterpart is harder to address because, unlike in longitudinal studies, there no independent replication. Figure 1 outlines a possible solution. The predictive target is the true prevalence, *P*, but in the case of informative non-response, the empirical prevalence depends on both *P* and *Q*.

Conflicts of interest: None declared.

References

Besag, J., York, J., & Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*(1), 1–20. https://doi.org/10.1007/BF00116466

Daniels, M. J., & Hogan, J. W. (2008). *Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis*. Chapman and Hall/CRC.

Diggle, P. J., Moraga, P., Rowlingson, B., & Taylor, B. (2013). Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Statistical Science*, *28*(4), 542–563. https://doi.org/10.1214/13-STS441

Greenland, S., & Morgenstern, H. (1990). Ecological bias, confounding and effect modification. *International Journal of Epidemiology*, *18*(1), 269–274. https://doi.org/10.1093/ije/18.1.269

Knutson, V., Aleshin-Guendel, S., Karlinsky, A., Msemburi, W., & Wakefield, J. (in press). Estimating country-specific excess mortality during the COVID-19 epidemic. *Annals of Applied Statistics*.

Neal, P. J., & Roberts, G. O. (2004). Statistical inference and model selection for the 1861 Hagelloch measles epidemic. *Biostatistics (Oxford, England)*, *5*(2), 249–261. https://doi.org/10.1093/biostatistics/5.2.249

Nicholson, G., Lehmann, B., Padellini, T., Pouwels, K. B., Jersakova, R., Lomax, J., King, R. E., Mallon, A.-M., Diggle, P. J., Richardson, S., Blangiardo, M., & Holmes, C. (2022). Local prevalence of transmissible SARS-CoV-2 infection: An integrative causal model for debiasing fine-scale targeted testing data. *Nature Microbiology*, *7*(1), 97–107. https://doi.org/10.1038/s41564-021-01029-0

Riley, S., Ainslie, K. E. C., Eales, O., Walters, C. E., Wang, H., Atchison, C., Fronterre, C., Diggle, P. J., Ashby, D., Donnelly, C. A., Cooke, G., Barclay, W., Ward, H., Darzi, A., & Elliott, P. (2021). Resurgence of SARS-CoV-2 in England: detection by community antigen surveillance. *Science*, *372*, 990–995.

The vote of thanks was passed by acclamation

https://doi.org/10.1093/jrsssa/qnad044
Advance access publication 11 April 2023

Arun Chind's contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Arun Chind

Proshen Health & Risk Consulting Ltd, 87 Dacre Avenue, Wakefield WF2 8AH, UK

Address for correspondence: Dr Arun Chind, Proshen Health & Risk Consulting Ltd, 87 Dacre Avenue, Wakefield WF2 8AH, UK. Email: arunpeter@yahoo.com

I have found the papers by Bhatt et al. and Storvik et al. illuminating and applaud their efforts that have assisted in containing the COVID pandemic. I appreciate that a member of this audience—from Brazil—acknowledged using methods from Bhatt et al. for their own work.

I have tried to replicate at least part of the results of Storvik et al. using their own instructions from the paragraph entitled 'Code and data', but was unable to get the code to work. While I appreciate that computational techniques themselves can be complex, unless the techniques presented in an article have a realistic chance of being replicated by other researchers, I fear that the underlying science may be inaccessible not only for current researchers but also for posterity. Such inaccessibility tends to undermine the very purpose of publication of research results.

I would be interested to know whether the official reviewers of the article by Storvik et al. (Professor Diggle, perhaps) had any success in replicating at least part of the results. If not, we have an uncomfortable situation where we have to take the authors' results as given.

In order to test the value of provided code and data sets of the *Journal of the RSS Series C*, I ran the code supplied by the authors of the first three articles that used *R* software (which I am familiar with), available on the web page: https://rss.onlinelibrary.wiley.com/hub/journal/14679876/

series-c-datasets/69_5. To my disappointment, a barrage of error messages greeted my efforts. Although use of a sample of three would not pass for a study design, I fear that I have stumbled upon a problem pattern involving author-supplied code and data that needs addressing. I believe that the progress of statistics should not be obfuscated or thwarted by researchers' *mathe-magic* or computational sleight of hand.

Whereas the reproduction number of the COVID-19 virus has rightly focused our attention, the matter of reproducibility of research studies remains a persistent problem in many areas of science (Baker, 2016). For the future, as a condition for publication, I propose that the *RSS* should consider requiring that at least the reviewers be able to replicate and vouch for the reproducibility of at least some if not all of researchers' results. I wish to see the day when authors who genuinely believe in the value of their research would put up a video (say, on *YouTube*) demonstrating how others could replicate their results.

Acknowledgments

I wish to acknowledge that following a formal request by the editor, just before the article by Storvik et al went to press, one of the paper's coauthors, Dr Palomares arranged a Microsoft Teams call with me. Dr Palomares shared his laptop computer screen and kindly took me through a cutdown version of the simulations and demonstrated the satisfactory working of the R code (available on GitHub) that produced representative simulation results.

Conflicts of interest: None declared.

Reference

Baker M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*, 452–454. https://doi.org/10.1038/533452a

https://doi.org/10.1093/jrsssa/qnad046
Advance access publication 5 April 2023

Paul J Birrell, Angelos Alexopoulos and Daniela De Angelis's contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Paul J. Birrell[1,2], Angelos Alexopoulos[2] and Daniela De Angelis[1,2]

[1]UK Health Security Agency, London, UK
[2]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Address for correspondence: Paul J. Birrell, UK Health Security Agency, 61 Colindale Avenue, London NW9 5EQ, UK; MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. Email: paul.birrell@ukhsa.gov.uk

We congratulate the authors on the work of this paper, particularly as we are all too familiar with the effort required to maintain pandemic surveillance in real-time, while generating inference to hit a consistently moving target. The authors generously cite related pre-pandemic work (Birrell et al., 2020) developed to provide a framework for online inference during a pandemic, particularly in the presence of what is termed a 'system shock', where there is either a clear step change in the pandemic dynamics or in the pandemic surveillance data itself.

However, this work focussed on estimation of static parameters and short-term epidemic forecasts. Storvik et al. are attempting to estimate a dynamic process within a triply stochastic model:

the reproduction number, $R_e(t)$, evolves according to a stochastic process and both the transmission dynamics and the model for reporting infections are also stochastic. Given this complexity, a pragmatic decision is made to implement a bootstrap particle filter (PF) for the estimation of model states (including $R_e(t)$). Filters of this type suffer sample degeneracy for historic model states, but perhaps this is not problematic when the focus of analysis is nowcasting and forecasting.

The absence of a jittering step to replenish the sample in the presence of degeneracy does mean that such PFs are not robust to system 'shocks'. One such 'shock' would be the national lockdowns implemented to interrupt transmission chains. In this case, however, the shock occurs in the very early stages of the pandemic and the impacted stochastic process ($R_e(t)$) is initiated from this point onwards much in the same way as we did in our response to the UK pandemic (Birrell et al., 2021).

This motivates a unifying framework that can be swiftly deployed in the event of a pandemic, extending Birrell et al. (2020) to models with stochasticity beyond observation error. The time course nature of the data promotes Sequential Monte Carlo (SMC), and in particular SMC-within-SMC, (SMC^2; Chopin et al., 2013; Chopin & Papaspiliopoulos, 2020). The PF in the static dimension, which utilises estimates of the marginal likelihood obtained from PF algorithms in the stochastic dimension, could be constructed using ideas from Birrell et al. (2020), guaranteeing a high effective sample size for the parameters of the mechanistic model, limiting the overall computational effort required. Unlike Markov chain Monte Carlo, SMC should pose fewer challenges to real-time inference, provided, as the authors claim, there really is a minimal need to refresh the sample.

Conflicts of interest: None declared.

References

Birrell, P. J., Blake, J., van Leeuwen, E., Gent, N., & De Angelis, D. (2021). Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. *Philosophical Transactions of the Royal Society, Series B*, *376*(1829), 20200279. https://doi.org/10.1098/rstb.2020.0279

Birrell, P. J., Wernisch, L., Tom, B. D., Held, L., Roberts, G. O., Pebody, R. G., & De Angelis, D. (2020). Efficient real-time monitoring of an emerging influenza pandemic: How feasible? *Annals of Applied Statistics*, *14*(1), 74–93. https://doi.org/10.1214/19-AOAS1278

Chopin, N., Jacob, P. E., & Papaspiliopoulos, O. (2013). SMC^2: An efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society Series B*, *75*(3), 397–426. https://doi.org/10.1111/j.1467-9868.2012.01046.x

Chopin, N., & Papaspiliopoulos, O. (2020). *An introduction to sequential Monte Carlo*. New York: Springer.

https://doi.org/10.1093/jrsssa/qnad047
Advance access publication 20 April 2023

Sanmitra Ghosh's contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Sanmitra Ghosh Mr.

MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Address for correspondence: Sanmitra Ghosh, MRC Biostatistics Unit, University of Cambridge, East Forvie Building, Forvie Site, Robinson Way, Cambridge Biomedical Campus, Cambridge, CB2 0SR, UK. Email: sanmitra.ghosh@mrc-bsu.cam.ac.uk

I like to thank the authors for their timely and inspirational work on real-time inference of the time-varying *reproduction number* R_t. Recently, a real-time inference scheme for a complex, yet deterministic,

transmission model was developed in Birrell et al. (2020). Such a deterministic model lacks the ability to capture the inherent stochasticity present in the dynamics of a population process. This was improved upon in Birrell et al. (2021) by introducing *environmental stochasticity*, where $R_t = f(\beta_t)$ was modelled as a function of a time-varying transmission potential β_t described as a piece-wise random-walk. It is commendable that the authors have considered a fully stochastic model where in addition to modelling R_t as a stochastic process (thus capturing environmental stochasticity), the *demographic stochasticity* is also considered by modelling the compartmental dynamics as a stochastic process.

Bayesian inference of a stochastic epidemic model is essentially a 'missing-data' problem that is generally solved using a data-augmentation-based Markov chain Monte Carlo (MCMC) scheme (O'Neill & Roberts, 1999). The authors have explored the particle MCMC algorithm (Andrieu et al., 2010) for this purpose. However, the computational burden was exorbitant and thus joint estimation of the parameters and the compartmental states was avoided.

In order to alleviate this computational hindrance, I suggest the usage of diffusion processes for representing the stochastic components (see Allen, 2008; Fuchs, 2013 for further details). Diffusion processes have differentiable approximations that enable the application of highly efficient MCMC samplers (Graham et al., 2019) which can rapidly explore the target density. Moreover, differentiability can be leveraged to apply variational inference (Ghosh et al., 2022; Li et al., 2020; Ryder et al., 2018) for additional scalability. Gradients can be evaluated using the *adjoint* technique of automatic differentiation (Chen et al., 2018; Ghosh et al., 2021; Li et al., 2020).

Let me point to a particular approximation that can be used to circumvent data-augmentation altogether. Consider a univariate Stratonovich stochastic differential equation (SDE):

$$\mathrm{d}X_t = a(X_t, \boldsymbol{\theta})\mathrm{d}t + b(X_t, \boldsymbol{\theta}) \circ \mathrm{d}W_t, \tag{1}$$

where $a(\cdot)$ and $b(\cdot)$ are the drift and diffusion terms with parameters $\boldsymbol{\theta}$, and W_t is a Brownian motion. The above SDE can be approximated with the following (random) ordinary differential equation (ODE) (Lyons et al., 2014):

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = a(X_t, \theta) + b(X_t, \theta) \sum_{i=1}^{N} Z_i \phi_i(t), \tag{2}$$

where $Z \sim \mathcal{N}(0, 1)$ and $\phi(t)$ is an orthonormal basis function. The ODE in (2) converges to (1) when $N \to \infty$ (Wong & Zakai, 1965). See Shmatkov (2006) for the convergence of the approximation of a multivariate SDE. Given $\{Z_i\}_{i=1}^{N}$, $\boldsymbol{\theta}$ the path of X_t is available deterministically and thus there is no need of data-augmentation. By utilising this approximation, gradient-based inference of diffusion was carried out in Ghosh et al. (2022).

I have recently used this technique to represent the transmission potential β_t as an approximate Brownian motion, within the model proposed in Birrell et al. (2021), replacing the piece-wise constant random-walk which required a computationally burdensome data-augmentation scheme for inference. This approximation made it computationally feasible to infer β_t at a high time resolution producing a smooth estimate of R_t, see Ghosh et al., (2023).

Conflict of interest: There are no conflicts to declare.

Data availability

This manuscript includes no presentation or analysis of data.

References

Allen L. J. S. (2008). An introduction to stochastic epidemic models. In F. Brauer, P. van den Driessche, & J. Wu (Eds.), *Lecture Notes in Mathematics: Vol. 1945. Mathematical Epidemiology* (pp. 81–30). Berlin: Springer.

Andrieu C., Doucet A., & Holenstein R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *72*(3), 269–342.

Birrell P., Blake J., Van Leeuwen E., Gent N., & De Angelis D. (2021). Real-time nowcasting and forecasting of Covid-19 dynamics in England: The first wave. *Philosophical Transactions of the Royal Society B*, *376*(1829), 20200279.

Birrell P. J., Wernisch L., Tom B. D. M., Held L., Roberts G. O., Pebody R. G., & De Angelis D. (2020). Efficient real-time monitoring of an emerging influenza pandemic: How feasible? *The Annals of Applied Statistics*, *14*(1), 74.

Chen T. Q., Rubanova Y., Bettencourt J., & Duvenaud D. K. (2018). Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett, (Eds.), *Advances in neural information processing systems* (pp. 6571–6583).

Fuchs C. (2013). *Inference for diffusion processes: With applications in life sciences*. Springer Science & Business Media.

Ghosh S., Birrell P., & De Angelis D. (2021). Variational inference for nonlinear ordinary differential equations. In A. Banerjee, & K. Fukumizu (Eds.), *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, PMLR 130 (pp. 2719–2727).

Ghosh S., Birrell P. J., & De Angelis D. (2022). Differentiable Bayesian inference of SDE parameters using a pathwise series expansion of Brownian motion. *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, PMLR 151 (pp. 10982–10998).

Ghosh, Sanmitra, Birrell, Paul J, De Angelis, Daniela, & Perkins, Alex. (2023). An approximate diffusion process for environmental stochasticity in infectious disease transmission modelling. *PLOS Computational Biology*, *19*(5), e1011088.

Graham M. M., Thiery A. H., & Beskos A. (2019). Manifold Markov chain Monte Carlo methods for Bayesian inference in diffusion models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *84*(4), 1229–1256.

Li X., Wong T.-K. L., Chen R. T. Q., & Duvenaud D. K. (2020). Scalable gradients for stochastic differential equations, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, PMLR 108 (pp. 3870–3882).

Lyons S. M. J., Särkkä S., & Storkey A. J. (2014). Series expansion approximations of Brownian motion for nonlinear Kalman filtering of diffusion processes. *IEEE Transactions on Signal Processing*, *62*(6), 1514–1524.

O'Neill P. D., & Roberts G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *162*(1), 121–129.

Ryder T., Golightly A., McGough A. S., & Prangle D. (2018). Black-box variational inference for stochastic differential equations, *Proceedings of the 35th International Conference on Machine Learning*, PMLR 80:4423-4432.

Shmatkov A. (2006). *Rate of convergence of Wong-Zakai approximations for SDEs and SPDEs* [Ph.D. thesis]. The University of Edinburgh.

Wong E., & Zakai M. (1965). On the convergence of ordinary integrals to stochastic integrals. *The Annals of Mathematical Statistics*, *36*(5), 1560–1564.

The following contributions were received in writing after the meeting.

https://doi.org/10.1093/jrsssa/qnad048
Advance access publication 20 October 2023

Alice Corbella, Anne M Presanis, Paul J Birrell and Daniela De Angelis's contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Alice Corbella[1,2], Anne M. Presanis[2], Paul J. Birrell[2,3] and Daniela De Angelis[2,3]

[1]Department of Statistics, University of Warwick, Coventry, UK
[2]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[3]UK Health Security Agency, London, UK

Address for correspondence: Alice Corbella, Department of Statistics, University of Warwick, Coventry CV4 7AL, UK; MRC Biostatistics Unit, University of Cambridge, UK. Email: alice.corbella@warwick.ac.uk

We congratulate Storvik and co-authors on this excellent work that combines elaborate models and cutting-edge inference in an impactful setting.

The model used to describe the dynamics that contribute to the generation of COVID-19 data comprises environmental, demographic, and observational randomness, with the final goal of appropriately allocating the stochasticity to each of these three components.

Among several innovations that are presented, we have particularly appreciated the presence of stochastic delays in the severity process and their handling in the Bayesian inference using sequential Monte Carlo (SMC). In our related work (Corbella et al., 2022), we highlight that the severity process plays a substantial role in generating randomness observable in the data, and that the delay between infection and the case becoming detectable via (severe) symptoms should not be ignored. A substantial body of related work to account for delays between infection and symptom onset/observation traces back to stochastic back-calculation methods used to estimate incidence of HIV/AIDS (Brookmeyer et al., 1994) and has advanced since (e.g., Brizzi et al., 2019; Marschner, 1994; Mezzetti & Robertson, 1999; Sweeting et al., 2005).

In some contexts, the use of multiple data streams, such as the community testing data and hospital admissions used by Storvik et al., should be accompanied by an assessment of potential dependency between the datasets. The authors have considered these two streams as independent signals of the same underlying process, which may be plausible for a context where those hospitalised are unlikely to have had time to be tested prior to admission. However, there are other situations (e.g., Corbella et al., 2022) where accounting for possible hidden dependence between data streams is important.

Lastly, we have enjoyed reading the description of the computational methods very much. In our experience, the implementation of the SMC sampler is often non-trivial but enables inference when richer models are assumed, such as the three models adopted for R_t. In our understanding, the authors have exploited results from a previous study on the same data (Engebretsen et al., 2021), to fix several parameters of the transmission model, enabling the fast inference using SMC. Does such use of the data in two stages not risk overstating the authors' certainty about R_t? Perhaps a two-phase approach, where the static parameters are inferred using data on the first wave, then are assumed known at the beginning of the second wave, to drive the SMC inference, might be more realistic.

Conflicts of interest: None declared.

Funding

This research was funded in whole or in part by EPSRC grant New Approaches to Bayesian Data Science: Tackling Challenges from the Health Sciences [AC, DDA: EP/R018561/1], and by the UKRI Medical Research Council [DDA, AMP, PJB: Unit Programme number MC/UU/00002/11].

Data availability

This work is entirely theoretical, there is no data underpinning this publication.

References

Brizzi, F., Birrell, P. J., Plummer, M. T., Kirwan, P., Brown, A. E., Delpech, V. C., Gill, O. N., & De Angelis, D. (2019). Extending Bayesian back-calculation to estimate age and time specific HIV incidence. *Lifetime Data Analysis*, *25*(4), 757–780. https://doi.org/10.1007/s10985-019-09465-1

Brookmeyer, R., Gail, M. H., & Gail, M. H. (1994). *AIDS epidemiology: A quantitative approach*. Oxford University Press on Demand.

Corbella, A., Presanis, A. M., Birrell, P. J., & De Angelis, D. (2022). 'Inferring epidemics from multiple dependent data via pseudo-marginal methods', arXiv, arXiv:2204.08901, preprint: not peer reviewed.

Engebretsen, S., Palomares, A. D., Rø, G., Kristoffersen, A. B., Lindstrøm, J. C., Engø-Monsen, K., Chan, L. Y., Dale, Ø., Midtbø, J. E., Stenerud, K. L., & Di Ruscio, F. (2021). 'Regional probabilistic situational awareness and forecasting of COVID-19', medRxiv 21265166, preprint: not peer reviewed, https://doi.org/10.1101/2021.10.25.21265166

Marschner, I. C. (1994). Using time of first positive HIV test and other auxiliary data in back-projection of AIDS incidence. *Statistics in Medicine*, *13*(19–20), 1959–1974. https://doi.org/10.1002/sim.4780131908
Mezzetti, M., & Robertson, C. (1999). A hierarchical Bayesian approach to age-specific back-calculation of cancer incidence rates. *Statistics in Medicine*, *18*(8), 919–933. https://doi.org/10.1002/(SICI)1097-0258(19990430)18:8<919::AID-SIM89>3.0.CO;2-7
Sweeting, M. J., De Angelis, D., & Aalen, O. O. (2005). Bayesian back-calculation using a multi-state model with application to HIV. *Statistics in Medicine*, *24*(24), 3991–4007. https://doi.org/10.1002/sim.2432

https://doi.org/10.1093/jrsssa/qnad049
Advance access publication 12 April 2023

Heejong Bong, Valerie Ventura and Larry Wasserman's contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Heejong Bong, Valérie Ventura and Larry Wasserman

Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Address for correspondence: Larry Wasserman, Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: larry@stat.cmu.edu

The authors discuss hierarchical Hawkes process models for assessing the effect of interventions on epidemics. We congratulate the authors on this interesting work. Below, we raise five points, which we elaborate on in Bong et al. (2022a, 2022b).

1. The authors focus on a Bayesian approach, which is useful for including prior information but which does have some disadvantages: it requires specifying priors and does not yield confidence intervals with valid frequentist coverage and can obscure identifiability issues. The model, in its most general form, is not identified, as there are potentially more parameters than data points. The dangers of non-identifiability in epidemic models are discussed in Gallo et al. (2022). We suggest using identified models fitted by maximum likelihood, and using the sandwich method to obtain robust standard errors. This yields valid asymptotic confidence intervals and does not require priors. Non-asymptotic, exact, confidence intervals can also be obtained by inverting tests or using universal inference (Wasserman et al., 2020). Prior information can be included by inverting an integrated likelihood ratio test.
2. What is the effect of the intervention on the outcome Y? This is a causal question which needs to be answered using the tools of causal inference. Let $Y_t(\overline{a}_t)$ denote the value that Y_t would take if the intervention variables $(A_1, \ldots, A_t)$ were set to $\overline{a}_t = (a_1, \ldots, a_t)$. The mean of this counterfactual, $\mathbb{E}[Y_t(\overline{a}_t)]$, is given by Robins' (1986) g-formula. We give an explicit (rather complicated) expression $g(\overline{a}_t)$ for the g-formula in the authors' semi-mechanistic model in Bong et al. (2022b).
3. Marginal structural models (MSMs) (Robins, 2000) offer an alternative approach to assess the impact of interventions. They are semiparametric models that specify directly the treatment effect $\mathbb{E}[Y_t(\overline{a}_t)]$ and do not require modelling the epidemic process. Bonvini et al. (2022a, 2022b) combined a model similar to the author' model with the MSM approach by applying the g-formula to the model to get $g(\overline{a}_t; \theta) = \mathbb{E}[Y_t(\overline{a}_t)]$ as mentioned above and then using this as a MSM. This method is simple, robust to model misspecification, provides straightforward inference, and avoids phantom bias (Bates et al., 2021; Robins, 1986; Robins & Wasserman, 1997) which is a subtle problem that occurs when we answer causal questions for time series using parametric models.

4. To model several geographic regions simultaneously, the authors propose using a hierarchical model, which is reasonable. An alternative is to get parameter estimates separately for each region and apply shrinkage methods to the resulting estimators. This can be done nonparametrically and allows one to combine information without invoking more modelling assumptions.
5. The authors correctly point out that mobility is a mediator and, as such, we can decompose the causal effect of an intervention into direct and indirect effects. It should be noted that there is debate about how to define mediation effects and there numerous definitions. In Bong et al. (2022b), we find explicit expressions for these various mediation parameters in the semi-mechanistic model.

Conflict of interest: None declared.

References

Bates S., Kennedy E., Tibshirani R., Ventura V., & Wasserman L. (2021). 'Causal inference with orthogonalized regression', arXiv, arXiv:2201.13451, preprint: not peer reviewed.

Bong H., Ventura V., & Wasserman L. (2022a). Causal inference for some epidemic models.

Bong H., Ventura V., & Wasserman L. (2022b). Frequentist inference for semi-mechanistic epidemic models with interventions.

Bonvini M., Kennedy E. H., Ventura V., & Wasserman L. (2022a). Causal inference for the effect of mobility on Covid-19 deaths. *The Annals of Applied Statistics*, *16*(4), 2458–2480. https://doi.org/10.1214/22-AOAS1599

Bonvini M., Kennedy E. H., Ventura V., & Wasserman L. (2022b). Sensitivity analysis for marginal structural models.

Gallo L., Frasco M., Latora V., & Russo G. (2022). Lack of practical identifiability may hamper predictions in Covid-19 epidemic models. *Science Advances*, *8*(3), 1–13. https://doi.org/10.1126/sciadv.abg5234

Robins J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modelling*, *7*(9–12), 1393–1512. https://doi.org/10.1016/0270-0255(86)90088-6

Robins J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials* (pp. 95–133). Springer.

Robins J. M., & Wasserman L. (1997). Estimation of effects of sequential treatments by reparam-eterizing directed acyclic graphs. In *Proceedings of the thirteenth conference on uncertainty in artificial intelligence, UAI'97* (pp. 409–420). Morgan Kaufmann Publishers Inc.

Wasserman L., Ramdas A., & Balakrishnan S. (2020). Universal inference. *Proceedings of the National Academy of Sciences of the Unites States of America*, *117*(29), 16880–16890. https://doi.org/10.1073/pnas.1922664117

https://doi.org/10.1093/jrsssa/qnad054
Advance access publication 5 April 2023

Sawitree Boonpatcharanon, Jane Heffernan and Hanna Jankowski's contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Sawitree Boonpatcharanon[1], Jane Heffernan[2] and Hanna Jankowski[2]

[1]Chulalongkorn University, Bangkok, Thailand
[2]York University and Canadian Centre for Disease Modelling, Toronto, Canada

Address for correspondence: Hanna Jankowski, York University and Canadian Centre for Disease Modelling, Toronto, Canada. Email: hkj@yorku.ca

We congratulate the authors on a timely, important, and well-written paper. The effective reproductive ratio is a key characteristic that can be used during a disease outbreak to understand the spread of the disease and to gauge effectiveness of public health measures. Furthermore, real-time estimation of the effective reproductive ratio is of particular importance, so that real-time responses can be made by public health authorities. The COVID-19 pandemic has brought various measures, especially non-pharmaceutical ones, to the forefront of the public's attention. However, public health measures have always been of great importance in monitoring and managing disease outbreaks. As zoonotic or re-emerging disease outbreaks are expected to happen from time to time, we can expect this work to have long-term impact.

The authors propose a version of a compartmental SEIR (Susceptible-Exposed-Infectious-Recovered) disease model, developed in Engebretsen et al. (2021). Their model uses both the number of cases reported based on PCR (Polymerase Chain Reaction) testing (which assumes that only a proportion of cases have been tested) as well as hospitalisation data. Certain model parameters are held static throughout, either assumed or estimated from other sources. These parameters include π_0 and π_1 which determine the proportion of cases that submit to PCR testing; as well as the compartmental model parameters θ which determine underlying disease dynamics. The developed method, notably, allows also to estimate the actual vs. reported number of cases; another key characteristic of interest.

One difficulty in public health management of COVID-19 has been that various inputs, such as θ and π_0, π_1, have changed considerably over the two years of the pandemic. For example, in the province of Ontario, Canada, PCR testing was very low at the beginning of the pandemic (first half of 2020) and also starting at the height of the Omicron outbreak (early 2022 until time of writing). Indeed, starting in 2022, only select individuals from target groups (e.g., those that are severely immunocompromised) are eligible for PCR testing. The model proposed by the authors allows for considerable flexibility in modelling such scenarios, however, certain parameter choices could affect performance and accuracy of the method. Have the authors tested their methods where the proportion of the population receiving PCR testing is quite small? When PCR testing is so low, it would be particularly relevant to obtain estimates of actual vs. reported cases.

Conflict of interest: None declared.

Reference

Engebretsen S., Ro G., Diz-Lois Palomares A., Engo-Monsen K., Kristoffersen A. B., Di Ruscio F., Frigessi A., & de Blasio B. F. (2021). 'Spatial modelling of the early-phase of the COVID-19 epidemic in Norway', medRxiv, Cold Spring Harbor Laboratory Press. https://www.medrxiv.org/content/10.1101/2021.10.25.21265166v1

https://doi.org/10.1093/jrsssa/qnad050
Advance access publication 6 April 2023

Hans R Künsch and Fabio Sigrist's contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Hans R. Künsch[1] and Fabio Sigrist[2]

[1]Department of Mathematics, ETH Zurich, Zurich, Switzerland
[2]Institute of Financial Service, Lucerne University of Applied Sciences and Arts, Lucerne, Switzerland

Address for correspondence: Hans R. Künsch, Department of Mathematics, ETH Zurich, CH-8092 Zurich, Switzerland. Email: kuensch@stat.math.ethz.ch

The authors are to be congratulated for an interesting and timely paper that combines epidemiological modelling with advanced statistical methodology. We are currently working on a related project, developing an Markov chain Monte Carlo algorithm for estimating effective daily reproduction numbers using the model of Cori et al. (2013) (see also p. 3 in the paper under discussion). The standard procedure as described, e.g., in Huisman et al. (2022) estimates first daily infections from confirmed cases by deconvolution and then reproduction numbers from inferred infections whereas our algorithm estimates infections and reproduction numbers jointly based on a coherent statistical model. In our current implementation, the acceptance rate for the number of infections decreases to very low values for long observation periods. To avoid this, we split the data into overlapping smaller windows and merge the posterior samples from different windows. We have also developed an sequential Monte Carlo algorithm for the same model, but even with a more efficient proposal than the bootstrap, sample depletion is a problem. A new observation of confirmed cases has a substantial effect on the smoothing distribution of infections several days earlier, and the auxiliary particle filter can account for this only via resampling, and not via changing the values in the sample. Since our setup differs from the one in the paper not only in the model and the data sources but also in other specifications a direct comparison is not possible.

The authors assume that the detection probability of an infected case on some day depends only on the number of tests on that day; see (5) in their paper. We think that the severity of symptoms would be a more important covariate as individuals with no or only mild symptoms are much more likely not to take a test.

In epidemic modelling, one has to make many choices: A more complex model is potentially more realistic, but it suffers from a large number of unknown parameters and/or additional assumptions. In view of the many choices that have to be made and the obvious weak points of all assumptions, an honest communication of the uncertainty is challenging. A potential advantage of the model by Cori et al. (2013) is that it is simpler and relies on fewer assumptions compared to an SEIR (susceptible-exposed-infected-recovered) model. In general, we think that sensitivity studies and comparison of results obtained with methods from different groups are important to assess uncertainty.

Conflict of interest: None declared.

References

Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and soft-ware to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, *178*(9), 1505–1512.

Huisman, J. S., Scire, J., Angst, D. C., Li, J., Neher, R. A., Maathuis, M. H., Bonhoeffer, S., & Stadler, T. (2022). Estimation and worldwide monitoring of the effective reproductive number of SARS-CoV-2. *eLife, 11*: e71345. https://doi.org/10.7554/eLife.71345

https://doi.org/10.1093/jrsssa/qnad051
Advance access publication 6 April 2023

Andrew B Lawson's contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Andrew B. Lawson[1,2]

[1]Public Health Sciences, College of Medicine, Medical University of South Carolina, Charleston, SC, USA
[2]Usher Institute, University of Edinburgh, Edinburgh, UK

Address for correspondence: Andrew B. Lawson, Medical University of South Carolina. Email: lawsonab@musc.edu

I commend the authors on a very thoughtful paper related to modelling COVID-19. I have a few comments and questions about the paper that I think is important to raise.

First of all, the paper does not provide any examples of the model fitting to real data. This is surprising given the paper is destined for JRSSA. It would have been much more persuasive if examples of the time series of counts and mobility are provided and the resulting model fits displayed. It's clear from our own modelling that goodness of fit (GOF) can vary over time during the pandemic and often models achieving good relative GOF (e.g., overall Watanabe Akaike Information Criterion) do not fit well locally (either at peak times or in given regions) (Lawson & Kim, 2021, 2022).

Second, the Flaxman et al. paper from 2020 (2020b) is cited in a number of places as giving detail about modelling. However, that paper is focused on mortality and non-pharmaceutical interventions (NPIs) and not on cases. Hence the models must in some way be different. No detail is given about how they differ. Is a Poisson data model or an negative binomial (NB) model used? In the paper it is stated (Section 5):

> '*This could be a Poisson distribution, where there is no auxiliary parameter. Using a quasi-Poisson or negative binomial instead allows for overdispersion*'.

This statement is not relevant within a Bayesian modelling framework as overdispersion is easily accommodated within Poisson data models via higher level random terms (uncorrelated or correlated heterogeneity).

Mobility and NPIs: It is unfortunate that in the paper no mention is made of the source of mobility data or indeed how the stated averaging of the mobility data is performed. Is it Google daily data (e.g., https://www.google.com/covid19/mobility/) and if so, which of the six indices was used in Figure 1. I assume the correlation result achieved (85%) would highly depend on the smoothing applied to the mobility data.

Confounding: although the mention of confounding is made there appears to be little emphasis on looking for residual confounding. This also relates to spatial referencing in the paper. Regional effects are mentioned in Section 6, but no specific modelling of space-time is discussed. Spatial effects are important in public health (PH) intervention planning, as well as in the assessment of residual confounding [see Lee et al. (2016) for general discussion]. A spatially referenced reproduction number was developed by Rotejanaprasert et al. (2019).

Conflicts of interest: None declared.

References

Lawson, A. B., & Kim, J. (2021). Space-time COVID-19 Bayesian SIR modeling in South Carolina. *PLos One*, *16*(3), e0242777. https://doi.org/10.1371/journal.pone.0242777

Lawson, A. B., & Kim, J. (2022). Bayesian space-time SIR modeling of Covid-19 in two US states during the 2020-2021 pandemic. *Plos One 17*(12):e0278515. https://doi.org/10.1371/journal.pone.0278515

Lee, E., Asher, J., Goldlust, S., Kraemer, J., Lawson, A. B., & Bansal, S. (2016). Mind the scales: Harnessing spatial big data for infectious disease surveillance and inference, *The Journal of Infectious Diseases*, *214*(Suppl 4), S409–S413. https://doi.org/10.1093/infdis/jiw344

Rotejanaprasert, C., Lawson, A. B., & Iamsirithaworn, S. (2019). Spatiotemporal multi-disease transmission dynamic measure for emerging diseases: An application to Dengue and Zika integrated surveillance in Thailand. *BMC Medical Research Methodology*, *19*(1):200. https://doi.org/10.1186/s12874-019-0833-6

https://doi.org/10.1093/jrsssa/qnad052
Advance access publication 5 April 2023

Jorge Mateu and Álvaro Briz-Redón's contribution to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Jorge Mateu[1] and Álvaro Briz-Redón[2]

[1]Department of Mathematics, University Jaume I, Castellón, Spain
[2]Department of Statistics and Operations Research, University of Valencia, Valencia, Spain

Address for correspondence: Jorge Mateu, Department of Mathematics, University Jaume I, E-12071 Castellón, Spain. Email: mateu@uji.es

Discussion

It is difficult to outline good contributions to the statistical modelling of COVID-19 data, but these two papers are to be congratulated on their valuable and thought-provoking contributions within this rapidly developing field. We would like to focus our attention on several aspects that are arguably of interest to the community.

Bhatt et al. consider that R_t can be estimated in each region in separate models. Clearly, this reproductive number can be better estimated if space–time structure is considered as, though a pragmatic assumption, it seems not very realistic that the reproductive number cannot vary in space and time, and can be affected by neighbouring structures. Another aspect is considering an alternative mechanistic form for the space–time intensity function (see Briz-Redón et al., 2022) bringing into play a deterministic part (where covariates of all types play a role) and an analytical form for the interaction. This seems a promising way to explore the behaviour of the spread of COVID-19.

Storviv et al. consider a stochastic susceptible-exposed-infectious-recovered (SEIR) model whose output is incorporated into their main statistical model to capture R_t variation. This approach is novel and convenient to control for the time-varying size of the susceptible population. Besides, the proposal of Storviv et al. for the time-varying reproduction number can be easily extended to space–time by considering a low-rank approximation based on splines as in Martínez-Beneito et al. (2022). Proceeding this way, there is no need for such a large amount of involved parameters and Bayesian inference is easily adopted. We argue that heterogeneity across areas or counties should be considered and not avoided or simplified. Again, mobility patterns are crucial in the spread of COVID-19 and a natural way to outline its importance comes well explained in Slater et al. (2022). We wonder if a simpler while more model-based approach would be a good alternative to such mobility modelling.

Finally, we would like to highlight two common aspects of these contributions. First, they explicitly account for certain features of COVID-19 data, mainly the existence of underreporting and the temporal delay of the observations. This is something that has been overlooked in numerous related studies. Second, the Bayesian nature of both methodologies enables estimating $P(R_t > 1)$ directly. Measuring the uncertainty around the reproduction number could better guide decision-making in future epidemic events.

Conflicts of interest: None declared.

References

Briz-Redón, A., Iftimi, A., Mateu, J., & Romero-García, C. (2022). A mechanistic spatio-temporal modeling of COVID-19 data. *Biometrical Journal*, *65*(1), e2100318. https://doi.org/10.1002/bimj.202100318

Martínez-Beneito, M. A., Mateu, J., & Botella-Rocamora, P. (2022). Spatio-temporal small area surveillance of the COVID-19 pandemics. *Spatial Statistics*, *49*, 100551. https://doi.org/10.1016/j.spasta.2021.100551

Slater, J. J., Brown, P., Rosenthal, J. S., & Mateu, J. (2022). Capturing spatial dependence of COVID-19 case counts with cellphone mobility data. *Spatial Statistics*, *49*, 100540. https://doi.org/10.1016/j.spasta.2021.100540

The authors replied later, in writing, as follows:

https://doi.org/10.1093/jrsssa/qnad053
Advance access publication 5 April 2023

Bhatt, Ferguson, Flaxman, Gandy, Mishra, and Scott's reply to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Samir Bhatt[1], Neil Ferguson[1], Seth Flaxman[2], Axel Gandy[2], Swapnil Mishra[1] and James A. Scott[2]

[1]MRC Centre for Global Infectious Disease Analysis, Imperial College, London, UK
[2]Department of Mathematics, Imperial College, London, UK

Address for correspondence: Axel Gandy, Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. Email: a.gandy@imperial.ac.uk

We would like to thank the Royal Statistical Society, and all of the discussants and editors for facilitating an important discussion and for all of their important contributions during the pandemic.

Professor Diggle gives insightful comments on problems that arise through discretisation of space, and when different spatial scales are mixed in a model, as occurs commonly for pragmatic reasons. The effect of this in epidemiological models would be of interest to explore in more detail.

As to Prof. Diggle's point of what reproduction numbers are exactly estimating; indeed the precise interpretation can differ amongst different models, they usually aim to give an average number of infections caused by an infected individual; the key difference is of course what group this is being averaged over, and if these groups become small then estimates can potentially become very noisy. To aid decision makers, reporting results at some aggregation level will probably always be necessary.

As Prof. Diggle correctly points out, whether case numbers (or other observations) are representative of the underlying epidemic is a key concern, particularly due to informed missingness. For example, limited availability of testing in the early parts of the pandemic led to unreliable case numbers. In our work, we often found that as long as the missingness is spatially and temporally stable, it is still possible to estimate reproduction numbers; however, when it comes to projecting population saturation due to acquired immunity, i.e. when the epidemic will turn, then it is indeed crucial to tie the observed counts to the underlying infections; population surveys, such as REACT seem to be a necessary (albeit imperfect) tool for this. We would also like to point out that our framework allows us to coherently integrate different observations, i.e. infections, case counts, hospitalisations, and deaths among others to provide a more reliable estimate (Pakkanen et al., 2023).

Finally, Prof. Diggle's point about discretising space is an important area of future research. While the model-based geostatistics framework pioneered by Prof. Diggle is a sensible starting point, we caution that spatial correlation is likely to be highly complex and driven by an interaction of mobility, the economy, culture, behaviour, and other aspects. As a result, simple covariance functions are unlikely to fully resolve the spatial aspect.

Professor Richardson rightly points out that ascertainment bias is important and that randomised surveillance seems to be the main way of estimating the amount of bias. As mentioned above, assuming stability in the ascertainment bias, it is still possible to estimate reproduction numbers. Combining data from different sources is indeed possible in our model; conflicting evidence may point to deficiencies in the underlying model—thus it might sometimes be beneficial to present estimates from different types of data separately; differences in these estimates could lead to important insights about the state of the epidemic and possible model improvements, from including age structure (Monod et al., 2021) to modelling variants of concern (Faria et al., 2021).

A limitation of the two-stage approach is indeed the propagation of uncertainty. A joint approach as taken in Flaxman et al. (2020) is always preferable but at a considerably larger computational cost with the possibility of harder to sample posteriors. The two-stage approach therefore represents a compromise. One of the challenges in our model is that the key object of interest is $R(t)$, for which no ground truth exists. Hence, it is not possible to directly assess residual confounding for $R(t)$. As Prof. Richardson points out, causal interpretations almost always rely on strong assumptions, and thus should be used cautiously.

We agree with Prof. Richardson that caution should be taken with shrinkage priors; a sensible choice is the horseshoe prior by Piironen and Vehtari (2017) that balances variable selection and collinearity. As to the source of the mobility data, we used the Community Mobility Report provided by Google (https://www.google.com/covid19/mobility/).

Dr Chind points out the value of research's reproducibility, which is an important issue. Thankfully, it is becoming standard that code is being provided with papers, and we always strive to make code available in our own research. The policies of the journals of the RSS do indeed strongly encourage this. As pointed out, successfully executing the code provided with articles can sometimes be difficult; one key problem is that the programming languages and packages used in the code evolve. One promising approach for long-term reproducibility is to provide code in containers that fix the environment in which the code runs (see e.g. Clyburne-Sherin et al., 2019).

Professor Lawson raises several points. Indeed, models can have varying degree of fit, particularly around peaks and early on in the pandemic. This can be problematic, as estimates can be heavily influenced by factors not accounted for in models. This is why models need to have sufficient flexibility to adapt to the observed data to provide fits throughout an epidemic. One example of this is that observed counts are generally overdispersed compared to Poisson counts and this is why we generally work with negative binomial observation models. As pointed out, a similar effect could be achieved in a Bayesian setting by including a latent variable for each observation; due to computational efficiencies we did not choose to do this.

Professor Mateu and Dr Briz-Red raise a series of very important points. We again stress a critical area of research is into the spatial aspect of spread, both in terms of fine grained data (such as mobility) but also new analytical models with a mechanism for the spatial spread of contagion.

The contribution by Dr Bong, Prof. Ventura, and Prof. Wassermann raises very interesting and valid points. The issue of identifiability is very important; it might indeed be easier to discuss it in a frequentist setting but it would be present in both frequentist and Bayesian models. Even very simple models, with few parameters, lead to identifiability issues, e.g. a model based on cases cannot distinguish between fast initial growth of an epidemic due to high R_0 vs. many imported infections. In our experience, the inclusion of epidemiologically informed priors in a Bayesian framework is the most effective way to ensure identifiability.

We agree that causal questions could be answered using different tools, all of which have their own assumptions, advantages, and drawbacks. We look forward to reading their derivation of the g-formula and Marginal Structural Models for our model. A formal consideration of mediation in our model is important, and could go beyond mobility to consider, for example, social factors such as attitudes and risk perceptions. We hope their causal approaches will generate future avenues of research, and be put to the test in real settings.

The complexities of mechanistic models of infectious disease spread, implemented within a statistical framework, requires a careful assessment of the mechanism, robust inference, attention to causality and confounding, and an accounting of identifiability. Pandemic response leaves little time to fully explore the interplay among these factors and their impact on estimates. We hope

that statistical research undertaken while we await the next pandemic can be used to refine analytic methods and decisively map the data sources, analytical approaches, diagnostics, and questions so as to begin resolving the issues discussed here.

Conflict of interest: none declared.

References

Clyburne-Sherin A., Fei X., & Green S. A. (2019). Computational reproducibility via containers in psychology. *Meta-psychology*, *3*. https://doi.org/10.15626/MP.2018.892

Faria N. R., Mellan T. A., Whittaker C., Claro I. M., Candido D. D. S., Mishra S., Crispim M. A., Sales F. C., Hawryluk I., McCrone J. T., & Hulswit R. J. (2021). Genomics and epidemiology of the P. 1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*, *372*(6544), 815–821. https://doi.org/10.1126/science.abh2644

Flaxman S., Mishra S., Gandy A., Unwin H. J. T., Mellan T. A., Coupland H., Whittaker C., Zhu H., Berah T., Eaton J. W., & Monod M. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Monod M., Blenkinsop A., Xi X., Hebert D., Bershan S., Tietze S., Baguelin M., Bradley V. C., Chen Y., Coupland H., Filippi S., Ish-Horowicz J., McManus M., Mellan T., Gandy A., Hutchinson M., Unwin H. J. T., van Elsland S. L., Vollmer M. A. C., …Ratmann O. (2021). Age groups that sustain resurging COVID-19 epidemics in the United States. *Science*, *371*(6536), eabe83. https://doi.org/10.1126/science.abe8372

Pakkanen M. S., Miscouridou X., Penn M. J., Whittaker C., Berah T., Mishra S., Mellan T. A., & Bhatt S. (2023). Unifying incidence and prevalence under a time-varying general branching process. *Journal of Mathematical Biology (accepted)*. https://arxiv.org/abs/2107.05579

Piironen J., & Vehtari A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, *11*(2), 5018–5051. https://doi.org/10.1214/17-EJS1337SI

https://doi.org/10.1093/jrsssa/qnad055
Advance access publication 15 May 2023

Storvik, Palomares, Engebretsen, Rø, Engø-Monsen, Kristoffersen, de Blasio and Frigessi's reply to the Discussion of 'The Second Discussion Meeting on Statistical aspects of the Covid-19 Pandemic'

Geir Storvik[1,2], Alfonso Diz-Lois Palomares[3], Solveig Engebretsen[2], Gunnar Øyvind Isaksson Rø[3], Kenth Engø-Monsen[4], Anja Bråthen Kristoffersen[3], Birgitte Freiesleben de Blasio[3,5] and Arnoldo Frigessi[5,6]

[1]Department of Mathematics, University of Oslo, Oslo, Norway
[2]Norwegian Computing Center, Oslo, Norway
[3]Department of Method Development and Analytics, Norwegian Institute of Public Health, Oslo, Norway
[4]Telenor Research, Fornebu, Norway
[5]Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway
[6]Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

Address for correspondence: Arnoldo Frigessi, PB 1122 Blindern, N-0317 Oslo, Norway. Email: arnoldo.frigessi@medisin.uio.no

We are grateful to the Research section of the Royal Statistical Society for allowing us to present our article and for having organised this discussion. We highly appreciate the valuable and interesting comments, criticisms, and questions from all the discussants.

We wrote and submitted a first version of our article at the end of 2020, when there was limited knowledge about the disease and on how the Covid-19 pandemic would progress, while health care systems across the world were under immense pressure. In Norway at that time, our models delivered weekly estimates of reproduction numbers and short-term forecasts of hospital occupancy, for each region of the country. The Norwegian health authorities used our results actively in their decisions. Today, almost two and a half years later, the SARS-CoV-2 virus is present in all over the world, vaccines and natural immunity after infection give good protection, and the health systems are in general not under special pressure because of the disease. Multiple variants of the virus are circulating, and the world health authorities are trying to monitor their capacity to escape immunity and the threat posed on the health of various groups of the population. Our models are still run regularly, now informed by much less precise data, and give useful indications of the current situation. Most of all, mathematical models and inferential methods are now evaluated and sharpened, to be readily actionable when the next pandemic will occur. In this light, the present discussion indicates directions of research to be better equipped for the future.

Data and models

Richardson emphasised the importance of a good calibration of uncertainty and appreciated our comparison with the simpler approach EpiEstim and our prediction-based validation. As commented by several discussants, we under-quantify uncertainty because several hyperparameters were pre-estimated from other data sources and plugged-in into the SEIR-Sequential Monte Carlo (SMC) model, though often by sampling them repeatedly from their estimated uncertainty distribution. Some hyperparameters, in particular those involved in the R_t-model, were estimated together with the latent processes. In general, the uncertainty of predictions should be estimated more realistically, by including uncertainties of all hyperparameters, assuming informative priors whenever possible, though taking into account the computational time allowed in practice.

Diggle suggested to model the epidemic as a spatio-temporal point process, with geo-localised events in time. We could not try this approach, because our data were aggregated in space (at municipality level) and time (day), leading to our model. In the present paper, we used national daily data, with the aim to estimate a national daily reproduction number. While operative, our model was informed by daily data at the county level, estimating reproduction numbers county-wise. There we assumed an equation of the type (1a) for every county, with its own county-specific reproduction number, and county-specific likelihood similar to Eq. (1c), while assuming conditional independence between counties. We used Sequential Monte Carlo (SMC) also in this case, although this was computationally much more demanding. In particular, we discovered weight degeneracy due to conflict between data sources. To remedy, we used a power likelihood (Miller & Dunson, 2018) to scale down the influence of data, thus helping resolving the discord and improving convergence. Miller and Dunson (2018) argue that power likelihood is a robustification procedure with respect to model assumptions.

We agree with Diggle and Mateu and Briz-Redón that it is important to account for spatial heterogeneity and that even county levels could be too coarse. Large geographical units, like Norwegian counties, include both densely populated urban areas and rural and semi-abandoned areas. Furthermore, many local interventions were implemented at the municipality level as a response to local outbreaks (and there are 356 municipalities in Norway). In order to produce prediction at a local scale, it would have been possible to work at the municipality level, however at a very significant computational cost. In addition, we observed that county-wise predictions were much less stable in periods with low infection rates and near-zero new daily hospitalisations, something that would occur much more frequently in models at municipal level. We also noticed that privacy concern might prevent the availability of data at more precise spatial and temporal scales, in particular hospitalisation data.

We liked the suggestion of Mateu and Briz-Redón of a small area model taking spatio-temporal dependencies more directly into account through specification of appropriate precision matrices that can handle both spatial heterogeneity and human mobility patterns, in addition to temporal

dynamics. Ghosh proposed the use of stochastic differential equations (SDEs) as an alternative to compartmental models. This is certainly an interesting suggestion to explore, given the potential improvement in terms of computational cost. Because of the detailed space-time resolution of our mobility data, we opted for a more structured process, implementing a certain level of spreading, realised by returning individuals visiting a municipality A back to their municipality of origin B, after a period of mixing in A. Furthermore, contrary to SDEs, compartment models introduce mechanisms of infection which allow to estimate latent quantities, as underlined by Richardson. We agree with Richardson on the usefulness of many different models, so that results can be compared, strengthening trust in aggregated and confirmed predictions.

As brought up by several discussants, multiple sources of data can offer more information about parameters and latent variables but also pose challenges on how to estimate their precision and with this their weight in the likelihood, and on how to cope with conflicting information. As demonstrated in Li et al. (2023), combining different data sources is in principle easy within state-space models such as the ones we have considered. Richardson correctly underlined that misspecification of delay distributions might lead to conflicting information on the state of the epidemic. In order to reduce this risk, we regularly re-estimated the distributions of the delay between infection and positive test. A further and even more serious possible cause of mis-alignment of data sources occurred when the probability of hospitalisation after infection, or (to a smaller extent) the probability of being tested once infected, were wrongly estimated. This issue was also raised by Boonpatcharanon et al. We protected our analyses against such mistakes by re-estimating these probabilities regularly. In particular when new viral strains were becoming more frequent and predominant, we gradually updated the probability of hospitalisation, using results from early studies in other countries. We similarly gradually decreased the hospitalisation risks for vaccinated individuals, when vaccination or boosting became available. There are several parameters that describe the hospitalisation process in time. In Norway, we had access to the national registry with all the individual-level data for patients hospitalised with Covid-19. Using these, we estimated: the duration of hospital stay both in ICU and not; the time between the onset of symptoms and hospitalisation; the probability of ICU treatment given hospitalisation; among others. When we suspected a change in one or more of these parameters, we assumed a changepoint, estimating newer values only based on the more recent data, though with a trade-off between capturing recent changes and the availability of enough data points.

Our compartment model did not include age-structured compartments, in order to reduce the number of latent variables, well knowing that the propensity of infection and the risk of hospitalisation were clearly age-dependent. At the beginning of the pandemic, when test data were highly biased, we distributed the simulated disease incidence according to the demographic age profile in each county. During this period, we also only calibrated the model to hospital data and not the test data. When testing became generally available for the whole population, we started to use both data sources in the calibration. We soon started to observe inconsistencies between our models for hospitalisation and positive test data. These were mainly due to the hospitalisation risks changing over time, because of a shift in the age profile of the infected and tested, towards younger ages. We then distributed the simulated disease incidence according to the region-specific age profile of the infected, which we in turn estimated from positive test data instead than from demography. In general, several hyperparameters were calibrated using both external sources and the same hospital and test data. This approach does result in a double use of some data sources, which can lead to underestimation of the uncertainty. Certainly, a more principled and systematic approach would be desirable here.

Underestimation of case counts was pointed out by Diggle and by Mateu and Briz-Redon. We took into account the efforts made to detect infected individuals by modelling the probability of being tested if infected as a function of the total number of tests. This probability was estimated to be typically around 0.5 during the whole time period, confirming a large bias in published case counts. Corbella et al. and Künch and Sigrist suggested that the severity of symptoms should be included as covariate in a model of the probability to detect a positive case. Indeed, it would be interesting to test a model that uses the current estimate of the number of infected with symptoms or other surrogates of severity of the circulating strains, despite the consequent increase in the number of parameters to be estimated. Boonpatcharanon et al. questioned our assumption that parameters in the model of the probability of detection of cases did not change over time except through the number of

tests performed. We agree on this, and our use of beta-binomial distributions for the observations partly accounted for this. Furthermore, hospitalisation probabilities varied on a daily basis, derived from external data registries. In addition, we did not use test data when these were critically unreliable and incomplete, like before August 2020 (when the testing system was under-dimensioned) and after February 2022, when testing when symptomatic was not required anymore. In the period for which we used test data, the Norwegian testing infrastructure essentially did not reach saturation. A randomised prevalence study similar to the REACT study in England would certainly have been very useful also in Norway, for example as mentioned by Diggle to debiase testing data; unfortunately, we failed to convince the authorities to prioritise it.

Corbella et al. discussed the importance of taking into account dependence between data sets. Our assumption was that, conditional on the latent variables of all compartments, the test data and the hospitalisation data were independent. It is however true that every hospitalised case also has a positive test, though we have not implemented this directly in our forward simulations. The difference will be minor, as the number of daily new hospitalisations was very small compared to the number of positive cases.

Conflicts in data will typically result in large variabilities in the importance weights used in the SMC algorithm. This can be seen both as a weakness and a strength of the procedure. A weakness is that the performance of the algorithm will deteriorate, a strength in that extreme weights can give a warning that a conflict might be present. During operations, we explored the performance of the algorithm, and in several cases, we found errors in data because of extreme weights. More generally, conflicts between the two sources of data were frequent, because our models lacked the flexibility to account for heterogeneity occurring in time and space. To allow for this, several parameters would need to be indexed in time and space, making the inferential task likely unfeasible, due to the lack of data.

Computational aspects

Richardson correctly pointed out the extra challenges related to our choice of stochastic delays which, combined with possible long delays, led to many extra latent variables and computational challenges. She suggested reusing the same proposal for a number of iterations, essentially delaying resampling. We agree that this is an interesting idea to follow. Another (perhaps supplementary) approach is to reduce the number of latent variables by an approximate calculation of the likelihoods in Eq. (1c) without including the dummy variables z_t. Furthermore, our rule of mobility giving priority to commuting, makes the number of compartmental variables quadratic in the number of regions. Removing this rule would reduce the numbers to be linear in the number of regions, but at the cost of possible overestimation of the spread (Danon et al., 2009). We should also note that although the dimension of the latent process is high, there are very strong dependencies between many of the variables making the effective dimension significantly smaller.

As pointed out by Birrell et al., the bootstrap filter suffers from sample degeneracy when considering smoothed distributions $p(R_t, z_t|\boldsymbol{y}_{1:T})$ for $t \ll T$. This was the main reason for using fixed lag-smoothing distributions $p(R_t, z_t|\boldsymbol{y}_{1:\min\{t+24,T\}})$ instead, which suffer less problems of degeneracy. Due to the delays from infection to test/hospitalisation, some loss of efficiency is to be expected here. However, we did make some comparisons of the fixed lag-smoothing with the forward-filtering backwards smoothing algorithm (Doucet et al., 2000; Godsill et al., 2004), indicating that a lag of 24 days was sufficient. Birrell et al. also pointed out that 'shocks' in the process can lead to weight degeneracy. We saw this in particularly for the piece-wise model in August 2020 when very few particles were given significant weights (lower plot of Figure C.2) but also the change from the first static period to the drop in March 2020 was a computational challenge. Adaptive algorithms which either increase the number of particles when needed or consider more advance proposals should probably be adopted in such cases.

There has been several recent developments of SMC algorithms for high-dimensional state-space models, e.g. tempering (Del Moral et al., 2006), block particle filters (Doucet et al., 2006), the space-time particle filter (Beskos et al., 2017), nested SMC (Naesseth et al., 2019), and divide-and-conquer SMC (Lindsten et al., 2017). Such approaches would certainly improve on the efficiency but would require working in the code for the compartmental model, something we wanted to avoid in our setting, for practical reasons and also in order to control the comparison of results with other

models. However, if SMC methods should be more used more systematically for the analysis of spread of infectious diseases, such improvements should certainty be considered.

Estimation of static parameters is in particular a challenging task for SMC algorithms. In our case, we performed a pragmatic choice in that most parameters were fixed while a few (parameters within the $\{R_t\}$ process) were estimated through the sufficient statistics approach (Fearnhead, 2002; Storvik, 2002). As pointed out by Corbella et al., there is a risk in underestimating the uncertainty of R_t and of other latent variables. Assuming appropriate priors for all parameters, updated parameter information can in principle be obtained by sampling the parameters at Day 1 from the prior and propagate these through the dynamics within the SMC algorithm. However, due to the need of resampling within SMC, the set of unique parameter values will quickly degenerate. The trick of including artificial dynamics into the static parameters, as advocated by Liu and West (2001), is a computationally efficient approach and might be justified from a modelling point of view in that many of the parameters assumed to be static actually will change over the time period. More sophisticated algorithm such as the resample-move algorithm (Berzuini & Gilks, 2001) or, as suggested by Birrell et al., the SMC^2 algorithm (Chopin et al., 2013) will certainly improve results. For such approaches to work in practice, more efficient particle algorithms as discussed above will be necessary due to the extra computational requirements needed.

Künch and Sigrist discussed their own Markov Chain Monte Carlo (MCMC) and SMC implementations but mention challenges. In our setting, beta-binomial distributions instead of pure binomial distributions and stochastic delay distributions gave more flexibility in the likelihoods which we believe was important in order to obtain algorithms that worked satisfactorily.

Other issues

Diggle questioned whether R_t really varies that quickly. One reason for such rapid changes was the frequent and intense local outbreaks, which were often rapidly controlled by successful contact tracing and isolation. We agree with Diggle, that such daily brisk changes challenge the use of R_t-estimates in decision-making. In our weekly public reports, we actually presented weekly averages of R_t-estimates, which we believe were more useful for public health decision-making. In this scale, the estimated oscillations were later validated, as discussed in the *Results* section, by relating them to different events that occurred during the pandemic.

Künch and Sigrist discussed the need of sensitivity studies of the many assumptions. Although Richardson appreciated the sensitivity analysis that we *have* done, we certainly agree that more sensitivity studies would be useful. Bayesian model checking methods such as posterior predictive approaches (Gelman et al., 1996) can be useful though they can be very conservative, in particular for hierarchical models (Hjort et al., 2006; Steinbakk & Storvik, 2009), due to double use of data. Although Figure 4 in the paper is reassuring in that predictions fit data well, these figures can be seen as some kind of posterior predictive check and suffer from the weaknesses mentioned above. The predictions after the green dashed vertical line in Figure 4 are pure forecasts and allow a more reliable validation.

Several papers have recently been published on the analysis of Covid-19 data, which share many similarities with our work. Koyama et al. (2021) used SMC for estimation of time-varying reproduction numbers based on Hawkes processes for modelling delays. Peak et al. (2020) used SMC methods for estimating the comparative efficacy of individual quarantine and active monitoring on SARS-CoV-2. Jiang et al. (2021) considered the use of waste water, with inference based on compartmental models and SMC algorithms. Li et al. (2023) considered estimation of time-varying reproduction numbers (among many other quantities) using particle filters with a much larger compartmental model and many additional sources of data. All these papers, combined with our own paper, demonstrate the utility of state-space model formulations combined with SMC algorithms for online inference of infectious diseases. However, as pointed out by our discussants, more research is needed in order to utilise the many data sources available in a coherent way, and to make the computational algorithms both efficient and easy to use. Preparing these tools for the next pandemic must be our common goal.

Conflict of interest: None declared.

References

Berzuini C., & Gilks W. (2001). Resample-move filtering with cross-model jumps. In A. Doucet, N. de Freitas, N. Gordon (Eds), *Sequential Monte Carlo methods in practice. Statistics for engineering and information science*. New York, NY: Springer. https://doi.org/10.1007/978-1-4757-3437-9_6

Beskos A., Crisan D., Jasra A., Kamatani K., & Zhou Y. (2017). A stable particle filter for a class of high-dimensional state-space models. *Advances in Applied Probability*, *49*(1), 24–48. https://doi.org/10.1017/apr.2016.77

Chopin N., Jacob P. E., & Papaspiliopoulos O. (2013). SMC2: An efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(3), 397–426. https://doi.org/10.1111/j.1467-9868.2012.01046.x

Danon L., House T., & Keeling M. J. (2009). The role of routine versus random movements on the spread of disease in Great Britain. *Epidemics*, *1*(4), 250–258. https://doi.org/10.1016/j.epidem.2009.11.002

Del Moral P., Doucet A., & Jasra A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(3), 411–436. https://doi.org/10.1111/j.1467-9868.2006.00553.x

Doucet A., Briers M., & Sénécal S. (2006). Efficient block sampling strategies for sequential Monte Carlo methods. *Journal of Computational and Graphical Statistics*, *15*(3), 693–711. https://doi.org/10.1198/106186006X142744

Doucet A., Godsill S., & Andrieu C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, *10*(3), 197–208. https://doi.org/10.1023/A:1008935410038

Fearnhead P. (2002). Markov chain Monte Carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics*, *11*(4), 848–862. https://doi.org/10.1198/106186002835

Gelman A., Meng X.-L., & Stern H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*(4), 733–760.

Godsill S. J., Doucet A., & West M. (2004). Monte Carlo smoothing for nonlinear time series. *Journal of the American Statistical Association*, *99*(465), 156–168. https://doi.org/10.1198/016214504000000151

Hjort N. L., Dahl F. A., & Steinbakk G. H. (2006). Post-processing posterior predictive *p* values. *Journal of the American Statistical Association*, *101*(475), 1157–1174. https://doi.org/10.1198/016214505000001393

Jiang S., Maggard K., Shakeri H., & Porter M. D. (2021). An application of the partially observed Markov process in the analysis of transmission dynamics of Covid-19 via wastewater. In *2021 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 1–6). IEEE.

Koyama S., Horie T., & Shinomoto S. (2021). Estimating the time-varying reproduction number of COVID-19 with a state-space method. *PLoS Computational Biology*, *17*(1), e1008679. https://doi.org/10.1371/journal.pcbi.1008679

Li X., Patel V., Duan L., Mikuliak J., Basran J., & Osgood N. (2023). Real-time epidemiology and acute care need monitoring and forecasting for covid-19 via Bayesian sequential Monte Carlo-leveraged transmission models. *Preprints.org* 2023, 2023020078. https://doi.org/10.20944/preprints202302.0078.v1

Lindsten F., Johansen A. M., Naesseth C. A., Kirkpatrick B., Schön T. B., Aston J. A. D., & Bouchard-Côté A. (2017). Divide-and-conquer with sequential Monte Carlo. *Journal of Computational and Graphical Statistics*, *26*(2), 445–458. https://doi.org/10.1080/10618600.2016.1237363

Liu J., & West M. (2001). Combined parameter and state estimation in simulation-based filtering. In A. Doucet, N. de Freitas, N. Gordon (Eds), *Sequential Monte Carlo methods in practice. Statistics for engineering and information science* (pp. 197–223). New York, NY: Springer. https://doi.org/10.1007/978-1-4757-3437-9_6

Miller J. W., & Dunson D. B. (2018). Robust Bayesian inference via coarsening. *Journal of the American Statistical Association 114*(527), 1113–1125. doi: 10.1080/01621459.2018.1469995

Naesseth C. A., Lindsten F., & Schön T. B. (2019). High-dimensional filtering using nested sequential Monte Carlo. *IEEE Transactions on Signal Processing*, *67*(16), 4177–4188. https://doi.org/10.1109/TSP.2019.2926035

Peak C. M., Kahn R., Grad Y. H., Childs L. M., Li R., Lipsitch M., & Buckee C. O. (2020). Individual quarantine versus active monitoring of contacts for the mitigation of COVID-19: A modelling study. *The Lancet Infectious Diseases*, *20*(9), 1025–1033. https://doi.org/10.1016/S1473-3099(20)30361-3

Steinbakk G. H., & Storvik G. O. (2009). Posterior predictive *p*-values in Bayesian hierarchical models. *Scandinavian Journal of Statistics*, *36*(2), 320–336. https://doi.org/10.1111/j.1467-9469.2008.00630.x

Storvik G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, *50*(2), 281–289. https://doi.org/10.1109/78.978383

https://doi.org/10.1093/jrsssa/qnad056
Advance access publication 27 April 2023