



Bayesian monitoring of COVID-19 in Sweden

Robin Marin^a, Håkan Runvik^b, Alexander Medvedev^b, Stefan Engblom^{a,*}

^a Division of Scientific Computing, Department of Information Technology, Uppsala University, SE-751 05, Uppsala, Sweden

^b Division of Systems and Control, Department of Information Technology, Uppsala University, SE-751 05, Uppsala, Sweden

ARTICLE INFO

Keywords:

Bayesian forecasting
Public health situation awareness
Data-driven epidemics
Compartment-based model
Kalman filtering

ABSTRACT

In an effort to provide regional decision support for the public healthcare, we design a data-driven compartment-based model of COVID-19 in Sweden. From national hospital statistics we derive parameter priors, and we develop linear filtering techniques to drive the simulations given data in the form of daily healthcare demands. We additionally propose a posterior marginal estimator which provides for an improved temporal resolution of the reproduction number estimate as well as supports robustness checks via a parametric bootstrap procedure.

From our computational approach we obtain a Bayesian model of predictive value which provides important insight into the progression of the disease, including estimates of the effective reproduction number, the infection fatality rate, and the regional-level immunity. We successfully validate our posterior model against several different sources, including outputs from extensive screening programs. Since our required data in comparison is easy and non-sensitive to collect, we argue that our approach is particularly promising as a tool to support monitoring and decisions within public health.

Significance: Using public data from Swedish patient registries we develop a national-scale computational model of COVID-19. The parametrized model produces valuable weekly predictions of healthcare demands at the regional level and validates well against several different sources. We also obtain critical epidemiological insights into the disease progression, including, e.g., reproduction number, immunity and disease fatality estimates. The success of the model hinges on our novel use of filtering techniques which allows us to design an accurate data-driven procedure using data exclusively from healthcare demands, i.e., our approach does not rely on public testing and is therefore very cost-effective.

1. Introduction

The results in this paper stem from the work carried out within the cross-disciplinary research project CRUSH Covid at Uppsala University.¹ Starting in the fall 2020, every week the group published a widely circulated report covering the region's COVID status by, e.g., collecting data from PCR tests, mobile apps, wastewater analysis, and health care. Our contribution consisted of a Bayesian disease-spread model which provided decision support in the form of predictions of health care demands as well as additional epidemiological insight.

There has been a multitude of attempts to model and forecast the spread of the virus. A problem often encountered is that, although data might appear abundant, fitting a given model to large volumes of data of various quality does not necessarily imply a high prediction accuracy (Shinde et al., 2020). A related issue is to identify the right level of model granularity: several aspects of the disease transmission

are relevant and need to be modeled, from small scale *in vitro* properties to global interventions. Some modeling efforts therefore include multiple levels of resolutions to capture, e.g., global travel patterns (Davis et al., 2021) or local within-country dynamics (Gatto et al., 2020). Understanding how to combine the various scales can substantially benefit the fidelity of scenario generators (Jordan et al., 2021). It is fair to say that models which have been in actual use are understudied due to time constraints and therefore often lack a thorough uncertainty analysis (Edeling et al., 2021; Soltesz et al., 2020). With the frequent lack of high-quality data for the current state of the disease, *nowcasting* has been increasingly critical in decision making (Wu et al., 2020; Altmejd et al., 2023; Hawryluk et al., 2021; Wu et al., 2021).

Since the situation concerns modeling under data limitations and with potentially large process uncertainties our proposed solution consists of a Bayesian framework. This approach involves adapting a

* Corresponding author.

E-mail addresses: robin.marin@it.uu.se (R. Marin), hakan.runvik@it.uu.se (H. Runvik), alexander.medvedev@it.uu.se (A. Medvedev), stefane@it.uu.se (S. Engblom).

¹ <https://www.uu.se/forskning/projekt/crush-covid/>

<https://doi.org/10.1016/j.epidem.2023.100715>

Received 8 June 2022; Received in revised form 28 July 2023; Accepted 16 August 2023

Available online 24 August 2023

1755-4365/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

linear noise approximation (Fearnhead et al., 2014) which enables fusion of different data sources and supports a computationally cheap approximate likelihood function via linear filters. We also investigate the posterior model not only through the posterior predictive distribution (Gelman et al., 1995; Gabry et al., 2019), but also by estimating the bias introduced by the approximate likelihood using ideas from parametric bootstrap (Engblom et al., 2020).

Our disease spread model attempts to balance three key qualities: interpretability, quantifiable uncertainty, and forecasting accuracy. The posterior model was investigated through marginal estimators, by comparing our results to several other sources, and by bias estimates obtained via bootstrap arguments. As this paper demonstrates, the achieved accuracy and robustness are quite remarkable considering that no data from screening programs were used.

2. Material and methods

Below we first summarize the Swedish publicly available data, and then present the associated design decisions made in developing the computational model. An important technical contribution lies in the techniques which support a computationally efficient approximate likelihood via linear filters. Two ‘bootstrap’ procedures are also outlined: one for improving the temporal resolution of the reproduction number estimates, and one for bounding the inversion bias through the generation of synthetic data. Further technical details concerning the derivation of the linear filters, data pre-processing, and optimization algorithms are found in the Supporting Information (SI).

2.1. Swedish COVID-19 data

In Sweden, the publicly available time series data for the COVID-19 pandemic fall in one of two categories: hospital load and results from PCR testing. Cumulative national-level disease severity statistics have also been made available and updated approximately once a month throughout the pandemic. The 21 regional councils compile hospital data and report the number of patients undergoing inpatient or intensive care, and also the number of deceased individuals. These numbers are reported on a daily basis and have been judged to be of consistent quality over sufficiently long periods of time to be used in our modeling. We retrieve those data from the portal initiative *c19.se*, which in turn collects the data from the regional councils. For validation, we have compared with official public registries, including the Swedish Public Health Agency (PHA), the National Board of Health and Welfare (NBHW), and the Swedish Intensive Care Registry (SIR). There are occasional inconsistencies in the data which need to be filtered away; see the SI for our quite basic approach for this.

The Swedish daily incidence as reported from PCR testing has been poor in several periods of time due to restrictions and changes in testing recommendations (Winblad et al., 2022). In June 2020, the Swedish government appointed a commission to evaluate COVID-19 measures, including, among other things, the testing programs. They found that the time from booking a PCR test to receiving the test results exceeded six days across several regions during the period of time studied in this paper, with additional time delays in publishing incidence results on the regional and municipality levels (Almgren and Björk, 2021). For these reasons, we judged the incidence data to be unreliable and excluded it from our model. Note, however, the direct comparison with the incidence data in Fig. 5.

Self-reporting via mobile apps has been proposed as a cheap and fast alternative to PCR testing (Kennedy et al., 2022). However, the validity of the signal depends on symptoms that overlap with other respiratory infections. For example, the PHA noted a high occurrence of symptoms of acute respiratory infections by the start of the Swedish second wave in fall 2020 (Kennedy et al., 2022; Folkhälsomyndigheten, 2021d). Laboratory analyses of respiratory viruses later indicated a high incidence of common colds caused by rhinoviruses during the

same period (Karolinska universitetslaboratoriet, 2022). Another alternative data source is the surveillance of wastewater (Saguti et al., 2021; Galani et al., 2022). Due to the signal’s large relative noise ratio, we did not consider it in this study but left it for future work.

2.2. Bayesian COVID-19 model

Given the previous considerations of data sources, we formulate our model around data in the form of daily observations of patients under hospital care (H), intensive care (W), and reported deaths (D), all in the 21 Swedish regions (counties). Any other sources of data have been used for comparisons *a posteriori* only.

In the standard SEIR model, susceptible individuals S become exposed E (without symptoms), and after progressing to a symptomatic infectious state I , they become recovered R . Based on the available data, we extend the SEIR model with the states (H, W, D), and regard them as worsened states of the symptomatic infection such that only a certain fraction of the infected individuals will enter them. It is widely accepted that not all exposed individuals become symptomatic (Byrne et al., 2020) and hence we also extend the model by including an asymptomatic state A with no or very mild symptoms.

The transmission is usually driven by random or time-dependent contact intensities between the susceptible and the infected individuals. As in Anderson and May (1981), Widgren et al. (2018), Engblom et al. (2020), van den Driessche (2017), we rather consider *implicit spread* via an infectious pressure compartment ϕ , an environmental state variable which models the current force of infection of the virus and which decays exponentially with time. Given that the decay rate of the infectious pressure is comparably fast and since direct spread can be understood as a timescale separation limit of indirect spread, most results should be robust with respect to this particular modeling choice (see also Benson et al., 2021).

The COVID-19 model in Keeling et al. (2021) sources the (explicit spread) infectivity from both symptomatic and asymptomatic carriers. We additionally allow for pre-symptomatic spread by sourcing the infectious pressure from all individuals in the states (E, A, I). Originally, we intended for our framework to regularly simulate spread over a national network defined by commuting intensities between regions. However, in the end we did not routinely incorporate network spread in the model since with our available data, letting the regions function as independent nodes reduces the computational complexity while giving a very similar data fit.

The resulting model is summarized schematically in Fig. 1, while a detailed description of the model and all parameters are found in the SI. With increased model complexity follows explanatory power at the expense of poorer model identifiability. With this in mind, we exclude all model refinements that are either missing or are unreliably or incompletely reported in the data, e.g., age and gender as well as certain refined states of minor symptoms. We recognize the infection’s varying effects on different age groups (Folkhälsomyndigheten, 2021b) and so must accept that our results remain in an age-averaged regime.

Increased model complexity which, however, is required involves using a mix of dynamic and static parameters, since this allows the model to respond to functional changes such as societal interventions (Haug et al., 2020), vaccinations and virus mutations (Liu and Rocklöv, 2021). We thus let β_t , that is, the infection rate which is related to the reproduction number R_t , as well as the infection fatality rate (IFR) both be time-varying parameters. All in all the problem is then to determine the posterior distribution for 10 static and two dynamic parameters. The latter are assumed constant for periods of four weeks but are re-sampled independently for each such period.

Clearly, well-chosen priors are required to make the problem definite. Considerable work went into constructing and continuously updating our priors using published research and public registries; the final priors are displayed in Fig. 2, with a complete list of priors and prior predictive estimates given in the SI.

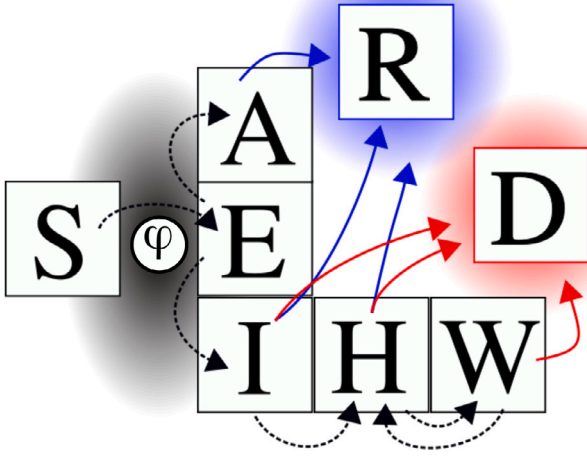


Fig. 1. Proposed COVID-19 compartment model with indirect transmission. (S)usceptible contract the virus by exposure to an infectious pressure (φ), become (E)xposed and then enter either the (A)symptomatic or the (I)nfectious state. Symptomatic severity follows: (I)nfectious to (H)ospitalized to (W)orsen (intensive care), and to (D)eath. Only states A, I, and H can (R)ecover from the disease. The observables are H, W, and D. The infectious pressure is sourced from individuals in the states (E, A, I) and decays exponentially with time.

2.3. Approximate likelihoods through linear filtering

We understand the compartment model as a continuous-time Markov chain (CTMC) over an integer lattice counting the number of individuals in the different compartments. Hence the waiting time for exiting one compartment is exponentially distributed of mean rate $1/\lambda$, and the number of individuals which exit in a small window of time $[t, t + \Delta t)$ is Poissonian $\sim \text{Po}(n_t \lambda \Delta t)$, with n_t the number of individuals at time t in the compartment. If the number of individuals is large enough, this transition can be approximated by the normal distribution $\sim \mathcal{N}(n_t \lambda \Delta t, n_t \lambda \Delta t)$, which can be directly translated into a contribution to the state update matrices (F_k, Q_k) of a discrete-time Kalman filter with the equations of state

$$x_{k+1} = F_k x_k + w_k, \quad w_k \sim \mathcal{N}(0, Q_k), \quad (1)$$

with k the discrete time index corresponding to days.

The linear filter allows for an approximation of the intractable likelihood of a parameter proposal $\tilde{\theta}$, namely the marginal filter likelihood,

$$p_{\tilde{\theta}}(y) = \prod_{k=0}^T \mathcal{N}(y_k | H_k \hat{x}_{k|k-1}, Q_k), \quad (2)$$

where y_k is the data at time k , H_k the observation matrix which maps the state to an observation, and where $\hat{x}_{k|k-1}$ is the Kalman state predictor at time k given data until time $k - 1$. In our application we rely on the likelihood to produce an approximate sample (θ_i) from the posterior using the Adaptive Metropolis (AM) algorithm (Haario et al., 2001). The role of the marginal likelihood is remindful of a *synthetic likelihood* (Wood, 2010) and we refer to the specific combination of a Kalman marginal Likelihood and Adaptive Metropolis as KLAM.

2.4. Improved transmission rate estimation

The achievable temporal resolution of the Bayesian parameter estimates provided by Metropolis sampling is limited by both computational complexity and the amount of data. The procedure described so far yields static reproduction number estimates for each four-week period and with comparably large spread. To obtain more fine-grained

estimates, a different approach is needed. Since the reproduction number is the dominating parameter of the dynamics, a more highly resolved estimate of the reproduction number is particularly useful. In terms of the parameters of the model, this corresponds to improving the estimation of β_t . Therefore, a daily estimate $\beta(t_k) = \beta_k$ is calculated using dynamic optimization techniques.

Dynamic optimization has been used for estimating the reproduction number of the COVID-19 outbreak also by others (Russo et al., 2020). When combined with an existing posterior distribution, care must be taken to avoid overconfidence from using the same data twice. For this reason, we do not attempt to derive an improved posterior distribution of β_t , but instead a single marginal time-dependent maximum likelihood estimate is sought, where the rest of the posterior distribution is “frozen” and consequently the parameter uncertainty is the same in absolute terms. The same logarithmic marginal likelihood that is used in the Kalman filter is utilized for this purpose, now understood as a quadratic cost function. Here, however, the deviation between measurements and the outputs from the mean-field dynamics is minimized, i.e., a shortened formulation compared to the filter is employed since the Kalman correction step is neglected. To avoid fast variations in β_k , a regularizing term penalizing square gradients is also added. The resulting optimization problem can be solved using standard techniques. Further details on the formulation of the optimization problem and its solution are provided in the SI.

Apart from providing more detailed information, this procedure yields improved confidence in the Bayesian workflow since it supports synthetic data with a known truth to be simulated in an off-line fashion. Our Bayesian inversion may thus be employed a second time in order to estimate bias or sensitivities for various estimates of interest, next to be described.

2.5. Assessing the quality of the approximate posterior

In real applications, a “true” or a “best” parameter posterior \mathbf{P}^* is usually unknown. Evaluating the stability and the quality of the approximate posterior $\tilde{\mathbf{P}}$ is unfortunately often overlooked. We suggest employing a *parametric bootstrap* approach as in Engblom et al. (2020) to assess the error between samples from the true and the approximated posterior $\tilde{E} := \tilde{\theta} - \theta^*$, where $\theta^* \sim \mathbf{P}^*$ and $\tilde{\theta} \sim \tilde{\mathbf{P}}$. Denote by $\theta^* := \mathbb{E}[\theta^*]$, the minimum mean square error estimator (MMSE) of an assumed truth. Decomposing the mean square error around this value we find

$$\tilde{e}^2 := \mathbb{E}[(\tilde{\theta} - \theta^*)^2] = \underbrace{\mathbb{E}[(\tilde{\theta} - \tilde{\theta})^2]}_{\text{Variance}} + \underbrace{(\tilde{\theta} - \theta^*)^2}_{\text{Square bias} =: \tilde{b}^2}, \quad (3)$$

where $\tilde{\theta} := \mathbb{E}[\tilde{\theta}]$ is the MMSE of $\tilde{\theta}$.

Formally, this still requires samples from the true posterior when estimating the bias. We approximate this via a bootstrapped estimator using a sample of N_{boot} synthetic data sets generated from the MMSE of the approximate posterior. The generative simulator requires daily estimates of (β_k) or else the synthetic data quickly drifts off compared to the observations, and thus this technique ultimately hinges on the highly resolved marginal estimator described above. Posterior samples may then be generated for each synthetic set, yielding now a set of samples $\hat{\theta}_i \sim \tilde{\mathbf{P}}$, which allows for the use of the now tractable bias estimator $\tilde{b}_i := \mathbb{E}[\hat{\theta}_i] - \mathbb{E}[\tilde{\theta}]$. Our final estimator is then an average over these synthetic sets; $\tilde{b}^2 \approx N_{\text{boot}}^{-1} \sum_i \tilde{b}_i^2$ in (3). While up to 196 000 samples from each posterior were used to compute point estimators and credible intervals (CrIs), bootstrap replicas are much more costly to process so we used $N_{\text{boot}} = 3$, and mainly relied on the bias estimator to diagnose non-robustness in point estimators. That is, a point estimator with CrI A of order $\alpha = 68\%$, say, and with bias estimate \tilde{b} is considered less robust whenever

$$\tilde{b} \geq 0.5 \text{diam}(A). \quad (4)$$

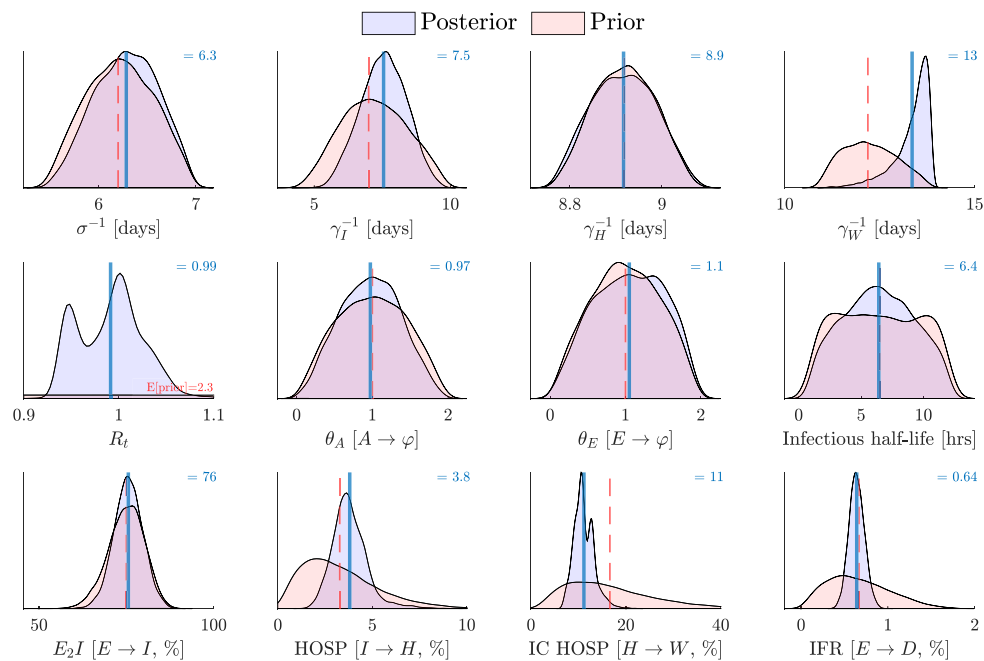


Fig. 2. Marginal priors (red) for the model parameters and the associated posterior distributions (blue) for the Swedish aggregate inferred from publicly available regional data from April 1, 2020 to May 30, 2021. The dashed lines indicate the prior means, and the full lines and numeric values the posterior means. *Top row:* latent period rate σ , and exit rates γ_X out of compartment $X \in \{I, H, W\}$. *Middle row:* reproduction number and contribution to the infectious pressure φ from compartments A and E , respectively, as well as its decay (half-life). *Bottom row:* fractions proceeding between the indicated compartments. The reproduction number R_t and the IFR are dynamic parameters and a temporal average is displayed here. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The bootstrap posterior densities can be used in aggregate form for bias estimation and check of estimator robustness as just described, or for a related check of credible interval robustness. Consider two CrI intervals A and B of order α , e.g., $\alpha = 68\%$, and where B is a bootstrap replicate of A . A basic measure of the robustness of A is the level of overlap between A and B and one can reasonably require the same overlap as the indicated order α , i.e., to require

$$\text{diam}(A \cap B) \geq \alpha \text{diam}(A \cup B). \tag{5}$$

If this criterion is satisfied, then a random variable drawn uniformly from $A \cup B$ has probability $\geq \alpha$ to also be in $A \cap B$. The robustness checks Eqs. (4)–(5) are explicitly reported in Table 2, but were also routinely employed when evaluating various results.

3. Results

Our model was used for weekly reported predictions within CRUSH Covid. Prior to each report we carried out a model updating procedure: new data were pulled from public repositories and screened for contradictory or incorrect values. The posterior was sampled by KLAM using an initialization either from an estimated initial state as described in the SI and developed for this very purpose, or simply using stored state samples. The latter allows for faster sampling as it reduces the burn-in period: about 24 h of compute for 21 regions and a year’s worth of data on a 4-core laptop was then reduced to a few hours. The final posterior model was queried for one-week-ahead forecasts with uncertainty bounds for the (H, W, D) triple. We also continuously evaluated the previous week’s predictions against the up-to-date data in the same round.

3.1. Posterior prediction

In Fig. 2, we display the prior distributions together with the resulting Swedish *aggregate* posterior, i.e., the population weighted average of the individual posterior of each of Sweden’s 21 regions. Several priors are clearly very similar to their respective posteriors,

e.g., the latent period rate σ and the symptomatic period rate γ_I . This is expected and simply indicates little information in the observations for these parameters relative to the prior. Our data also cannot improve on the prior for the share of spread from exposed (pre-symptomatic) and asymptomatic individuals (parameters θ_E and θ_A , respectively). This is simply due to the fact that we have no continuously reported data neither for E nor for A . On the same note we had to rely on a directed study (Byrne et al., 2020) to define the prior for the parameter E_2I , i.e., the fraction of exposed individuals who eventually develop symptoms.

Of more interest are the parameters that govern the fraction that transition to a worsened state of the disease: $I \rightarrow H$ and $H \rightarrow W$. Our results show that 3.8% [2.2, 5.9] (95% Credible Interval (CrI)²) of the symptomatic individuals require hospital care, and 11.2% [7.6, 16.1] of the hospitalized patients require intensive care. During the considered period, the Swedish Public Health Agency (PHA) published five point estimates of those same fractions. The relevant demographic average of these are $I \rightarrow H$: 3.0% (2.4, 3.6) and $H \rightarrow W$: 14.3% (9.8, 18.8) (mean ± 2 std, $n = 5$) respectively (Folkhälsomyndigheten, 2021b). Additionally, Salje et al. (2020) found similar estimates in France $I \rightarrow H$: 2.9% [1.7, 4.8] and $H \rightarrow W$: 19.0% [18.7, 19.4]. We could have used the earlier of those point estimates to improve on the corresponding priors, however, as validation possibilities are scarce, we decided to rather use them for this purpose instead.

The regional models were used for posterior predictions on a weekly basis, e.g., one 7-day ahead prediction for the Uppsala and Stockholm regions, respectively, and one in aggregate for the entire nation. In Fig. 3 this is exemplified over a longer period together with the actual outcome. The performance of the weekly predictions which were reported live ($N = 25$) is presented in Table 1. Each prediction included a mean with 68/95% CrI. The week after publishing, we evaluated the

² We indicate Credible Intervals (CrI) by square brackets [·,·] and Confidence Intervals (CI) by regular parentheses (·,·), unless otherwise specified.

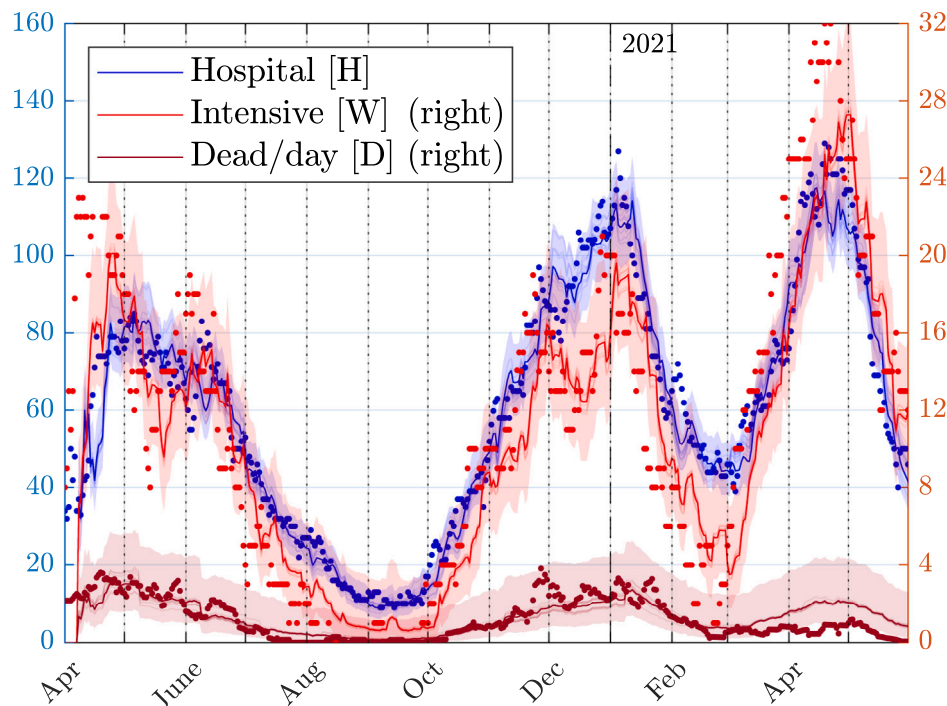


Fig. 3. Seven-day ahead prediction in the Uppsala region. Shaded area shows 68% CrI, and the points are the observations. The dotted vertical lines indicate the four-week periods used for the dynamic rates.

Table 1

Frequency (%) of all weekly reported predictions that fell inside of (68/95% CrI, 7 days ahead), evaluated on the following week (N = 25). Note that the Uppsala results are very close to the ideal 68/95% outcome.

	Hospital (H)	Intensive (W)	Death (D)
Uppsala (68%)	76	72	68
Stockholm	64	64	36
Sweden	88	44	32
Uppsala (95%)	100	100	96
Stockholm	84	92	68
Sweden	96	92	68

prediction against the then-available data. The predictions performed better in the medium sized region Uppsala and notably, the predictions for casualties were poor in the larger region Stockholm. The observed misfit in casualties eventually lead us to reconsider the role of the IFR parameter, initially just a static parameter, and we decided to let it be dynamic as described previously. In the live reporting in Table 1, only 9 of the 25 reports allowed for a dynamic IFR. Another possible reason for the poorer prediction performance in the Stockholm region could be that, since this region contains three large hospitals, the greater heterogeneity in terms of reporting makes identification more challenging. One can rightly question if smaller sub-regions should rather be modeled here, but we did not have access to the data to drive such a model.

In the SI we further compare the quality of the Kalman predictor with that of a simpler regression-based estimator. The latter provides accurate mean-square predictions and is very fast to evaluate. The overall advantage with our approach lies rather in that the Bayesian posterior model itself can be investigated for further epidemiological insight, next to be discussed.

3.2. Posterior hidden state estimation

The posterior model can also be used as a kind of “Bayesian twin” and estimate quantities that are otherwise very difficult to approach.

For example, we can readily estimate the number of individuals that have contracted the disease and survived; these individuals are the ones that could potentially have developed antibodies that are detectable in serology tests. In Fig. 4, we visualize several reported results for Stockholm (Castro Dopico et al., 2021; Folkhälsomyndigheten, 2021c) and our estimates. Those estimates compare very well, and notably so given that no screening data was used by our method.

The posterior model can also estimate the symptomatic incidence in the same vein. Fig. 5 illustrates our estimated symptomatic incidence and the reported number of positive RT-PCR tests by the PHA for Uppsala. As testing increases, the ratio between our estimate and the positive tests oscillates around one, indicating that the testing at that time captures most of the symptomatic infected.

These two examples demonstrate that our computational model can be used to effectively estimate hidden variables at the same regional resolution as supported by data. The financial cost for this kind of monitoring would of course be a tiny fraction compared to any alternatives based on testing.

3.3. Marginal risks

Recall that the IFR is defined as the proportion of deaths among infected individuals, i.e., including asymptomatic cases. By design, our model relies on an IFR which is constant over four weeks. Our estimated IFR for Sweden stayed relatively constant over the period May 2020–November 2020 at 0.69% [0.11, 1.5] (95% CrI). For comparison, in April 2020, the PHA published an early estimate of 0.58% (0.37, 1.05) (95% CI) for Stockholm, Sweden (Folkhälsomyndigheten, 2021a). However, this estimate relies on initial assumptions on the number of undetected cases that seem unjustified when compared to later findings (Rippinger et al., 2021; Irons and Raftery, 2021). It also assumes the relatively younger Stockholm demographics rather than the national one, and is therefore an underestimate of the national IFR. A later Sweden-wide IFR estimate of 0.76% [0.65, 0.87] was published in November 2020 (Garcia-Ptacek et al., 2021) and aligns well with our estimate above. Although our 95% CrI is comparatively wide, this is partially due to

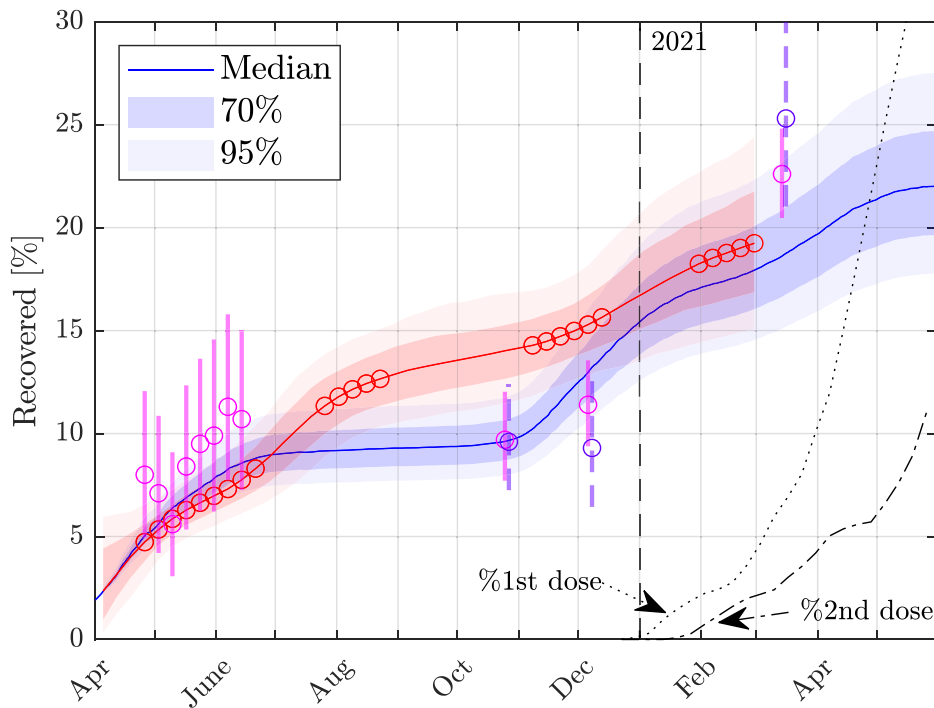


Fig. 4. The fraction of recovered individuals in the Stockholm region. Our model (blue), is set to 2.35% on April 5, 2020 (matching Castro Dopico et al., 2021). Serology-based Bayesian model predictions (Castro Dopico et al., 2021) (red) which reported 70/95% CrIs. Estimated mean prevalence of antibodies found in spared blood samples from outpatient care, with 95% CI (solid, pink) and blood donors (dashed, purple) (Folkhälsomyndigheten, 2021c). On 2020, December 27, the Swedish vaccination campaign started and both the number of delivered first doses (dotted) and second doses (dash-dot) are indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

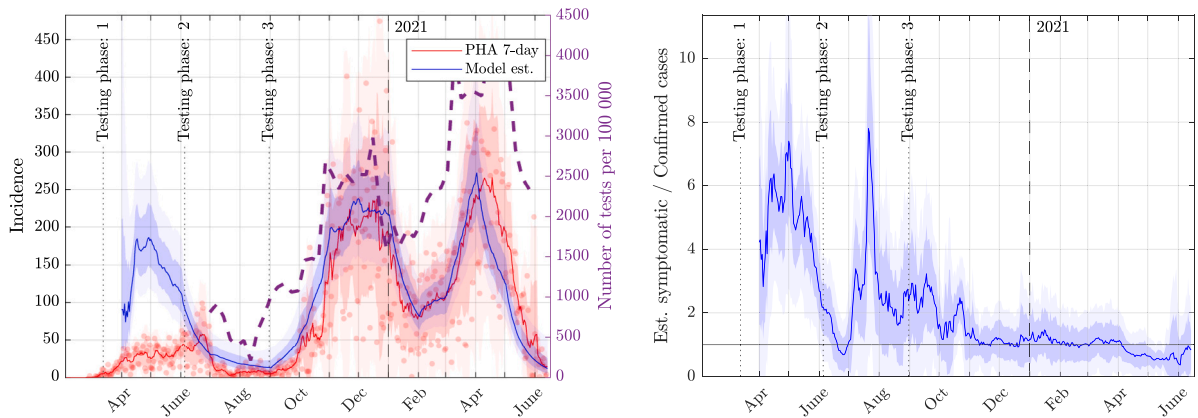


Fig. 5. *Left:* the estimated symptomatic incidence from our model (blue with [68,95%] CrIs shaded) and the confirmed cases by PHA (red line 7-day smoothed and uncertainty from rolling $\pm[1,2]$ standard deviations shaded, with data as points) for Uppsala. The right axis gives the number of tests per 100000 inhabitants per week (dashed). The total number of tests administered during Testing phase 3 was approximately 360 000. The dotted vertical lines mark the boundaries of testing phases as defined by policy changes affecting the testing volumes. *Right:* the ratio between the model’s symptomatic incidence and the confirmed cases incidence indicates the proportion of cases that are missed by testing. Values above one thus hint at an underreporting of symptomatic cases. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

modeling each period independently and without any regularization in the transition between periods.

Table 2 summarizes a few bi-monthly estimates for the period after October 2020. The first two entries in the table overlap the estimates from Anon (2022): Nov 2020 = [0.60, 1.46]%, and Jan 2021 = [0.56, 1.44]%. Clearly, the IFR was trending downwards and this is also known to be the case in Stockholm (Garcia-Ptacek et al., 2021) as well as for the world in general (Anon, 2022).

Of interest is also the *case fatality risk* (CFR) (Kelly and Cowling, 2013), i.e., the risk of death conditioned on being diagnosed with the disease. We more generally define CFR_X as the proportion of

deaths expected given a certain number of individuals in compartment $X \in \{I, H, W\}$. Note that CFR_I involves the number of symptomatic individuals which formally is not the same thing as the number of cases confirmed by testing.

From our posterior we may directly estimate the national average CFRs to be $\{[0.72, 1.3], [16, 19], [34, 36]\}$ % (95% CrIs) for CFR_I , CFR_H , and CFR_W . By comparison, Alimohamadi et al. (2021) offers the estimates $CFR_I = (1.0, 3.0)\%$, $CFR_H = (9.0, 17.0)\%$, and $CFR_W = (24.0, 51.0)\%$ (95% CI).

Another way to investigate these risks is by running the posterior filter across our data to produce an estimate for the number of deceased

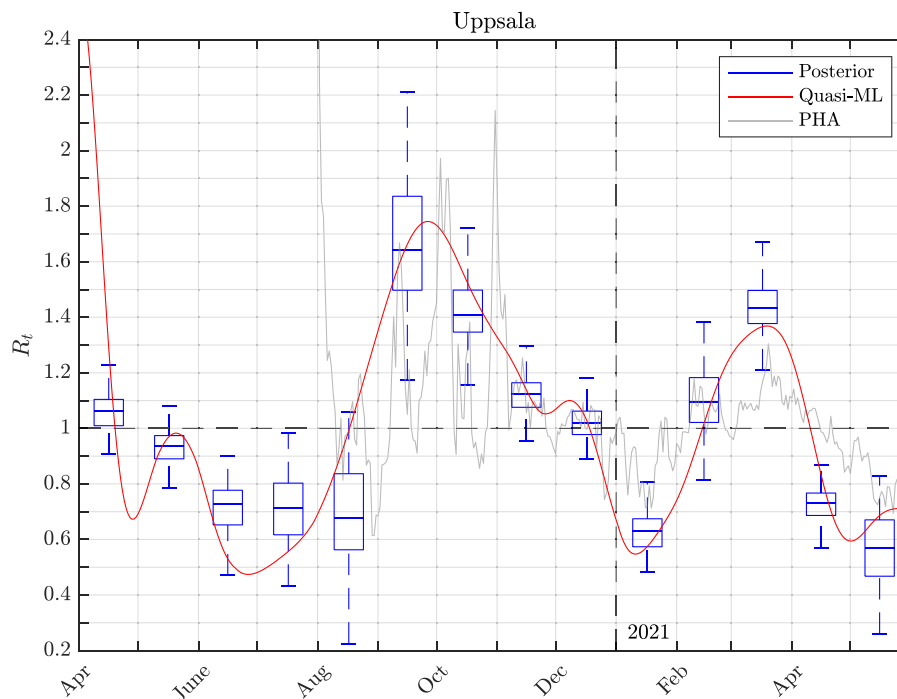


Fig. 6. Reproduction number estimates for the Uppsala region. The Bayesian posterior yields monthly estimates (box-plot) while the quasi-ML estimator is daily (solid). For comparison also the PHA’s estimates based on screening from of a total of 367200 RT-PCR tests during the period 5 August, 2020–May 26, 2021. Note that reproduction numbers are model specific and the comparison here is therefore mainly qualitative.

Table 2
Bi-monthly estimated IFR [%] with 68% CrI from October 2020 to May 2021.

	Stockholm	Uppsala	Sweden
Oct+Nov	^b 0.66 [0.34, 0.81]	0.49 [0.24, 0.74]	0.55 [0.27, 0.88]
Dec+Jan	^b 0.84 [0.71, 1.00]	^b 0.59 [0.28, 0.92]	0.88 [0.55, 1.30]
Feb+Mar	^{ab} 0.35 [0.18, 0.50]	0.47 [0.23, 0.70]	0.44 [0.22, 0.81]
Apr+May	0.35 [0.19, 0.49]	0.45 [0.22, 0.67]	0.35 [0.18, 0.56]

^aThe estimated bias is large compared to the 68% CrI (see Material and Methods).

^bThe 68% CrIs of the posterior and the bootstrap replicate do not share at least a 68% overlap (see Material and Methods).

per compartment. Until March 23, 2021 we find for all of Sweden that $\{D_I, D_H, D_W\} = \{[1\ 786, 8\ 652], [3\ 612, 8\ 096], [1\ 126, 3\ 627]\}$ (95% CrIs). From NBHW data we may estimate those same numbers to be $\{8\ 335, 4\ 102, 935\}$ [Socialstyrelsen \(2020\)](#), after a scaling of about 5% to arrive at the same total number of dead as in our dataset ($D = 13\ 372$). Apparently, our model overestimates the deaths under ICU and to some extent also at hospital, while our estimate for deaths outside hospital is underestimated compared to this data point. This could likely improve given more detailed data sources, including, e.g., improved records of COVID-related deaths and hospital care outcomes.

3.4. Reproduction number estimates

The *reproduction number* provides an essential insight into the future development of the spread of a disease in a population. An accurate estimate of this number is vital to support rational public health decision-making and to inform the general public. Temporal variations in the reproduction number are caused by socio-behavioral, environmental, and virological and biological factors ([Delamater et al., 2019](#)). The dynamics of the reproduction number is therefore significantly faster than those of most other parameters, which was also the motivation behind our development of an improved marginal estimator via dynamic optimization techniques as described previously.

Reproduction number estimates have been calculated in this way for the whole duration of the parametrized period for all regions in Sweden. In [Fig. 6](#), we present our results for Uppsala together with the testing-based estimate produced by the PHA for the same period. The cost-effectiveness of our estimate is apparent here since the results are similar, but the PHA estimate relies on costly incidence data.

4. Discussion

The pathogen SARS-CoV-19 and the subsequent pandemic resulted in an explosion in research related to COVID-19: research on data collection and interpretation, modeling and forecasting, as well as scenario generation. From the beginning of the outbreak, data have been instrumental in understanding the disease dynamics ([Li et al., 2020](#); [Lavezzo et al., 2020](#); [Wu et al., 2020](#)). With increasing amounts of cases and recorded patient data, statistical models for diagnosis and prognosis were quickly developed ([Wynants et al., 2020](#)). Combining in-patient data with other covariates, the use of digital technologies for disease surveillance is now possible ([Abad et al., 2021](#)), notably with an impact also for COVID-19 ([Menni et al., 2020](#); [Drew et al., 2020](#); [Rossman et al., 2020](#); [Kennedy et al., 2022](#)). Challenges connected to the collection and distribution of available data where met with specialized tools for decentralized publishing ([Guidotti and Ardia, 2020](#); [Reinhart et al., 2021](#)) and anonymized mobility data were also in active use ([Imran et al., 2022](#); [Ilin et al., 2021](#)).

We have devised a detailed Bayesian model for the regional dynamics of the COVID-19 pandemic in Sweden. A data-centric viewpoint is that, using *model-based data analysis*, we have gained a thorough insight into the progression of the COVID-19 pandemic in Sweden in general, with extra emphasis on the Uppsala Region. The proposed compartment-based model combined with the novel use of optimal linear filters turned out to be an effective information-theoretic epidemiological tool. The output quality as obtained in our work compares very well with official estimates gathered in test-based programs, cf. [Figs. 4–6](#). During the second and third waves in Sweden, at least 10 million RT-PCR tests were administered at a standard cost of 1400

SEK (\approx \$150) per test (Sveriges Kommuner och Regioner, 2021). The expenses for such a PCR test for all symptomatic-strategy quickly grows, underlining the economical advantages of our approach. Clearly, at the individual level there are several benefits with testing and the importance of screening as a means to collect initial statistics for the disease spread cannot be stressed enough. However, our approach remains very promising as a supporting tool to continue to monitor the situation when testing is limited due to risk-cost trade-offs.

The resulting model was further processed to output weekly predictions for health care demands, and also marginal estimators for important characteristics of the disease such as infection fatality rates and reproduction numbers. The latter output increased the confidence into the overall approach through the generation of synthetic data and parametric bootstrap techniques. Improved data that would have enabled a higher model precision include (1) a consistently managed incidence report from randomized testing (not necessarily high volume), and (2) a higher temporal resolution of hospitalization and intensive care risks as well as times for treatment in these respective categories. These statistics could both be collected at a relatively small cost but would likely improve the precision considerably.

Without relying on public testing strategies, our model-based approach provided improved *situation awareness* of the progression of the pandemic. The developed methodology is of highly general character and can therefore be expected to be useful in other contexts too. By virtue of the consistent Bayesian framework, uncertainties are transparently propagated under clear assumptions, also in the face of potentially hazardous situations. We argue that this quality makes the techniques developed herein particularly promising from a communicative perspective.

Our data streams are high in latency, but are on the other hand fairly low in noise. Low-latency signals, e.g., public screening, self-reporting mobile apps, or analysis of sewage water, are instead often more noisy or otherwise biased. Combining these different kinds of streams provides for excellent decision support and appears extremely promising for use in tracking regional epidemics at a near-daily resolution.

CRedit authorship contribution statement

Robin Marin: Developed the initial forward model, Adapted the Bayesian sampling techniques, Collected data and performed computations, Prepared and revised the manuscript. **Håkan Runvik:** Designed linear filters, Developed and operated the dynamic optimization approach, Drafted and revised the manuscript. **Alexander Medvedev:** Contributed to the dynamic optimization approach, Revised the manuscript. **Stefan Engblom:** Conceived the research, Developed the initial forward model, Designed linear filters, Developed priors and data pre-processing, Prepared and Revised the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All associated data, computer codes, and experimental scripts are available online on GitHub.

Acknowledgments

This work was financially supported by the Swedish Innovation Agency Vinnova, by the Swedish Research Council Formas (S. Engblom), and by the Swedish Research Council (H. Runvik and A. Medvedev).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.epidem.2023.100715>.

References

- Abad, Z.S.H., Kline, A., Sultana, M., Noaen, M., Nurmambetova, E., Lucini, F., Al-Jefri, M., Lee, J., 2021. Digital public health surveillance: A systematic scoping review. *NPJ Digit. Med.* 4 (1), 1–13. <http://dx.doi.org/10.1038/s41746-021-00407-6>.
- Alimohamadi, Y., Tola, H.H., Abbasi-Ghahramanloo, A., Janani, M., Sepandi, M., 2021. Case fatality rate of COVID-19: A systematic review and meta-analysis. *J. Prev. Med. Hyg.* 62 (2), E311. <http://dx.doi.org/10.15167/2421-4248/jpmh2021.62.2.1627>.
- Almgren, M., Björk, J., 2021. Kartläggning av skillnader i regionernas insatser för provtagning och smittspårning under coronapandemin. Technical Report Coronakommissionen S. 2020:09, ISBN: 978-91-525-0257-0, Underlagsrapport till SOU 2021:89 Sverige under pandemin.
- Altmejd, A., Rocklöv, J., Wallin, J., 2023. Nowcasting COVID-19 statistics reported with delay: A case-study of Sweden and the UK. *Int. J. Environ. Res. Publ. Health* 20 (4). <http://dx.doi.org/10.3390/ijerph20043040>.
- Anderson, R.M., May, R.M., 1981. The population dynamics of microparasites and their invertebrate hosts. *Philos. Trans. R. Soc.* 291 (1054), 451–524. <http://dx.doi.org/10.1098/rstb.1981.0005>.
- Anon, 2022. COVID-19 Forecasting Team. Variation in the COVID-19 infection–fatality ratio by age, time, and geography during the pre-vaccine era: A systematic analysis. *Lancet* [http://dx.doi.org/10.1016/S0140-6736\(21\)02867-1](http://dx.doi.org/10.1016/S0140-6736(21)02867-1).
- Benson, L., Davidson, R.S., Green, D.M., Hoyle, A., Hutchings, M.R., Marion, G., 2021. When and why direct transmission models can be used for environmentally persistent pathogens. *PLoS Comput. Biol.* 17 (12), 1–26. <http://dx.doi.org/10.1371/journal.pcbi.1009652>.
- Byrne, A.W., McEvoy, D., Collins, A.B., Hunt, K., Casey, M., Barber, A., Butler, F., Griffin, J., Lane, E.A., McAloon, C., et al., 2020. Inferred duration of infectious period of SARS-CoV-2: Rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ Open* 10 (8), e039856. <http://dx.doi.org/10.1136/bmjopen-2020-039856>.
- Castro Dopico, X., Muschiol, S., Christian, M., Hanke, L., Sheward, D., Grinberg, N., Rorbach, J., Bogdanovic, G., Mcinerney, G., Allander, T., et al., 2021. Seropositivity in blood donors and pregnant women during the first year of SARS-CoV-2 transmission in Stockholm, Sweden. *J. Intern. Med.* <http://dx.doi.org/10.1111/joim.13304>.
- Davis, J.T., Chinazzi, M., Perra, N., Mu, K., Pastore y Piontti, A., Ajelli, M., Dean, N.E., Gioannini, C., Litvinova, M., Merler, S., et al., 2021. Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave. *Nature* 600 (7887), 127–132. <http://dx.doi.org/10.1038/s41586-021-04130-w>.
- Delamater, P.L., Street, E.J., Leslie, T.F., Yang, Y.T., Jacobsen, K.H., 2019. Complexity of the basic reproduction number (R0). *Emerg. Infect. Dis.* 25 (1), 1. <http://dx.doi.org/10.3201/eid2501.171901>.
- Drew, D.A., Nguyen, L.H., Steves, C.J., Menni, C., Freydin, M., Varsavsky, T., Sudre, C.H., Cardoso, M.J., Ourselin, S., Wolf, J., et al., 2020. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 368 (6497), 1362–1367. <http://dx.doi.org/10.1126/science.abc0473>.
- Edeling, W., Arabnejad, H., Sinclair, R., Suleimenova, D., Gopalakrishnan, K., Bosak, B., Groen, D., Mahmood, I., Crommelin, D., Coveney, P.V., 2021. The impact of uncertainty on predictions of the CovidSim epidemiological code. *Nat. Comput. Sci.* 1 (2), 128–135. <http://dx.doi.org/10.1038/s43588-021-00028-9>.
- Engblom, S., Eriksson, R., Widgren, S., 2020. Bayesian epidemiological modeling over high-resolution network data. *Epidemics* 32, 100399. <http://dx.doi.org/10.1016/j.epidem.2020.100399>.
- Fearnhead, P., Giagos, V., Sherlock, C., 2014. Inference for reaction networks using the linear noise approximation. *Biometrics* 70 (2), 457–466. <http://dx.doi.org/10.1111/biom.12152>.
- Folkhälsomyndigheten, 2021a. The Infection Fatality Rate of COVID-19 in Stockholm. Technical Report, <https://www.folkhalsomyndigheten.se/contentassets>. (Online; Accessed 21 January 2022).
- Folkhälsomyndigheten, 2021b. Scenarier för fortsatt spridning. pp. 2–7, <https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/statistik-och-analyser/analys-och-prognoser>. (Online; Accessed 21 January 2022).
- Folkhälsomyndigheten, 2021c. Påvisning av antikroppar mot SARS-CoV-2 (1) i blodprov från öppenvården, (2) hos blodgivare. <https://www.folkhalsomyndigheten.se/contentassets>. (Online; Accessed 21 January 2022).
- Folkhälsomyndigheten, 2021d. Veckorapport om COVID-19, vecka 39. <https://www.folkhalsomyndigheten.se/globalassets/statistik-uppfoljning/smittsamma-sjukdomar/veckorapporter-covid-19/2021/covid-19-veckorapport-2021-vecka-39-final.pdf>. (Online; Accessed 09 February 2022).
- Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., Gelman, A., 2019. Visualization in Bayesian workflow. *J. R. Stat. Soc. Ser. A* 182 (2), 389–402. <http://dx.doi.org/10.1111/rssa.12378>.

- Galani, A., Aalizadeh, R., Kostakis, M., Markou, A., Alygizakis, N., Lytras, T., Adamopoulos, P.G., Peccia, J., Thompson, D.C., Kontou, A., et al., 2022. SARS-CoV-2 wastewater surveillance data can predict hospitalizations and ICU admissions. *Sci. Total Environ.* 804, 150151. <http://dx.doi.org/10.1016/j.scitotenv.2021.150151>.
- Garcia-Ptacek, S., Xu, H., Annetorp, M., Jerlardtz, V.B., Cederholm, T., Engström, M., Kivipelto, M., Lundberg, L.G., Metzner, C., Olsson, M., Nyvang, J.S., Öberg, C.S., Åkesson, E., Religa, D., Eriksdotter, M., 2021. Temporal trends in hospitalizations and 30-day mortality in older patients during the COVID pandemic from march 2020 to july 2021. <http://dx.doi.org/10.1101/2021.12.22.21268237>, medRxiv preprint medRxiv:2021.12.22.21268237.
- Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., Rinaldo, A., 2020. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc. Natl. Acad. Sci. USA* 117 (19), 10484–10491. <http://dx.doi.org/10.1073/pnas.2004978117>.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. *Bayesian Data Analysis*. Chapman and Hall/CRC, <http://dx.doi.org/10.1201/b16018>.
- Guidotti, E., Ardia, D., 2020. COVID-19 data hub. *J. Open Source Softw.* 5 (51), 2376. <http://dx.doi.org/10.21105/joss.02376>.
- Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. *Bernoulli* 7 (2), 223–242. <http://dx.doi.org/10.2307/3318737>.
- 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. *Nat. Hum. Behav.* 4 (12), 1303–1312. <http://dx.doi.org/10.1038/s41562-020-01009-0>.
- Hawryluk, I., Hoeltgebaum, H., Mishra, S., Miscouridou, X., Schnekenberg, R.P., Whittaker, C., Vollmer, M., Flaxman, S., Bhatt, S., Mellan, T.A., 2021. Gaussian process nowcasting: Application to COVID-19 mortality reporting. *Uncertainty Artif. Intell.* 1258–1268.
- Ilin, C., Annan-Phan, S., Tai, X.H., Mehra, S., Blumenstock, J.E., 2021. Public mobility data enables COVID-19 forecasting and management at local and global scales. *Sci. Rep.* 11 (1), 1–11. <http://dx.doi.org/10.1038/s41598-021-92892-8>.
- Imran, M., Qazi, U., Ofli, F., 2022. TBCOV: Two billion multilingual COVID-19 tweets with sentiment, Entity, Geo, and Gender Labels. *Data* 7 (1), 8. <http://dx.doi.org/10.3390/data7010008>.
- Irons, N.J., Raftery, A.E., 2021. Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. *Proc. Natl. Acad. Sci. USA* 118 (31), <http://dx.doi.org/10.1073/pnas.2103272118>.
- Jordan, E., Shin, D.E., Leekha, S., Azarm, S., 2021. Optimization in the context of COVID-19 prediction and control: A literature review. *IEEE Access* <http://dx.doi.org/10.1109/ACCESS.2021.3113812>.
- Karolinska universitetslaboratoriet, 2022. Luftvägspatogener prov analyserade av Karolinska Universitetslaboratoriet till och med vecka 5 2022. <https://www.karolinska.se/globalassets/global/2-funktioner/funktion-kul/klinisk-mikrobiologi/epidemiologi/rapport-influensa-och-rs-virus-och-andra-luftvagspatogener.pdf>. (Online; Accessed 09 February 2022).
- Keeling, M.J., Hill, E.M., Gorsich, E.E., Penman, B., Guyver-Fletcher, G., Holmes, A., Leng, T., McKimm, H., Tamborrino, M., Dyson, L., Tildesley, M.J., 2021. Predictions of COVID-19 dynamics in the UK: Short-term forecasting and analysis of potential exit strategies. *PLoS Comput. Biol.* 17 (1), 1–20. <http://dx.doi.org/10.1371/journal.pcbi.1008619>.
- Kelly, H., Cowling, B.J., 2013. Case fatality: Rate, ratio, or risk? *Epidemiology* 24 (4), 622–623. <http://dx.doi.org/10.1097/EDE.0b013e318296c2b6>.
- Kennedy, B., Fitipaldi, H., Hammar, U., Maziarz, M., Tsereteli, N., Oskolkov, N., Varotsis, G., Franks, C.A., Nguyen, D., Spiliopoulos, L., Adami, H.-O., Björk, J., Engblom, S., Fall, K., Grimby-Ekman, A., Litton, J.-E., Martinell, M., Oudin, A., Sjöström, T., Timpka, T., Sudre, C.H., Graham, M.S., du Cadet, J.L., Chan, A.T., Davies, R., Ganesh, S., May, A., Ourselin, S., Pujol, J.C., Selvachandran, S., Wolf, J., Spector, T.D., Steves, C.J., Gomez, M.F., Franks, P.W., Fall, T., 2022. App-based COVID-19 syndromic surveillance and prediction of hospital admissions in COVID symptom study Sweden. *Nat. Commun.* 13, 2110. <http://dx.doi.org/10.1038/s41467-022-29608-7>.
- Lavezzo, E., Franchin, E., Ciavarella, C., Cuomo-Dannenburg, G., Barzon, L., Del Vecchio, C., Rossi, L., Manganelli, R., Lorigian, A., Navarin, N., et al., 2020. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo. *Nature* 584 (7821), 425–429. <http://dx.doi.org/10.1038/s41586-020-2488-1>.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y., et al., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* <http://dx.doi.org/10.1056/NEJMoa2001316>.
- Liu, Y., Rocklöv, J., 2021. The reproductive number of the delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. *J. Travel Med.* <http://dx.doi.org/10.1093/jtm/taab124>.
- Menni, C., Valdes, A.M., Freidin, M.B., Sudre, C.H., Nguyen, L.H., Drew, D.A., Ganesh, S., Varsavsky, T., Cardoso, M.J., Moustafa, J.S.E.-S., et al., 2020. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat. Med.* 26 (7), 1037–1040. <http://dx.doi.org/10.1038/s41591-020-0916-2>.
- Reinhart, A., Brooks, L., Jahja, M., Rumack, A., Tang, J., Agrawal, S., Al Saeed, W., Arnold, T., Basu, A., Bien, J., et al., 2021. An open repository of real-time COVID-19 indicators. *Proc. Natl. Acad. Sci. USA* 118 (51), <http://dx.doi.org/10.1073/pnas.2111452118>.
- Rippinger, C., Bicher, M., Urach, C., Brunmeir, D., Weibrecht, N., Zauner, G., Sroczynski, G., Jahn, B., Mühlberger, N., Siebert, U., et al., 2021. Evaluation of undetected cases during the COVID-19 epidemic in Austria. *BMC Infect. Dis.* 21 (1), 1–11. <http://dx.doi.org/10.1186/s12879-020-05737-6>.
- Rossmann, H., Keshet, A., Shilo, S., Gavrieli, A., Bauman, T., Cohen, O., Shelly, E., Balicer, R., Geiger, B., Dor, Y., et al., 2020. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. *Nat. Med.* 26 (5), 634–638. <http://dx.doi.org/10.1038/s41591-020-0857-9>.
- Russo, L., Anastassopoulou, C., Tsakris, A., Bifulco, G.N., Campana, E.F., Toraldo, G., Sietto, C., 2020. Tracing day-zero and forecasting the COVID-19 outbreak in Lombardy, Italy: A compartmental modelling and numerical optimization approach. *PLoS One* (ISSN: 19326203) 15 (10), e0240649. <http://dx.doi.org/10.1371/journal.pone.0240649>.
- Saguti, F., Magnin, E., Enache, L., Churqui, M.P., Johansson, A., Lumley, D., Davidsson, F., Dotevall, L., Mattsson, A., Trybala, E., et al., 2021. Surveillance of wastewater revealed peaks of SARS-CoV-2 preceding those of hospitalized patients with COVID-19. *Water Res.* 189, 116620. <http://dx.doi.org/10.1016/j.watres.2020.116620>.
- Salje, H., Kiem, C.T., Lefrancq, N., Courtejoie, N., Bosetti, P., Paireau, J., Andronico, A., Hozé, N., Richet, J., Dubost, C.-L., et al., 2020. Estimating the burden of SARS-CoV-2 in France. *Science* 369 (6500), 208–211. <http://dx.doi.org/10.1126/science.abc3517>.
- Shinde, G.R., Kalamkar, A.B., Mahalle, P.N., Dey, N., Chaki, J., Hassanien, A.E., 2020. Forecasting models for coronavirus disease (COVID-19): A survey of the state-of-the-art. *SN Comput. Sci.* 1 (4), 1–15. <http://dx.doi.org/10.1007/s42979-020-00209-9>.
- Socialstyrelsen, 2020. Avlidna och COVID-19. <https://www.socialstyrelsen.se/statistik-och-data/statistik/statistik-om-covid-19/statistik-over-antal-avlidna-i-covid-19>. (Online; Accessed 09 February 2022).
- Soltész, K., Gustafsson, F., Timpka, T., Jaldén, J., Jidling, C., Heimerson, A., Schön, T.B., Spreco, A., Ekberg, J., Dahlström, Ö., et al., 2020. The effect of interventions on COVID-19. *Nature* 588 (7839), E26–E28. <http://dx.doi.org/10.1038/s41586-020-3025-y>.
- Sveriges Kommuner och Regioner, 2021. Meddelande från styrelsen - Överenskommelse mellan Staten och Sveriges Kommuner och Regioner om ökad nationell testning och smittspårning för COVID-19, 2021. <https://skr.se/download/18.32563d7d1784aa279e298a/1618741812524/WEBB-17-OK-Statens-SKR-Covid-19.pdf>. (Online; accessed 01 April 2022).
- van den Driessche, P., 2017. Reproduction numbers of infectious disease models. *Infect. Dis. Model.* 2 (3), 288–303. <http://dx.doi.org/10.1016/j.idm.2017.06.002>.
- Widgren, S., Engblom, S., Emanuelson, U., Lindberg, A., 2018. Spatio-temporal modelling of verotoxigenic *Escherichia coli* O157 in cattle in Sweden: Exploring options for control. *Vet. Res.* 49 (1), 1–13. <http://dx.doi.org/10.1186/s13567-018-0574-2>.
- Winblad, U., Swenning, A.-K., Spangler, D., 2022. Soft law and individual responsibility: A review of the Swedish policy response to COVID-19. *Health Econ. Policy Law* 17 (1), 48–61. <http://dx.doi.org/10.1017/S1744133121000256>.
- Wood, S.N., 2010. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* 466 (7310), 1102–1104. <http://dx.doi.org/10.1038/nature09319>.
- Wu, J.T., Leung, K., Lam, T.T., Ni, M.Y., Wong, C.K., Peiris, J., Leung, G.M., 2021. Nowcasting epidemics of novel pathogens: Lessons from COVID-19. *Nat. Med.* 27 (3), 388–395. <http://dx.doi.org/10.1038/s41591-021-01278-w>.
- Wu, J.T., Leung, K., Leung, G.M., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *Lancet* 395 (10225), 689–697. [http://dx.doi.org/10.1016/S0140-6736\(20\)30260-9](http://dx.doi.org/10.1016/S0140-6736(20)30260-9).
- Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M.M., Dahly, D.L., Damen, J.A., Debray, T.P., et al., 2020. Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. *BMJ* 369, <http://dx.doi.org/10.1136/bmj.m1328>.