Revealing the Transmission Dynamics of COVID-19: A Bayesian Framework for Rt Estimation

Xian Yang
Hong Kong Baptist University
Shuo Wang
Imperial College London https://orcid.org/0000-0002-2947-8783
Yuting Xing
Imperial College London
Ling Li
University of Kent
Richard Yi Da Xu
University of Technology Sydney
Karl Friston
University College London https://orcid.org/0000-0001-7984-8909
Yike Guo (yikeguo@hkbu.edu.hk)
Hong Kong Baptist University https://orcid.org/0000-0002-3075-2161

Article

Keywords: reproduction number (Rt), transmission dynamics, COVID-19, Bayesian framework, Data Assimilation

Posted Date: March 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-137557/v1

Revealing the Transmission Dynamics of COVID-19: A Bayesian Framework for R_t Estimation

Xian Yang[1,2#], Shuo Wang[2#], Yuting Xing[2], Ling Li[3], Richard Yi Da Xu[4], Karl J. Friston[5] and Yike Guo[1,2*]

[1]Department of Computer Science, Hong Kong Baptist University, Hong Kong Special Administrative Region, China

[2]Data Science Institute, Imperial College London, UK

[3]School of Computing, University of Kent, UK

[4]Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

[5]Institute of Neurology, University College London, UK

[#] Contribute equally [*] Corresponding author: yikeguo@hkbu.edu.hk

ABSTRACT

In epidemiological modelling, the instantaneous reproduction number, R_t, is important to understand the transmission dynamics of infectious diseases. Current R_t estimates often suffer from problems such as lagging, averaging and uncertainties demoting the usefulness of R_t. To address these problems, we propose a new method in the framework of sequential Bayesian inference where a Data Assimilation approach is taken for R_t estimation, resulting in the state-of-the-art 'DARt' system for R_t estimation. With DARt, the problem of time misalignment caused by lagging observations is tackled by incorporating observation delays into the joint inference of infections and R_t; the drawback of averaging is improved by instantaneous updating upon new observations and a model selection mechanism capturing abrupt changes caused by interventions; the uncertainty is quantified and reduced by employing Bayesian smoothing. We validate the performance of DARt through simulations and demonstrate its power in revealing the transmission dynamics of COVID-19.

INTRODUCTION

Epidemic modelling is important for understanding the evolving transmission dynamics and responding to the emerging pandemic[1–8]. In particular, the instantaneous reproduction number R_t has drawn extensive attention and become an essential metric, indicating the trajectory of prevalence[9]. R_t is defined as the average number of secondary cases that could be generated by a primary case, if conditions remained the same thereafter time t. There are two major applications of R_t: as a trend indicator for the nowcast of transmission[10], and as a quantitative metric for the retrospective assessment of intervention impacts[6,11]. Both applications depend on a reliable system estimating the time-varying R_t from appropriate epidemiological observations with accuracy and timeliness. Inappropriate interpretation or imprecise estimation of R_t could furnish misleading information. Several systems [9,12–15] have been proposed to estimate R_t, however, in practice, this remains a challenging task due to the following issues[16]:

a. **Lagging observations.** Given a mathematic model of transmission dynamics, R_t estimation is to infer the time-varying parameters of the model based on epidemiological observations, where the number of infections could be the ideal data source. However, the actual infection number is unknown and can only be inferred from other epidemiological observations (e.g., the daily confirmed cases). Such observations are lagging behind the infection events due to inevitable time delays between an individual being infected and reported (e.g., days for symptom onset[17]). Direct R_t estimation from lagging observations without adjusting for the time delay results in the temporal inaccuracy[16,18]. To address this problem, a two-step strategy, first estimating infections from epidemiological observations with a temporal transformation followed by R_t estimation using epidemic models, has been commonly used in practice[18]. The simple temporal shift of observations by the mean observation delay turns out not sufficient for the relatively long observation delay nor the rapidly changing transmission dynamics, which are seen in the COVID-19 pandemic[18]. Backward convolution method (i.e., subtracting time delay, with a given distribution, from each observation time) leads to an over-smooth reconstruction of infection and bias for R_t estimation[9]. More advanced deconvolution methods[19] through inversing the observation process are mathematically more accurate but sensitive to the optimisation procedure (e.g., stopping criterion) of the ill-posed inverse problem. In addition, the estimated result of infection number is often calculated as a point estimate for the R_t estimation so that the

uncertainty from the observation process is discarded[20]. As an alternative to the two-step strategy, we are investigating a new Bayesian approach of jointly estimating both infection number and R_t with uncertainty from epidemiological observations with explicitly parameterising observation delay.

b. Averaging inference. There are two general paradigms to deal with the challenge of estimating time-varying parameters: 1) reformulating the problem into an inference of static or quasi-static parameters, so that various methods for static parameter estimation can be used; 2) developing inference methods for explicit time-varying parameter estimation. For the first approach, the evolution of R_t is usually parameterised with several static parameters (e.g., the decay rate of an exponential R_t[14]). The quasi-static method is to assume slow R_t evolution that could be treated as static within a short period. For example, Cori et al.[13] proposed a sliding-window method 'EpiEstim' using a segment of observations for averaging inference, with the assumption that R_t keeps the same within the sliding window. This assumption does not apply to the rapidly changing transmission dynamics with interventions, and the window size affects the temporal and quantitative accuracy of R_t estimation. Best practices of selecting the sliding window are still under investigation[18]. Instead of adopting a local sliding window, Flaxman et al.[11] defined several intervention periods according to intervention measures, assuming a constant R_t within each period. This requires additional information about the intervention timeline, which could be not accurate, and the abrupt change is blurred. Although the technical challenges are substantially reduced by inferring static parameters rather than a high-dimensional R_t time series, the fixed-form R_t assumption or the averaging inference is not able to describe the switching transmission dynamics that were common due to the intervention measures imposed during the COVID-19 pandemic. In contrast to window-based methods, data assimilation[21] with sequential Bayesian inference[22,23] is a window-free alternative that has been less explored for R_t estimation. Applying sequential Bayesian inference, data assimilation supports instantaneous updating model states upon the arrival of new observations. Moreover, the Bayesian model selection mechanism[24] can be used for modelling the switching transmission dynamics under interventions, avoiding the drawback of averaging inference. This motivates us to employ the sequential Bayesian approach for R_t estimation.

c. **Uncertainty.** The confidence of R_t estimation is of equal importance compared to the estimate itself, especially when providing evidence for policymaking. The uncertainty of R_t estimates come from different sources, including the intrinsic uncertainties of epidemic modelling, data observation and inference process. Firstly, the uncertainty of epidemiological parameters affects the final R_t estimates. For example, R_t estimation is found sensitive to the distribution of generation time intervals[18]. Secondly, the uncertainty, resulting from systematic error (e.g., weekend misreporting) and random error (e.g., observation noise), in the observation process should be properly measured. During the COVID-19 pandemic, we have seen different reporting standards across countries and regions, with different levels of uncertainty. This is not only the uncertainty of observations, but also that of the time delay between the infection event and the observation that affects R_t estimates. Thirdly, the uncertainty could be enlarged or smoothed in the inference process. For example, the use of a sliding window could smooth the R_t estimation but may simultaneously miscalculate the uncertainty, due to the overfitting within the sliding-window. To provide a reliable credible interval (CI) of R_t estimates, the aforementioned three types of uncertainty should all be handled and reported as part of the final R_t estimates. As a state-of-the-art package, EpiEstim (Version 2)[25] allows users to account for the uncertainty from epidemiological parameters by resampling over a range of plausible values. However, the uncertainty from imperfect observations and the side effects associated with the sliding window is not addressed. Recently, 'EpiNow'[15] was proposed to integrate the uncertainty of observation process, but the inference is still based on the sliding window. In this work, we deal with model and data uncertainty in the data assimilation framework[21] with a Bayesian smoothing mechanism to enable both the latest and historical observations to be continuously integrated into R_t estimation to alleviate spurious variability.

In order to tackle these practical issues, we propose a comprehensive Bayesian data assimilation system, namely 'DARt' (Data Assimilation for R_t estimation), for joint estimates of infections and R_t together with their uncertainty. The evolution of the transmission dynamics is described by a hierarchical transition process, which is informed by the newly observed data through an observation process formulated with explicit observation delay. A model selection mechanism is built in the transition process to detect abrupt changes under interventions. The performance of the DARt system is validated and compared to the state-of-the-art EpiEstim system through simulations and real-world applications, showing its power of

estimation and adequacy for practical use. We have made the system available online for broad use in R_t estimation for both research and policy assessment.

Results

1. DARt: A Data Assimilation System for R_t Estimation

With DARt, we jointly estimate the number of incident infections (j_t), the instantaneous reproduction number (R_t), and the change indicator (M_t) of the switching dynamics of R_t. This is achieved by the inference of the latent state $\boldsymbol{X}_t$, consisting of R_t, j_t and M_t, with all available observations $C_{1:T}$ up to the latest observation time T within a data assimilation framework (see Methods). In contrast to inferring the 'pseudo' dynamics (i.e., reformulating into a static/quasi-static problem), our proposed DARt system directly estimates the 'real' time-varying R_t by assimilating information from the observation with the model forecast.

As a data assimilation system, DARt consists of three main components: 1) a **state transition model** – describing the evolution of latent states based on an epidemic renewal process; 2) an **observation function** – linking the underlying latent states to lagging epidemiological observations; and 3) a **sequential Bayesian inference engine**, as depicted in Figure 1 – updating the state estimation with ongoing observations using *forward Bayesian filtering* (Figure 1 (A)) and refining historical state estimates using all observations with *backward Bayesian smoothing* (Figure 1 (B)).

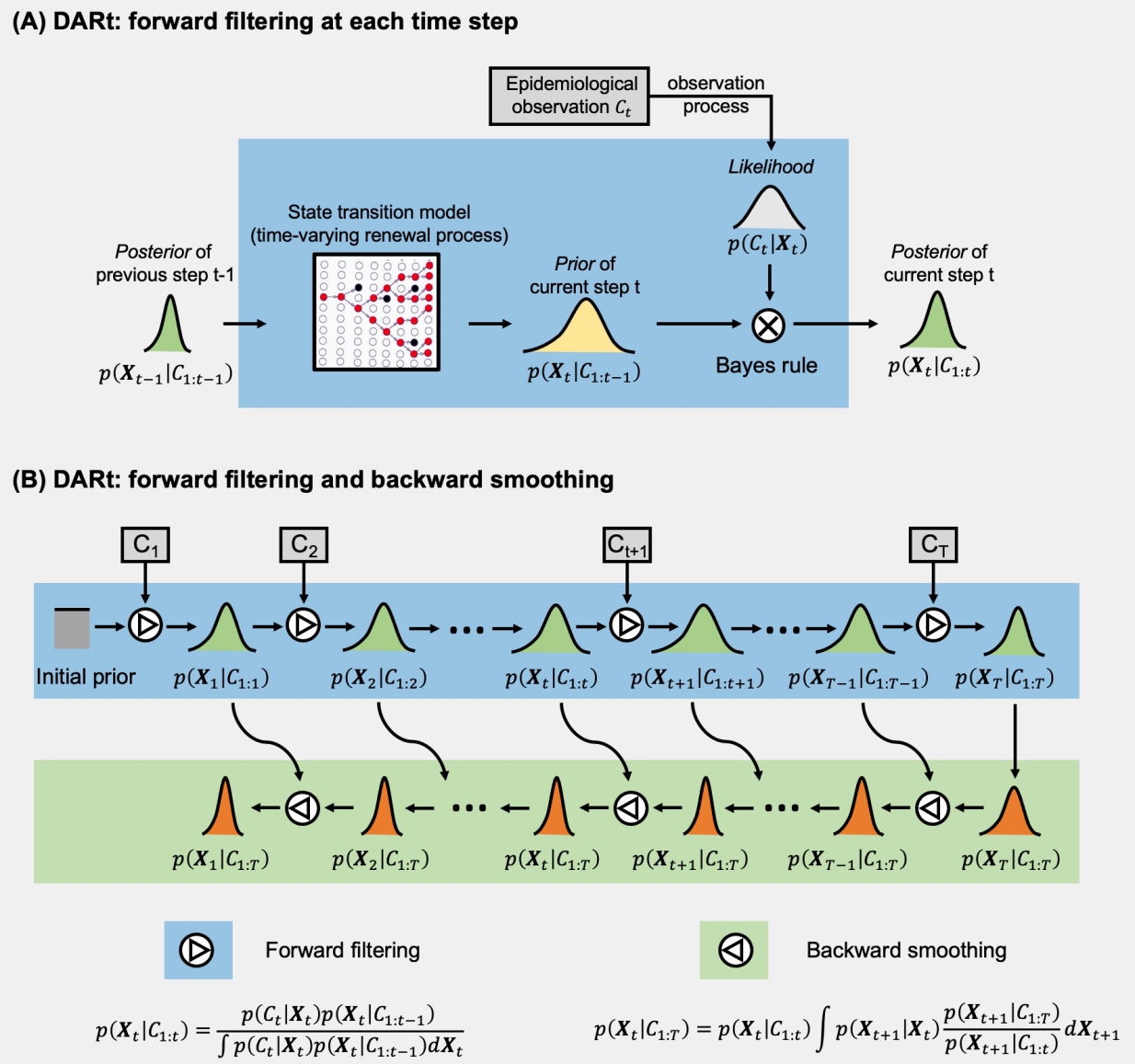


Figure 1. Illustration of the inference of DARt system for R_t estimation. The latent state $\boldsymbol{X}_t$ includes the number of infections (j_t), the instantaneous reproduction number (R_t), and the change indicator (M_t) of abrupt change. The epidemiological observation is denoted as C_t, and is linked to the latent state via the observation function. For each time step, the estimation of the latent state $p(\boldsymbol{X}_t|C_{1:t})$ are constantly updated according to ongoing reported observations using sequential Bayesian updating with forward filtering and backward smoothing. **(A) Forward filtering at each time step**. The posterior state estimation $p(\boldsymbol{X}_{t-1}|C_{1:t-1})$ estimated from previous step $t-1$ is transformed as the *prior* $p(\boldsymbol{X}_t|C_{1:t-1})$ for the current step t, where a time-varying renewal process is used as the state transition model as detailed in Figure 7. Together with the *likelihood* $p(C_t|\boldsymbol{X}_t)$ obtained from epidemiological observation at the current step, the *posterior* of the current

step $p(\boldsymbol{X}_t|C_{1:t})$ is estimated. At the same time, as shown in **(B), backward smoothing** is to compute $\{p(\boldsymbol{X}_t|C_{1:T})\}_{t=1}^{T}$, taking account of all the observations $C_{1:T}$ up to the time T by applying a Bayesian smoothing method (see Methods).

- **State transition model**

As shown in Figure 1(A), the latent state $\boldsymbol{X}_t$ is updated with a state transition model based on a time-varying renewal process[26,27]. Besides R_t, j_t, we introduce an auxiliary binary latent variable M_t to indicate the switching dynamics of R_t under interventions without assuming a pre-defined evolution pattern of R_t (e.g., constant or exponential decay). $M_t = 0$ indicates a smooth evolution of R_t corresponding to minimal or consistent interventions; $M_t = 1$ indicates an abrupt change of R_t corresponding to new interventions or outbreak. The smooth evolution of R_t is modelled as a Gaussian random walk while the abrupt change is implemented through resetting the R_t memory by assuming a uniform probability distribution for the next time step of estimation. That provides an automatic way of framing a new epidemic period as manually done in Ref[11]. The transition of M_t is modelled as a discrete Markovian process with fixed transition probabilities controlling the sensitivity of change detection. Compared with the widely used two-step strategy (i.e., j_t and R_t are estimated successively), the joint inference of R_t and j_t provides a comprehensive picture of the evolving epidemics (see details in Methods).

- **Observation function**

An observation function is designed as a convolution process, reflecting the fact that observation C_t at time t is linked with a series of infection numbers of lagging days by a time delay distribution: $C_t = \sum_k \varphi_k j_{t-k}$ where the kernel φ_k is the probability that an individual infected is observed after k days (see details in Methods). According to the observation types, different observation delays such as incubation time, testing and reporting delay can be chosen by selecting the corresponding kernel φ_k. Therefore, through the observation function, the lagging relationship between the latent state and observations are explicitly modelled to handle the issue of temporal inaccuracy. This makes our R_t estimates aligned with the timeline, which is important for the retrospective analysis of policy impacts.

- **Sequential Bayesian inference engine**

In DARt, the Bayesian inference has two phases: **forward filtering** and **backward smoothing**. The forward filtering uses the up-to-date prior from the state transition model and the likelihood determined by the latest observation to update the current latent state, by computing its posterior following the Bayes rule. The backward smoothing works by looking back to refine the previous state estimation when more observations have been accumulated to reduce the uncertainty of R_t estimation.

Forward filtering. For each new time step t, the latent state $\boldsymbol{X}_t$ is continuously updated upon the new observation C_t, following Bayes rule: $p(\boldsymbol{X}_t|C_{1:t}) \propto p(\boldsymbol{X}_t|C_{1:t-1})\, p(C_t|\boldsymbol{X}_t)$, where $p(\boldsymbol{X}_t|C_{1:t-1}) = \int p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}) p(\boldsymbol{X}_{t-1}|C_{1:t-1})\, d\boldsymbol{X}_{t-1}$ is the prior forecasted by the state transition model from the previous step $t-1$, and $p(C_t|\boldsymbol{X}_t)$ is the likelihood determined by the new observation C_t under the forecasted state $\boldsymbol{X}_t$. This approach differs from the fixed prior in the Bayesian inference of static parameters (e.g., the fixed gamma distribution adopted in 'EpiEstim'). This filtering mechanism, as shown in Figure 1 (A), computes the posterior distribution of the latent state by assimilating the forecast from the forward transition model with the information from the new epidemiological observations. For the implementation of this Bayesian updating process, we adopt a particle filter method[23] to efficiently approximate the posterior distribution through Sequential Monto Carlo (SMC) sampling (see details in Method). This eschews any fixed-form assumptions for the posterior – of the sort used in variational filtering and dynamic causal modelling[28].

Backward smoothing. The estimated result $p(\boldsymbol{X}_t|C_{1:t})$ from aforementioned forward filtering only includes the past and present information flows, corresponding to the prior $p(\boldsymbol{X}_t|C_{1:t-1})$ and likelihood $p(C_t|\boldsymbol{X}_t)$, respectively. The R_t estimation would be accurate if all infections related to R_t are fully observed in $C_{1:t}$. However, this is certainly not the case due to observation delay. As a result, the R_t estimate from forward filtering is with large uncertainty inherited from the likelihood where the present observations include little up-to-date infection information. In order to reduce the uncertainty in R_t estimation from forward filtering, DARt adopts the Bayesian backward smoothing technique, estimating the latent state at a time t retrospectively, given all observations available till time T ($T > t$). Compared with other R_t estimation methods, DARt takes the advantage of additional information to smooth inference results with reduced uncertainty caused by incomplete observations. More specifically, the

smoothing mechanism can be described as: given a sequence of observations $C_{1:T}$ up to time T and filtering results $p(\boldsymbol{X}_t|C_{1:t})$, for all time $t < T$, the state estimates are smoothed as: $p(\boldsymbol{X}_t|C_{1:T}) = p(\boldsymbol{X}_t|C_{1:t}) \int p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) \frac{p(\boldsymbol{X}_{t+1}|C_{1:T})}{p(\boldsymbol{X}_{t+1}|C_{1:t})} d\boldsymbol{X}_{t+1}$, where $p(\boldsymbol{X}_{t+1}|C_{1:T})$ is the smoothing results at time $t+1$ where $\int p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) \frac{p(\boldsymbol{X}_{t+1}|C_{1:T})}{p(\boldsymbol{X}_{t+1}|C_{1:t})} d\boldsymbol{X}_{t+1}$ is the smoothing factor, as shown in Figure 1(B). In this way, all the relevant observations are fully exploited to enable us to reduce the uncertainty of R_t estimation.

In summary, comparing with the sliding-window (i.e., averaging inference) approaches, our sequential Bayesian updating mechanism of DARt features an instantaneous R_t estimation and smoothing uncertainty through the utilisation of all available observations.

2. Validation through simulation

Due to the lack of ground-truth R_t in real-world epidemics, we present a set of experiments based on synthetic data to establish the face validity of DARt system. Figure 2 illustrates the design of simulation experiments where a synthesised R_t, is adopted as the ground truth to validate its estimated $\hat{R}_t$. We also estimated R_t using the state-of-the-art R_t estimation package EpiEstim[25] to compare the effectiveness in overcoming three aforementioned issues (i.e., lagging, averaging and uncertainty).

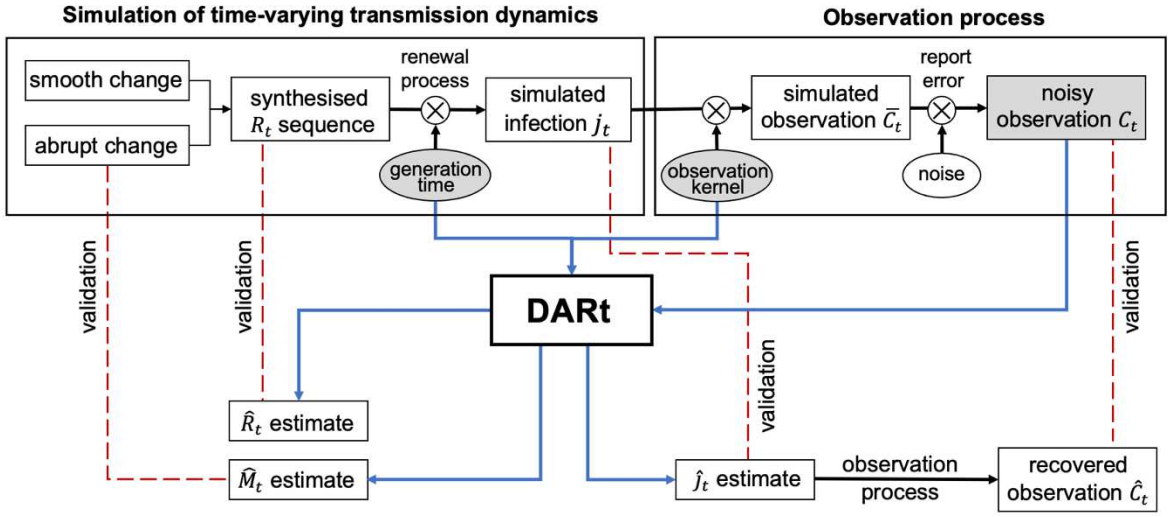


Figure 2. Validation experiment of the DARt system on simulated data. First, the ground-truth R_t sequence is synthesised using piecewise Gaussian random walk split by two

abrupt change points, simulating the drop of mean R_t from 3.2 to 1.6 and then 0.8 under two intensive interventions. The sequence of incident infection j_t is simulated based on a renewal process parameterised by the synthesised R_t and starting with 5 infections on the first day. The observation process includes applying a convolution kernel representing the probabilistic observation delay to obtain the expectation of observation $\bar{C}_t$ and adding Gaussian noise representing the reporting error to obtain the noisy 'real' observation C_t. The inputs (in grey) to the DARt system are the distributions of generation time, observation kernel and simulated noisy observation C_t. The system outputs are the estimated $\hat{R}_t$, estimated $\hat{j}_t$ and change indicator $\hat{M}_t$. These outputs are compared with the synthesised R_t, j_t and the time of abrupt changes. Also, the observation function is applied to the estimated $\hat{j}_t$ to compute the recovered observation $\hat{C}_t$ with uncertainty, which is compared to the 'real' observation. This provides an indirect way to validate the correctness of inference results. These comparisons results are shown in Figure 3.

Data synthesis. We first generated a synthesised R_t curve from a piece-wise Gaussian random walk mimicking the scenario of two successive interventions (Figure 3 (A)). To approximate the early stage of exponential growth, the simulation started with $R_0 = 3.2$, which reflects the basic reproduction number of COVID-19 and followed a Gaussian random walk $R_{t+1} \sim Gaussian(R_t, (0.05)^2)$. At $t = 20$, we set R_{20}=1.6 indicating the mitigation outcome of soft interventions. After soft interventions, the epidemic is still being uncontrolled with the evolution of R_t resuming to the Gaussian random walk as above. At $t = 30$, R_t experienced another abrupt decrease to a value under 1, where we set R_{30}=0.8 to indicate the suppression effects of intensive interventions (e.g., lockdown). Afterwards, the epidemic is being controlled, and the evolution of R_t follows the same random walk as above.

With the simulated R_t curve, we followed the renewal process using the generation time distribution as reported by Ferretti et al.[3] (i.e., the Weibull distribution with shape and scale equal to 2.826 and 5.665 days respectively) to simulate the infected curve j_t starting with 5 infections at $t = 0$. Then we generated the lagging observation curve of onset cases $\bar{C}_t$ using the incubation time distribution[3] (i.e., the lognormal distribution with log mean and standard deviation of 1.644 and 0.363 days respectively) as the observation time delay. Throughout this paper, both the generation time distribution and the incubation time distribution are truncated and normalised, where values smaller than 0.1 are discarded. To simulate the real-world noisy observations, we added Gaussian noise as $C_t \sim Gaussian(\bar{C}_t, 9\bar{C}_t)$.

Validation procedure and results. As shown in Figure 2, DARt takes inputs from simulated observation C_t, the distribution of generation time and observation delay, and produces the estimated outputs ($\hat{R}_t$, daily infection $\hat{\jmath}_t$ and the change indicator $\widehat{M}_t$), which are validated via comparing with the synthesised R_t, j_t, and the dates of abrupt change, respectively. We also obtained the recovered observation $\hat{C}_t$ (i.e., the observation result of estimated daily infection $\hat{\jmath}_t$) and checked its consistency with the 'real' observation C_t.

To compare with the state-of-the-art EpiEstim package[25], we input the infection curve and the distribution generation time to EpiEstim to estimate R_t with a 7-day sliding window following the recommended practice. It is noted that two approaches of preparing infection curve input for EpiEstim were taken as common practice: 1) 'plug-and-play' use[1] by taking observations C_t as the infection without adjusting for the observation delay; 2) the two-step strategy[29] that shift C_t backwards in time by the median observation delay (5 days in the simulation). We implemented both practices and compared with DARt.

All validation results are shown in Figure 3.

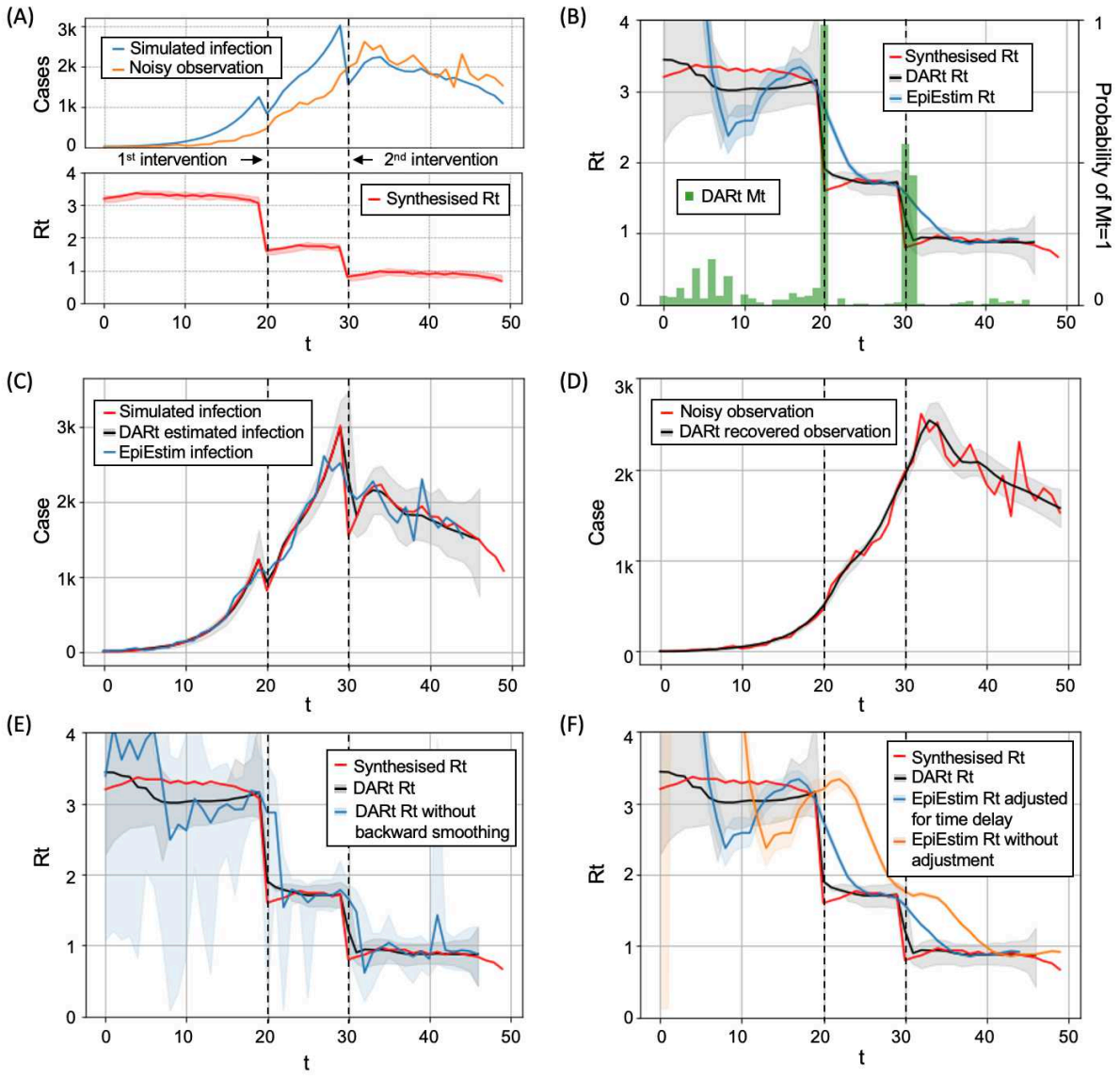


Figure 3. Validation results. (A) Synthesised R_t, simulated j_t and C_t curves from the evolution of transmission dynamics under two successive interventions. Validation results are listed as follows. **Correctness of R_t estimation: (B)** shows the comparison of the synthesised R_t with the DARt estimated R_t, as well as and that of the EpiEstim with backward shifted C_t as its infection. The comparison indicates that DARt is more robust to observation noise and the estimated R_t matches the synthesised R_t better with relatively less degree of fluctuations. Moreover, DARt can respond to abrupt changes instantaneously demonstrating its ability to overcome the weakness of averaging. The probabilities of having abrupt changes are captured by M_t as shown in green bars. Even under observation noise, DARt can still acutely detect abrupt changes. **Correctness of j_t estimation**: **(C)** shows the simulated j_t , DARt estimated j_t, and j_t of EpiEstim. We

can find that the DARt estimated j_t with 95% CI match well the simulated j_t. In contrast, the EpiEstim result diverts from j_t, indicating that, with the shifting the observation curve to approximate infection numbers, EpiEstim results still suffer from the inaccuracy of approximation. **Accuracy in recovering observations C_t**: **(D)** compares the distribution of recovered C_t from DARt with the simulated C_t curve. The recovered C_t with 95% CI can generally match well the simulated C_t without undesired local fluctuations. **Effectiveness of DARt smoothing**: **(E)** illustrates the effectiveness of backward smoothing by comparing the DARt estimated R_t results with and without smoothing, showing the expected smoothing effect of estimated R_t with reduced CI. **Temporal accuracy of DARt**: **(F)** shows the results from EpiEstim with and without shifting the observation curve backwards. Although the time misalignment problem can be alleviated by temporal shifting in EpiEstim, its inferred R_t cannot match well to the synthesised R_t curve, failing to reflect two abrupt changes in a timely fashion.

3. Applicability to real-world data

We applied DARt to estimate R_t in four different regions during the emerging pandemic. Each region represents a distinct epidemic dynamic, allowing us to test the effectiveness and robustness of DARt in each scenario. 1) Wuhan: When COVID-19 had the outbreak in the city, the government responded with very stringent interventions such as a total lockdown. By studying the R_t evolution there during the early few months of this year, we can check the capability of the system in detecting the expected abrupt change of R_t. 2) Hong Kong: during the year, the daily increase of reported cases in Hong Kong has been remaining at a low level for most of the time with the maximum daily value under 200. As no stringent interventions have been introduced in such a city with high-density population, this is an ideal scenario of studying smooth change of R_t with the risk of sudden resurgence. 3) United Kingdom: UK is undergoing an intense secondary outbreak of COVID-19 with daily infections exceeding 10,000. Instantaneous R_t estimation is very meaningful for this scenario for examining the impacts of various proposed policies. We also applied DARt to estimate the distinct epidemic curves of typical cities in England to reflect the local epidemic dynamics (see Supplementary Materials). 4) Sweden: Sweden is a representative of countries that have less stringent intervention policies and has a clear miss-reporting pattern repeated weekly. This makes Sweden an ideal case to examine the robustness of DARt with considerable observation noise.

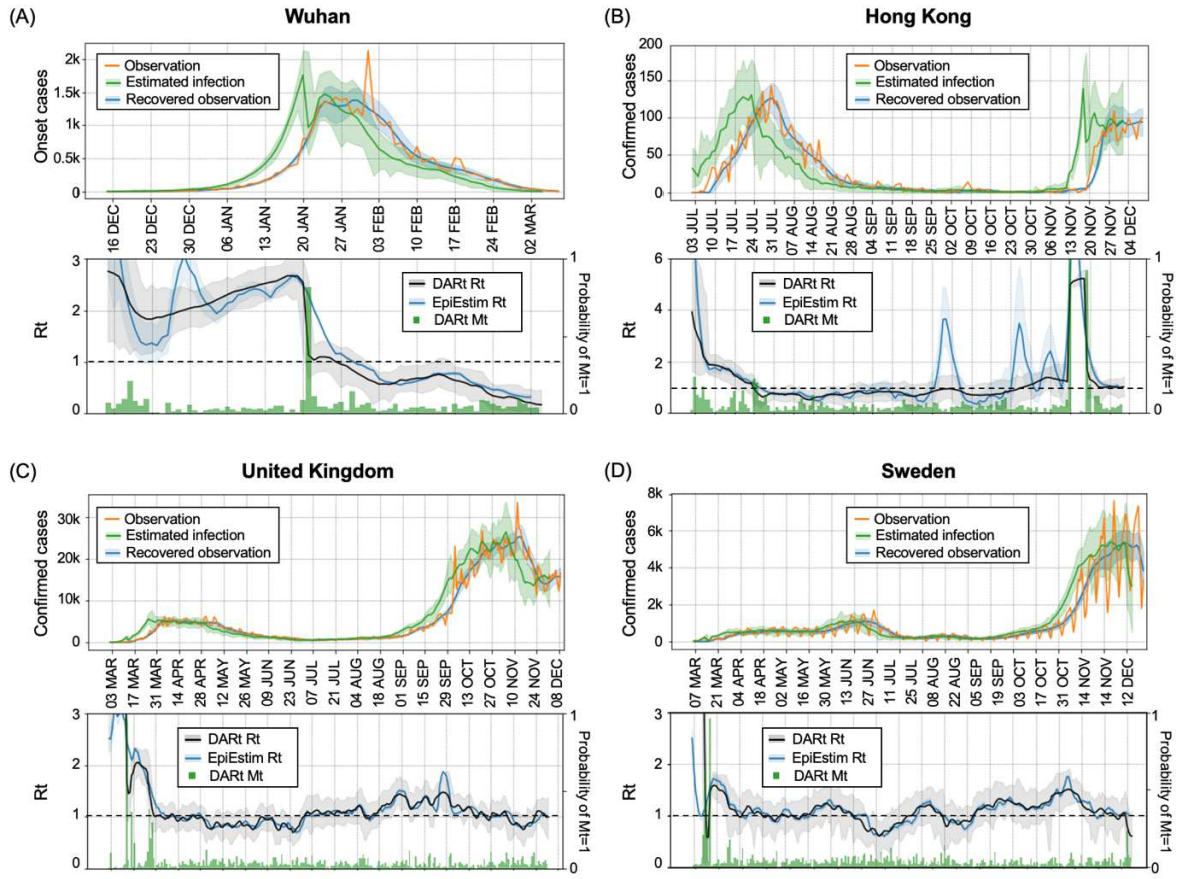


Figure 4. Epidemic dynamics in Wuhan **(A)**, Hong Kong **(B)**, United Kingdom **(C)** and Sweden **(D).** The top row of each subplot shows the number of daily observations (in yellow), the recovered daily observations (in blue) and the recovered daily infections (in green). The consistency between the yellow line and blue line indirectly validated the soundness of R_t results. The bottom row compares the DARt R_t estimation (in black) with the EpiEstim results (in blue). $R_t = 1$ is plotted in dash line for reference. The distributions of all estimates from both DARt and EpiEstim are with 95% CI. The probabilities of having abrupted changes (M_t=1) are shown in green bars. For Wuhan, in contrast with EpiEstim having a smooth decrease in R_t around the lockdown time, the R_t curve from DARt experienced a sharp decrease showing the instantaneous response to the intervention, which is also detected by M_t. For Hong Kong, DARt can generate a more stable R_t than EpiEstim, especially during September and October when the number of cases remained at a very low level. For the UK and Sweden, R_t curves from DARt and EpiEstim largely follow the same trend. However, CIs of R_t from EpiEstim are quite narrow (even cannot be easily observed from this figure), indicating the underestimated uncertainty of estimation by EpiEstim.

Epidemic dynamics in these four regions. The inference results for R_t and reconstructed observations for these four countries are shown in Figure 4. For Wuhan, the observation data are the number of onset cases compiled retrospectively from epidemic surveys, while for Hong Kong, UK and Sweden, the observation data are the number of reported confirmed cases. Throughout this paper, we use the onset-to-confirmed delay distribution from Ref[30] together with the distribution of incubation time proposed in Ref[3] to approximate the observation delay. As the ground-truth R_t is not available, we validate the results by checking whether the recovered distributions of observation well cover the observation curve of C_t. As shown in the top panel of each subplot in Figure 4, the CIs of recovered C_t distributions (in blue) covered most parts of the original observations (in yellow), showing the reliability of our R_t estimation.

Figure 4 (A) shows the results using Wuhan's onset data[1]. We observe that there was a sharp decrease in Wuhan's R_t after 21st of Jan 2020, which is also indicated by M_t as the probability of abrupt changes peaked at this time (in green bars). Around that time, a strict lockdown intervention has been enforced in Wuhan. This sharp decrease is likely to be the result of this intervention, indicating its impact. The small offset between the exact lockdown date and the time of sharp decrease might be due to noisy onset observations and approximated incubation time distribution. After lockdown, R_t decreased smoothly, indicating that people's awareness of the disease and the precaution measures taken had made an impact. Since the beginning of Feb 2020, the value of R_t remained below 1 for most of the time with the enforcement of quarantine policy and increases in hospital beds to accept of all then diagnosed patients. It is noted that the onset curve has a peak on 1st of Feb 2020, due to a major correction in reporting standard. Neither R_t nor j_t curve from our model suffers from this fluctuation, showing its robustness delivered by the smoothing mechanism. The results from Wuhan show our switching mechanism can overcome the issue of averaging and automatically detect sharp changes in epidemiological dynamics.

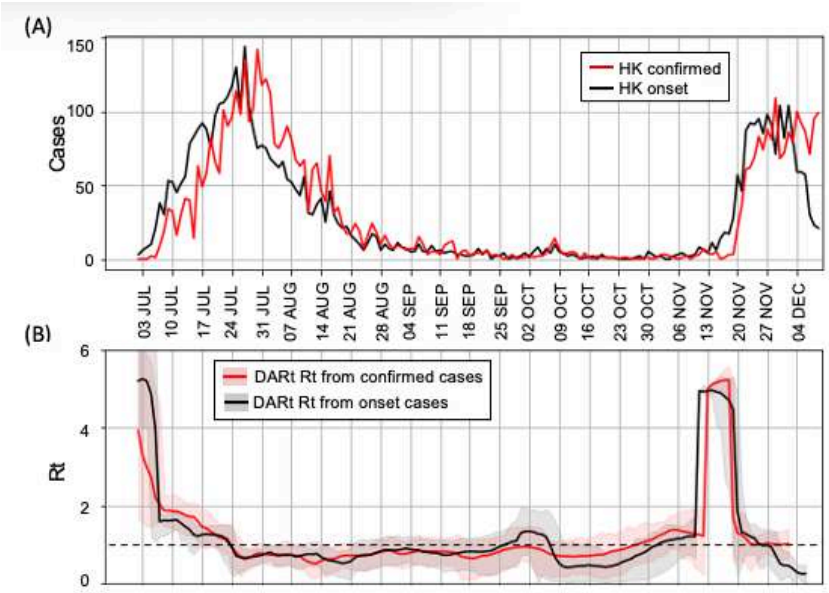
Figure 4 (B) shows the inferred results from Hong Kong reported confirmed cases[31] since early July. In Hong Kong, the number of infections remains low for the most time and the government has continuously imposed soft interventions: for example, on 11th of July, the Hong Kong government forbidden social gatherings with > 8 people in restaurants and pubs; from 24 July, all citizens are required to wear masks in indoor places. We can see that the R_t curve has remained largely under 1 until November. In the middle of November 2020, a dramatic increase in R_t emerged indicating a new wave of outbreak. By looking at the

probabilities represented in green bars of Figure 4 (B), we can find that the start of the new wave can be also detected by M_t. The change in R_t we have found is consistent with the official reports about a newly imported case has triggered a new outbreak.

Figure 4 (C) shows the inference results from the United Kingdom's reported confirmed cases[32]. After a continuous decline during May and June 2020, the second wave of COVID-19 cases is evidenced. Around the end of June, R_t started climbing above 1 for the subsequent dates and reached almost 1.5 at the end of August. At the beginning of September, shortly after the reopening of schools at full capacity, a new wave of epidemic outbreak occurred. Then, a series of interventions were introduced from late September and a national lockdown has enforced on 5th of November, where R_t curves kept decreasing and became lower than 1 after the lockdown. With the end of lockdown on 2nd of December, the R_t was again around 1. The match between the time of changes in R_t and the dates of interventions indicates the effectiveness of DARt in policy impact monitoring. Moreover, there was a sharp increase in the number of reported confirmed cases on 4th of October 2020 caused by miss reporting between September 25 and October 2. The estimated R_t didn't experience a spike at that time, indicating the robustness of DARt at the presence of observation noise.

Figure 4 (D) shows the inferred epidemic dynamics in Sweden from the daily reported data[33]. We find that the daily reported cases in Sweden have shown periodic drop every 7 days that are likely caused by misreporting. By looking at the results, we can find that the influence of such periodic fluctuations has been smoothed by DARt to provide a consistent R_t curve.

Comparing DARt performance with different types of observations. Figure 5 shows the results of R_t estimation using onsets and confirmed cases (Figure 5 (A)) as observations in Hong Kong to estimate R_t, respectively. We choose Hong Kong for illustration purpose since both onset and reported confirmed cases are publicly accessible. The CIs of inferred results from two different observations largely overlapped (Figure 5 (B)). For the emerging outbreak started in mid-November, as the record of onset cases is not completely documented by the government yet (especially for the recent week) and the tails of onset and confirmed curves in Figure 5 (A) are different. This result reflects that, with a proper observation kernel according to the observation types (e.g., onset, confirmed cases), DARt can make consistent estimation from a wide range of observation types.



401

402 **Figure 5.** Comparison of estimated R_t curves of Hong Kong using different observations.
403 **(A)** Observations of daily onset (black) and confirmed cases (red) used as the input are
404 compared in. **(B)** R_t estimations with 95% CI inferred from onsets and confirmed cases
405 using DARt are in black and red, respectively. $R_t = 1$ is plotted in dash line for reference.

406 **Sensitivity to systematic observation errors.** For Sweden, we observe a systematic
407 observation error that – on every Monday – the number of reported cases is significantly smaller
408 than the other days. This kind of observation noise could induce unnecessary fluctuations to R_t
409 curves. Therefore, we used Sweden's data to further illustrate the robustness of our scheme in
410 the presence of noise. The blue curves in Figure 6 (A), (B) and (C) are the results from EpiEstim
411 with the sliding-window size set to 7, 5 and 3 days, respectively. For comparison, the estimated
412 R_t curve from DARt is shown as the black curve in each subplot. We found that EpiEstim is
413 sensitive to the choice of window size in this case. In contrast, DARt does not require a pre-
414 setting of window size. To illustrate the robustness of our scheme to observation noise, we made
415 imputed reported numbers on all Mondays by the average of their neighbouring values. The
416 inference results are shown in Figure 6 (D), where the green curve is from the imputed
417 observations with systematic observation errors alleviated and the black curve is the results
418 directly using the raw observations. The results of DARt between using the raw observation
419 and imputed observation are almost indistinguishable, indicating its robustness at the presence
420 of observation noise.

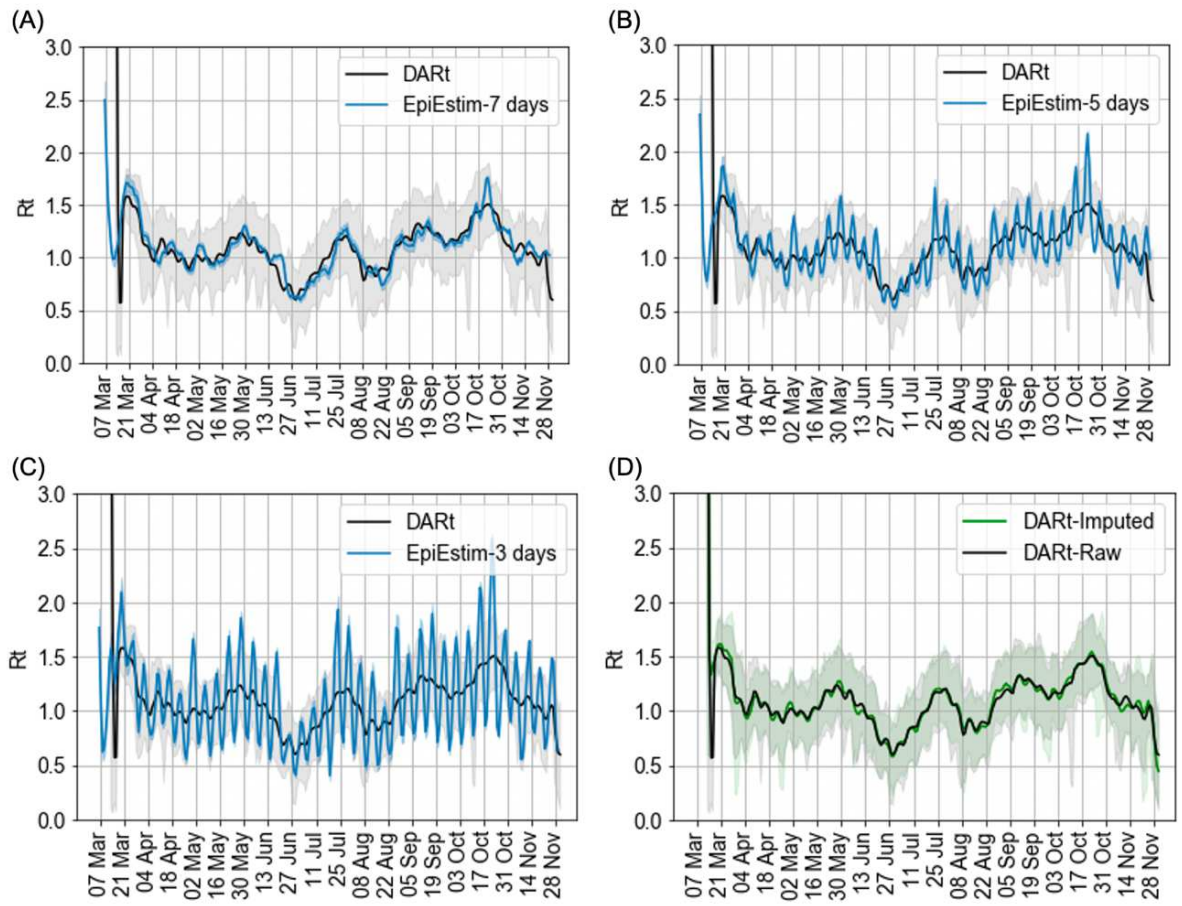


Figure 6. Inference results for Sweden with different settings. **(A)-(C)** compare results from DARt (black) and EpiEstim (blue) with sliding window ranging from 7, 5 and 3 days, respectively. These subplots show EpiEstim is sensitive to the setting of sliding window width. **(D)** compares the DARt results using raw reported data (black) with using imputed reported data (green). The similar results of DARt between using the raw observation and imputed observation as the input indicate the robustness of DARt at the presence of observation noise.

To summarise, DARt has been applied to four different regions for revealing the transmission dynamics of COVID-19 to demonstrate its real-world applicability and effectiveness. Consistent with the findings in the simulation study, DARt has shown its advantage over the popular EpiEstim in the following aspects:

- Instantaneity: Unlike EpiEstim which is sensitive to the width of the sliding window, DARt is window-free following the sequential Bayesian inference approach. DARt

equips the change indicator to detect the sharp change R_t, while EpiEstim is difficult to reflect the abrupt epidemic change due to its averaging mechanism.

- Robustness: With Bayesian smoothing, the R_t curve from DARt is more stable at the presence of observation noise, comparing with that of EpiEstim. Moreover, during the period of low infection numbers (such as the period at the beginning of October in Hong Kong), EpiEstim might generate pseudo peaks in R_t while DARt can avoid such fluctuation.

- Temporal accuracy: EpiEstim requires manually shifting the observation curve backwards by the median observation delay for R_t estimation, while DARt performed a joint estimation of R_t and j_t by explicitly encoding the lag into observation kernels. This alleviates the bias of shifting and inappropriate settings and interpretation from the 'plug-and-play' use of EpiEstim by non-experts.

Discussion and Conclusion

In this paper, we have proposed DARt by adopting a Bayesian inference scheme for estimating R_t. Our work provides a state-of-the-art R_t estimation tool supporting a wide range of observations. In the system, epidemic states can therefore be updated using newly observed data, following a data assimilation process in the framework of sequential Bayesian belief updating. For the model inference, a particle filtering/smoothing method is used to approximate the R_t distribution in both forward and backward directions of time, ensuring the R_t at each time step assimilates information from all time points. By taking the Bayesian approach, we have emphasised the uncertainty in R_t estimation by accommodating observation uncertainty in likelihood mapping and introduced Bayesian smoothing to incorporate sufficient information from observations. Our method provides a smooth R_t curve together with its posterior distribution. We have demonstrated that inferred R_t curves can explain different observations accurately. Our work is not only important in revealing the epidemical dynamics but also useful in assessing the impact of interventions. The sequential inference mechanism of R_t estimation takes into account the accuracy of time alignment and provides an abrupt change indicator. This method offers a promising method for intervention assessment,

comparing the approach of directly incorporating interventions as co-factors into epidemic model[11,34].

We have made some approximations to facilitate the implementation. First, the observation time and generation time distributions are truncated into fixed and identical length. Theoretically, these two distributions can be of any length, while most values are quite small. In our state transition model, one variable of the latent state is a vectorised form of infection numbers over a period. The purpose of vectorisation is to facilitate implementation by making the transition process to be Markovian. The length of this vector variable is determined by the length of effective observation time and the generation time distributions. Truncating these two distributions to a limited length, by discarding small values, would facilitate the vectorisation. Apart from truncation, we have assumed that these two distributions do not change during the prevalence of disease. However, as we have discussed previously[20], introducing interventions, such as an increased testing capacity, would affect the observation time. The distribution of generation time would also change as the virus is evolving. It is possible to extend our model with time-varying observation function. For example, the testing capability and time-varying mortality rate could also be considered in the observation process.

Second, we approximate the variance of observation error empirically. Given the variance of observations is unknown and could change over time across different regions, the standard deviation of the Gaussian likelihood function is not set to a fixed value in our scheme. We estimate the region-specific time-varying observation variance from the observational data. Although the empirical estimation yielded reasonable results for the four regions and cities in the UK (see Supplementary Materials), it may generate some implausible results in some scenarios; for example, when the epidemic is growing or resurging explosively, leading to an overestimation of observation variance. An adaptive error variance inference should be developed to tackle this issue.

The third approximation is implicit in the use of a particle filter to approximate the posterior distributions over model state variables – including R_t – with a limited number of samples (i.e., particles). Particle filtering makes no assumptions about the form of posterior distributions. On the contrary, the variational equivalent of particle filter, namely variational filtering[28] provides an analytical approximation to the posterior probability and can be regarded as limiting solutions to an idealised particle filter, with an infinite number of particles[35]. As not only the

mean value of R_t but also its estimation uncertainty is important for advising governments on policymaking, an analytical approximation to help properly quantify uncertainty is desirable.

Fourthly, change detection is approximated by the change indicator M_t. The change indicator is included as part of the latent state and inferred during particle filtering. This work opens a venue to explore variational Bayesian inference for switching state models[36]. Crucially, variational procedures enable us to assess model evidence (a.k.a. marginal likelihood) and hence allow automatic model selection. Examples of Variational Bayes and model comparison to optimise the parameters and structure of epidemic models can be found in previous studies[37]. These variational procedures can be effectively applied to change detection.

Finally, the method outlined in the paper can, in principle, be applied to generative epidemic models that include more latent states that underwrite the renewal process; for example, contact rates, transmission strengths, et cetera. We envisage that such models would be considered from observational, spatial-temporal and social perspectives. From the observational aspect, multiple epidemic curves are generally available (e.g., daily onsets, deaths or confirmed cases). This allows using different kinds of data to inform model parameters (and structure). This sort of modelling may call for a generative model that explicitly includes the latent states generating the data at hand (e.g., hospital admissions). Dynamic causal models[37] are potential candidates here because they extend conventional (SEIR) models to include spatial location, mobility, hospitalisation et cetera. From the spatial-temporal perspective, one could construct a homogeneously mixed spatial-temporal model with connected regions that share the same model structure but with distinct model parameters[38]. Mobility information could then be used to inform inter-regional spread, when suitably parameterised. From the socio-behavioural aspect, one could build a comprehensive model by including epidemic-relevant behavioural factors into the model, especially human mobility trends. This usually entails modelling differential contact rates between subpopulations (or populations in specific locations) as a function of other latent states: for example, modelling social distancing as a non-linear function of the prevalence of infection[39]. Mobility is reflected in the use of public transportation, people's average walking distance, people's attitude towards disease (cautious or passive), and people's lifestyle (e.g., work from home, take-away). All of these metrics are, to a greater or lesser extent, available as empirical constraints on suitably structured generative models. By considering different modelling factors, the estimation results of R_t should be more accurate and furnish more precise credible intervals.

In conclusion, our work provides a practical scheme for accurate and robust R_t estimation. It opens a new avenue to study epidemic dynamics within the Bayesian framework. We provide an open-source R_t estimation package as well as an associated Web service that may facilitate other people's research in computational epidemiology and the practical use for policy development and impact assessment.

Methods

The DARt system estimates R_t in the framework of data assimilation with a sequential Bayesian inference approach. Three main components of DARt: 1) a **state transition model** - describing the evolution of the latent state; 2) an **observation function** – defining an observation process describing the relationship between the latent state and observations; 3) a **sequential Bayesian engine**: statistical reason time-varying model parameters with uncertainty by assimilating prior state information provided by the transition model and the newly available observation. In this section, we first introduce the renewal process for modelling epidemic dynamics, which is the fundamental of our state transition model. We then describe the observation function, linking a sequence of infection numbers with the observation data. Next, we present a detailed state transition model and propose the sequential Bayesian update module.

1. Renewal process for modelling epidemic dynamics

Common R_t estimation methods include compartment model-based methods (e.g., SIR and SEIR [40]) and time-since-infection models based on renewal process[25]. Their relationships are discussed in Supplementary Materials Section 1.1. Comparative studies have been done in Ref[18] to show that EpiEstim, one of the renewal process-based methods, outperforms other methods in terms of accuracy and timeliness. Given the renewal process, the key transition equation derived from the process is:

$$j_t = R_t \sum_{k=1}^{T_w} w_k \, j_{t-k} \tag{1}$$

where j_t is the number of incident infections on day t, T_w is the time span of the set $\{w_k\}$, and individual w_k is the probability that the secondary infection case occurs k days after the primary infection, describing the distribution of generation time[9]. The profile of w_k is related to the biological characteristics of the virus and is generally assumed time-independent during the epidemic. Considering the simplicity and performance of applying the renewal process to model epidemic dynamics, our work adopts Equation (1) as the basic transition function for joint estimation of R_t and j_t.

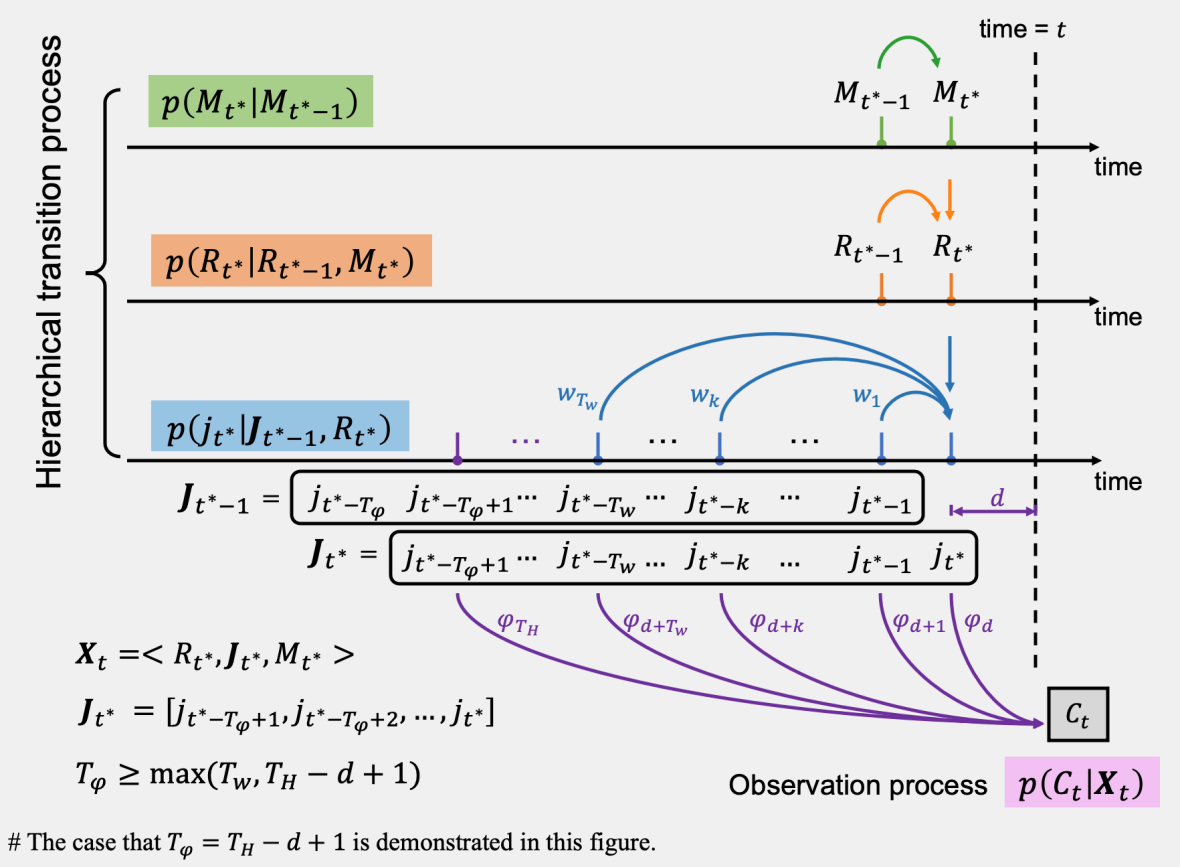
2. Observation process

In epidemiology, the daily infection number j_t cannot be measured directly but is reflected in observations such as the case reports of onset, confirmed and deaths. There is an inevitable time delay between the real date of infection and the date reporting, due to the incubation time, report delay, etc. Taking account of this time delay, we model the observation process as a convolution function between kernel φ, and the infection number in T_H most recent days.

$$C_t = \sum_{k=d}^{T_H} \varphi_k \, j_{t-k} \tag{2}$$

where C_t is the observation data, and φ_k is the probability that an individual infected is observed on day k. T_H is the maximum dependency window. It is assumed that the past daily infections before this window do not affect the current observation C_t. Since there is a delay between observation and infection, we suppose the most recent infection that can be observed by C_t is at the time $t - d$, where d is a constant determined by the distribution of observation delay.

To accommodate various observation types (e.g., the number of daily reported cases, onsets, deaths and infected cases), DARt will choose the appropriate time delay distributions accordingly. For example, for the input of onsets, the infected-to-onsets time distribution is chosen to be the kernel in the observation function. For the input of daily reported cases, the infected-to-onset and the onset-to-report delays are used together as the kernel in the observation function. These delay distributions can be either directly obtained from literature or learned from case reports that contain individual observation delays[3]. Detailed descriptions of the observation functions for different epidemic curves can be found in Supplementary Materials (Section 1.2).



583

584 **Figure 7. Illustration of the hierarchical transition process and observation process.** The
585 most recent infection that can be observed by C_t is at the time $t^* = t - d$ where d is a constant
586 determined by the distribution of observation delay. Suppose T_φ is the length of the vector
587 $\boldsymbol{J}_{t^*} = [j_{t^*-T_\varphi+1}, j_{t^*-T_\varphi+2}, \ldots, j_{t^*}]$ such that C_t is only relevant to $\boldsymbol{J}_{t^*}$ and j_{t^*} only depends on
588 $\boldsymbol{J}_{t^*-1}$ via the renewal process. Therefore, $T_\varphi \geq \max(T_w, T_H - d + 1)$. The case that $T_\varphi =$
589 $T_H - d + 1$ is depicted in this figure.

590 ## 3. Sequential Bayesian Inference

591 In Figure 8, we illustrate the Bayesian inference scheme of DARt with the following
592 descriptions.

593 ▪ **State transition model**

594 In our model, indirectly observable variables j_t and R_t are included in the latent state. The state
595 transition function for R_t is commonly assumed to follow a Gaussian random walk or constant
596 within a sliding window as implemented in EpiEstim. This kind of simplification is not capable
597 of capturing an abrupt change in R_t under stringent intervention measures. To address this

problem, we introduce an auxiliary binary latent variable M_t to characterise and switch between two distinct evolution patterns of R_t – smooth transition (Mode I, $M_t = 0$) and abrupt change (Model II, $M_t = 1$):

$$p(R_t|R_{t-1}, M_t) \sim \begin{cases} \mathcal{N}(R_{t-1}, {\sigma_R}^2) & M_t = 0 \quad \text{Mode I} \\ \mathrm{U}[0, R_{t-1} + \Delta] & M_t = 1 \quad \text{Mode II} \end{cases} \tag{3}$$

where $\mathcal{N}(R_{t-1}, {\sigma_R}^2)$ is a Gaussian distribution with the mean value of R_{t-1} and variance of ${\sigma_R}^2$, describing the random walk with the randomness controlled by σ_R. $\mathrm{U}[0, R_{t-1} + \Delta]$ is a uniform distribution between 0 and $R_{t-1} + \Delta$ allowing sharp decrease while limiting the amount of increase. This is because we assume that R_t can have a big decrease when intervention is introduced but it is unlikely to increase dramatically as the characteristics of disease would not change instantly.

The transition of the change indicator M_t, is modelled as a discrete Markovian process with fixed transition probabilities:

$$p(M_t = 0|M_{t-1}) = p(M_t = 0) = \alpha \tag{4a}$$

$$p(M_t = 1|M_{t-1}) = p(M_t = 1) = 1 - \alpha \tag{4b}$$

where α is a value close to and lower than 1. The above function means that the value of M_t is independent of M_{t-1}, while the probability of Mode II (i.e., $M_t = 1$) is quite small. This is because it is unlikely to have frequent abrupt changes in R_t.

For the incident infection j_t, the state transition can be modelled based on Equation (1) as $p\left(j_t \middle| j_{t-1}, \dots, j_{t-T_w}\right)$. To make the transition process to be Markovian, we vectorise the infection numbers as follows. Suppose the infection numbers $\{j_{t-k}\}_{k=d}^{T_H}$ that can be observed in C_t are all contained in $\boldsymbol{J}_{t^*} = [j_{t^*-T_\varphi+1}, j_{t^*-T_\varphi+2}, \dots, j_{t^*}]$, where $t^* = t - d$, and the length of this vector T_φ is larger than or equal to $T_H - d + 1$. We also require T_φ to be not smaller than T_w. Therefore, all the historical information needed to infer j_{t^*} is available from $\boldsymbol{J}_{t^*-1}$, i.e., $\boldsymbol{J}_{t^*}$ only depends on $\boldsymbol{J}_{t^*-1}$ (i.e., being Markovian). The state transition process and observation process are illustrated in Figure 7.

The latent state in our model is then defined as $\boldsymbol{X}_t =< R_{t^*}, \boldsymbol{J}_{t^*}, M_{t^*} >$, which contribute to C_t at time t. The state transition function of $\boldsymbol{J}_{t^*}$ is therefore Markovian:

$$p(\boldsymbol{J}_{t^*}|\boldsymbol{J}_{t^*-1}, R_{t^*}) = Poisson(j_{t^*};\ R_{t^*} \sum_{k=1}^{T_w} w_k\, j_{t^*-k}\,) \prod_{m=1}^{T_\varphi - 1} \delta(\boldsymbol{J}_{t^*}^{(m)}, \boldsymbol{J}_{t^*-1}^{(m+1)}) \quad (5)$$

where $\boldsymbol{J}_{t^*}^{(m)}$ is the m-th component of the latent variable $\boldsymbol{J}_{t^*}$ and $\delta(x, y)$ is the Kronecker delta function (please refer to Supplementary Materials 2.3 for more details). With Equation (3)-(5), the latent state transition function $p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})$ can be obtained as a Markov process:

$$p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}) = p(\boldsymbol{J}_{t^*}|\boldsymbol{J}_{t^*-1}, R_{t^*}) p(R_{t^*}|R_{t^*-1}, M_{t^*}) p(M_{t^*}|M_{t^*-1}) \quad (6)$$

- **Forward Filtering**

We formulate the inference of the latent state $\boldsymbol{X}_t =< R_{t^*}, \boldsymbol{J}_{t^*}, M_{t^*} >$ with the observations C_t as within a data assimilation framework. Sequential Bayesian approach (also called 'filtering') is adopted to infer the time-varying latent state, which updates the posterior estimation using the latest observations following the Bayes rule.

Let us denote the observation history between time 1 and t as $C_{1:t} = [C_1, C_2, \dots, C_t]$. Given that previous estimation $p(\boldsymbol{X}_{t-1}|C_{1:t-1})$ and new observation C_t, we would like to update the estimation of $\boldsymbol{X}_t$, i.e., $p(\boldsymbol{X}_t|C_{1:t})$ following the Bayes rule with the assumption that $C_{1:t}$ is conditionally independent of $C_{1:t-1}$ given $\boldsymbol{X}_t$:

$$p(\boldsymbol{X}_t|C_{1:t}) = \frac{p(C_t|\boldsymbol{X}_t) p(\boldsymbol{X}_t|C_{1:t-1})}{\int\ p(C_t|\boldsymbol{X}_t) p(\boldsymbol{X}_t|C_{1:t-1}) d\boldsymbol{X}_t} \quad (7)$$

where $p(\boldsymbol{X}_t|C_{1:t-1})$ is prior and $p(C_t|\boldsymbol{X}_t)$ is the likelihood. The prior can be written in the marginalised format:

$$p(\boldsymbol{X}_t|C_{1:t-1}) = \int p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}) p(\boldsymbol{X}_{t-1}|C_{1:t-1})\, d\boldsymbol{X}_{t-1} \quad (8)$$

where $\boldsymbol{X}_t$ is assumed to be conditionally independent of $C_{1:t-1}$ given $\boldsymbol{X}_{t-1}$, and the transition $p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})$ is defined in Equation (6) based on the underlying renewal process. The likelihood

$p(C_t|\boldsymbol{X}_t)$ can be calculated assuming the observation uncertainty follows a Gaussian distribution:

$$p(C_t|\boldsymbol{X}_t) \sim \mathcal{N}(H(\boldsymbol{X}_t), {\sigma_C}^2) \tag{9}$$

where H is the observation function with a kernel chosen accordingly to the types of observations and ${\sigma_C}^2$ is the variance of observation error estimated empirically. To show the benefits of using this Gaussian likelihood function, we show the simulation results of using Poisson likelihood without considering the observation noise. Results can be found in Supplementary Figure 2, where the estimations fluctuate dramatically under noisy observation.

By substituting Equation (8) into Equation (7), we obtain the iterative update of $p(\boldsymbol{X}_t|C_{1:t})$ given the transition $p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})$ and likelihood $p(C_t|\boldsymbol{X}_t)$:

$$p(\boldsymbol{X}_t|C_{1:t}) = \frac{p(C_t|\boldsymbol{X}_t)\int p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})p(\boldsymbol{X}_{t-1}|C_{1:t-1})\,d\boldsymbol{X}_{t-1}}{\iint p(C_t|\boldsymbol{X}_t)\int p(\boldsymbol{X}_t|\boldsymbol{X}_{t-1})p(\boldsymbol{X}_{t-1}|C_{1:t-1})\,d\boldsymbol{X}_{t-1}d\boldsymbol{X}_t} \tag{10}$$

- **Backward Smoothing**

In DARt, we adopt backward smoothing to infer the latent state at a certain time, given all observations relevant to the state. Based on the filtering results of $p(\boldsymbol{X}_t|C_{1:t})$ for $t \in \{1,..,T\}$, we can obtain the smoothing results $p(\boldsymbol{X}_t|C_{1:T})$, where T is the total number of observations. To assimilate the information from subsequent observations, we use a standard backward pass method:

$$p(\boldsymbol{X}_t|C_{1:T}) = p(\boldsymbol{X}_t|C_{1:t})\int p(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t)\frac{p(\boldsymbol{X}_{t+1}|C_{1:T})}{p(\boldsymbol{X}_{t+1}|C_{1:t})}d\boldsymbol{X}_{t+1} \tag{11}$$

which provides the iterative backward calculation of $p(\boldsymbol{X}_t|C_{1:T})$ from time T to time t. More details can be found in Supplementary Materials (Section 2). For computational convenience, particle methods are used to approximate the posterior distributions (details in Supplementary Materials, Section 2.4). Detailed parameter settings can be found in Supplementary Materials, Section 4.

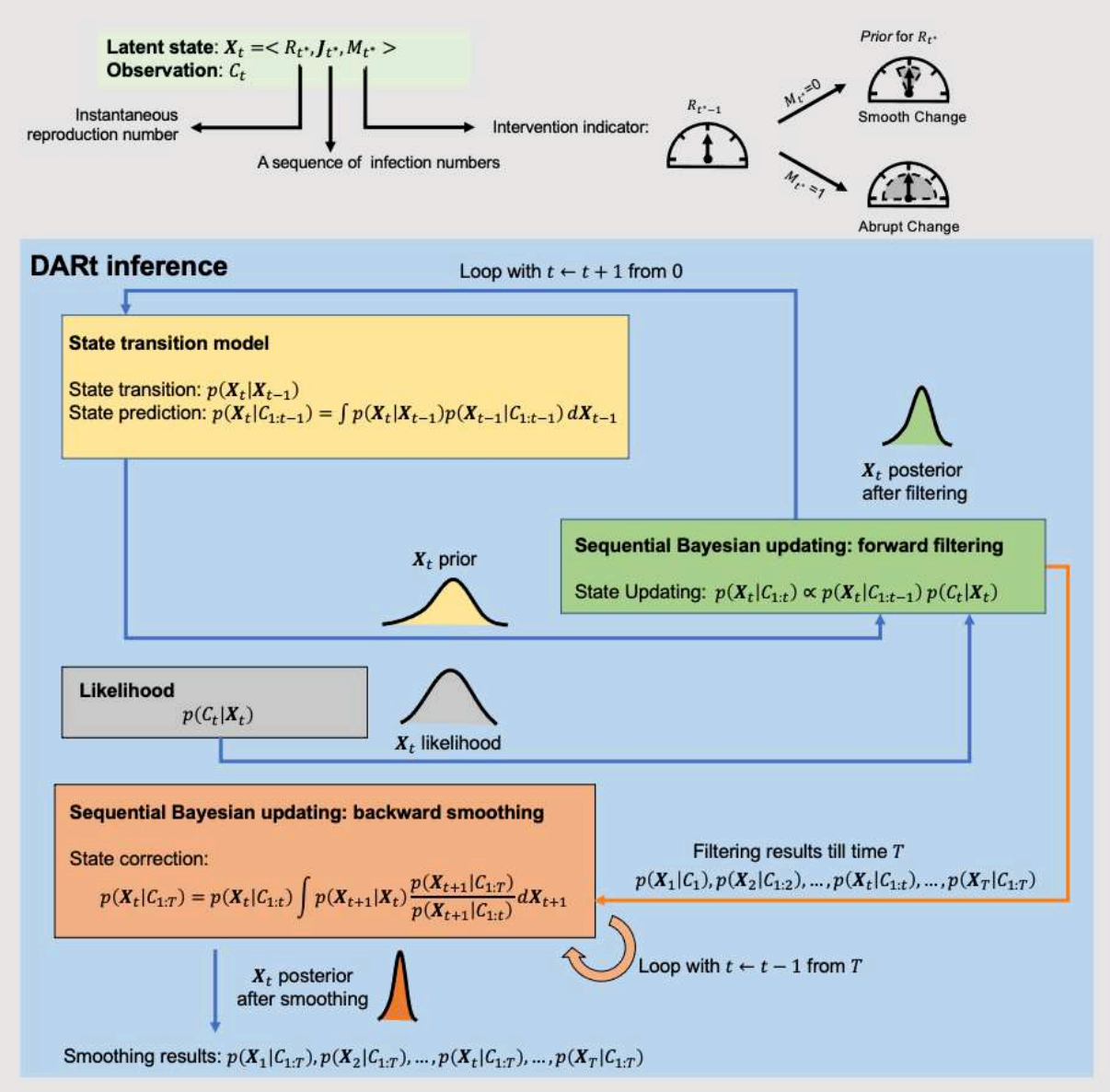


Figure 8. Three components of DARt: state transition model, observation function and sequential Bayesian update module with two phases (forward filtering and backward smoothing). The latent state that can be observed in C_t are defined as $\boldsymbol{X}_t =< R_{t^*}, \boldsymbol{J}_{t^*}, M_{t^*} >$ where R_{t^*} is the instantaneous reproduction number, M_{t^*} is a binary state variable indicating different evolution patterns of R_{t^*}, $\boldsymbol{J}_{t^*} = [j_{t^*-T_\varphi+1}, j_{t^*-T_\varphi+2}, \ldots, j_{t^*}]$ is a vectorised form of infection numbers j_t, t^* indicates the most recent infection that can be observed at time t is from the time t^* due to observation delay, and T_φ is the length of the vector $\boldsymbol{J}_{t^*}$ such that C_t is only relevant to $\boldsymbol{J}_{t^*}$ and j_{t^*+1} only depends on $\boldsymbol{J}_{t^*}$ via the renewal process.

Data availability

We obtained daily onset or confirmed cases of four different regions (Wuhan, Hong Kong, Sweden, UK) from publicly available resources[1,31–33] (accessed on 11th of December 2020). For Wuhan, we adopted the daily number of onset patients from the retrospective study[1] (from the middle of December to early March). For UK data, we downloaded the daily report cases (cases by date reported) from the official UK Government website for data and insights on Coronavirus (COVID-19)[32] (from early March to the end of November 2020). Data for UK Cities were also downloaded from the same resource[32] (from early August to the end of November). For Sweden data, we downloaded the daily number of confirmed cases from the European Centre for Disease Prevention and Control[33] (from early March to the end of November 2020). For Hong Kong, we downloaded the case reports from government website[31] (from early July to the end of November 2020), including descriptive details of individual confirmed case of COVID-19 infection in Hong Kong. For those asymptomatic patients whose onset date are unknown, we applied their reported date as their onset date, and for those whose onset date is unclear, we simply removed and neglected these records. Only local cases and their related cases are considered, while imported cases and their related cases are excluded.

Code availability

We are releasing DARt as open-source software for epidemic research and intervention policy design and monitoring. The source code of our method and our web service are publicly available online (https://github.com/Kerr93/DARt).

References

1. Pan, A. *et al.* Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA* **323**, 1915 (2020).
2. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
3. Ferretti, L. *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science (80-.).* **368**, eabb6936 (2020).
4. Neil M, F. *et al.* Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imp. Coll. COVID-19 Response Team* (2020).
5. Cowling, B. J. *et al.* Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Heal.* (2020).
6. Lai, S. *et al.* Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* (2020).
7. Reiner, R. C. *et al.* Modeling COVID-19 scenarios for the United States. *Nat. Med.* (2020).
8. Chang, S. L., Harding, N., Zachreson, C., Cliff, O. M. & Prokopenko, M. Modelling transmission and control of the COVID-19 pandemic in Australia. *Nat. Commun.* **11**, 5710 (2020).
9. Fraser, C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One* **2**, (2007).
10. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697 (2020).
11. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 1–5 (2020).
12. Wallinga, J. & Teunis, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* **160**, 509–516 (2004).
13. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512 (2013).
14. Bettencourt, L. M. A. & Ribeiro, R. M. Real Time Bayesian Estimation of the

733 Epidemic Potential of Emerging Infectious Diseases. *PLoS One* **3**, e2185 (2008).
734 15. Abbott, S. *et al.* Estimating the time-varying reproduction number of SARS-CoV-2
735 using national and subnational case counts. *Wellcome Open Res.* **5**, 112 (2020).
736 16. Adam, D. A guide to R-the pandemic's misunderstood metric. *Nature* **583**, 346–348
737 (2020).
738 17. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19.
739 *Nat. Med.* (2020).
740 18. Gostic, K. *et al.* Practical considerations for measuring the effective reproductive
741 number, Rt. *medRxiv.* (2020).
742 19. Goldstein, E. *et al.* Reconstructing influenza incidence by deconvolution of daily
743 mortality time series. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21825–21829 (2009).
744 20. Wang, S. *et al.* A Bayesian Updating Scheme for Pandemics: Estimating the Infection
745 Dynamics of COVID-19. *IEEE Comput. Intell. Mag.* **15**, 23–33 (2020).
746 21. Asch, M., Bocquet, M. & Nodet, M. *Data assimilation: methods, algorithms, and*
747 *applications*. (2016).
748 22. Chen, Z. Bayesian filtering: From Kalman filters to particle filters, and beyond.
749 *Statistics (Ber).* **182**, 1–69 (2003).
750 23. Doucet, A. & Johansen, A. M. A tutorial on particle filtering and smoothing: Fifteen
751 years later. in *Handbook of nonlinear filtering* vol. 12 3 (2009).
752 24. Mark, C. *et al.* Bayesian model selection for complex dynamic systems. *Nat. Commun.*
753 **9**, (2018).
754 25. Thompson, R. N. *et al.* Improved inference of time-varying reproduction numbers
755 during infectious disease outbreaks. *Epidemics* **29**, (2019).
756 26. Grassly, N. C. & Fraser, C. Mathematical models of infectious disease transmission.
757 *Nat. Rev. Microbiol.* **6**, 477–487 (2008).
758 27. Vynnycky, E. & White, R. *An introduction to infectious disease modelling*. (OUP
759 oxford, 2010).
760 28. Friston, K. J. Variational filtering. *Neuroimage* **41**, 747–766 (2008).
761 29. Xihong Lin's Group. Visualizing COVID-19's Effective Reproduction Number (Rt).
762 http://metrics.covid19-analysis.org/.
763 30. Leung, K., Wu, J. T., Liu, D. & Leung, G. M. First-wave COVID-19 transmissibility
764 and severity in China outside Hubei after control measures, and second-wave scenario
765 planning: a modelling impact assessment. *Lancet* **395**, 1382–1393 (2020).
766 31. HK Centre of Health Protection. Latest local situation of COVID-19. (2020).

32. Department of Health and Social Care. Coronavirus (COVID-19) in the UK: Cases. https://coronavirus.data.gov.uk/cases.
33. European Centre for Disease Prevention and Control. Daily number of new reported cases of COVID-19 by country worldwide. https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide.
34. Wu, J. T. *et al.* Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510 (2020).
35. Friston, K., Stephan, K., Li, B. & Daunizeau, J. Generalised filtering. *Math. Probl. Eng.* (2010).
36. Ghahramani, Z. & Hinton, G. E. Variational learning for switching state-space models. *Neural Comput.* (2000).
37. Friston, K. J. *et al.* Testing and tracking in the UK: A dynamic causal modelling study. *Wellcome Open Res.* **5**, 144 (2020).
38. Friston, K. J. *et al.* Second waves, social distancing, and the spread of COVID-19 across America. *arXiv* (2020).
39. Friston, K. J., Costello, A. & Pillay, D. 'Dark matter' , second waves and epidemiological modelling. *medRxiv* (2020).
40. Cintrôn-Arias, A., Castillo-Chavez, C., Bettencourt, L. M. A., Lloyd, A. L. & Banks, H. T. The estimation of the effective reproductive number from disease outbreak data. *Math. Biosci. Eng.* **6**, 261–282 (2009).

Author Contributions

X.Y., S.W. and Y.G. conceived the overall study. X.Y., S.W., Y.G. and R.X. designed the method and analysed the data. X.Y. and Y.X. contributed to the experimental design and algorithm implementation. Y.X. delivered the tool and Web service. X.Y., S.W, L.L., K.F. and Y.G. wrote the manuscript. K.F., R.X. and L.L provided statistical guidance. All authors have discussed the results, edited and approved the contents of the manuscript.

Figures

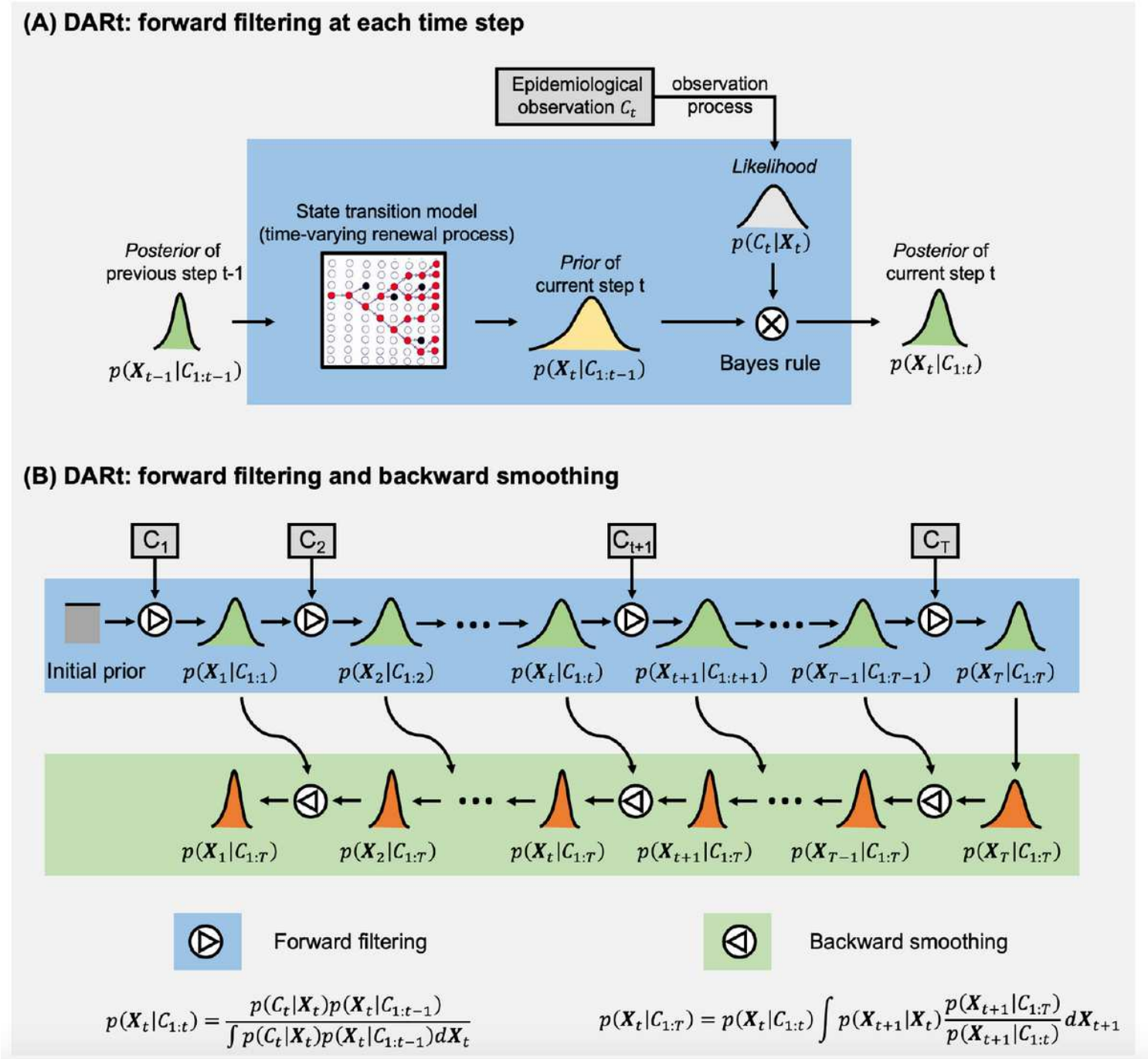


Figure 1

[See manuscript PDF file for full caption] Illustration of the inference of DARt system for t estimation.

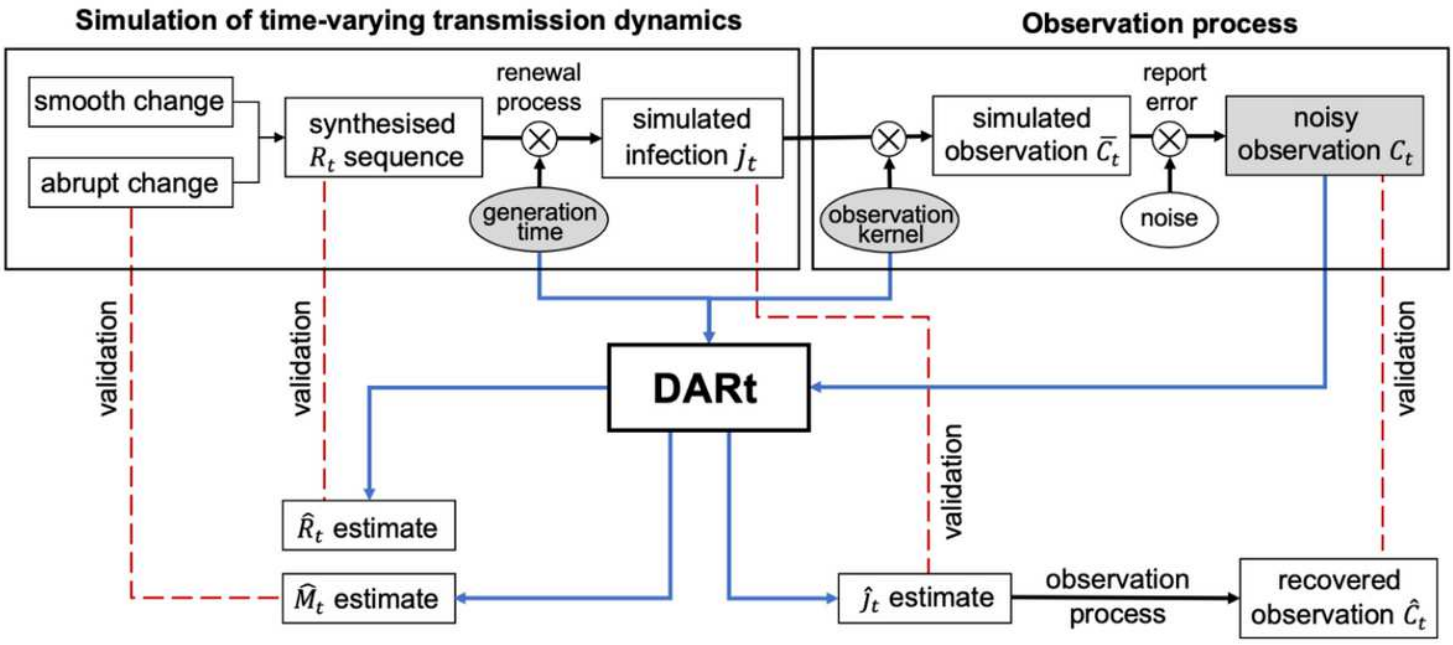


Figure 2

[See manuscript PDF file for full caption]. Validation experiment of the DARt system on simulated data.

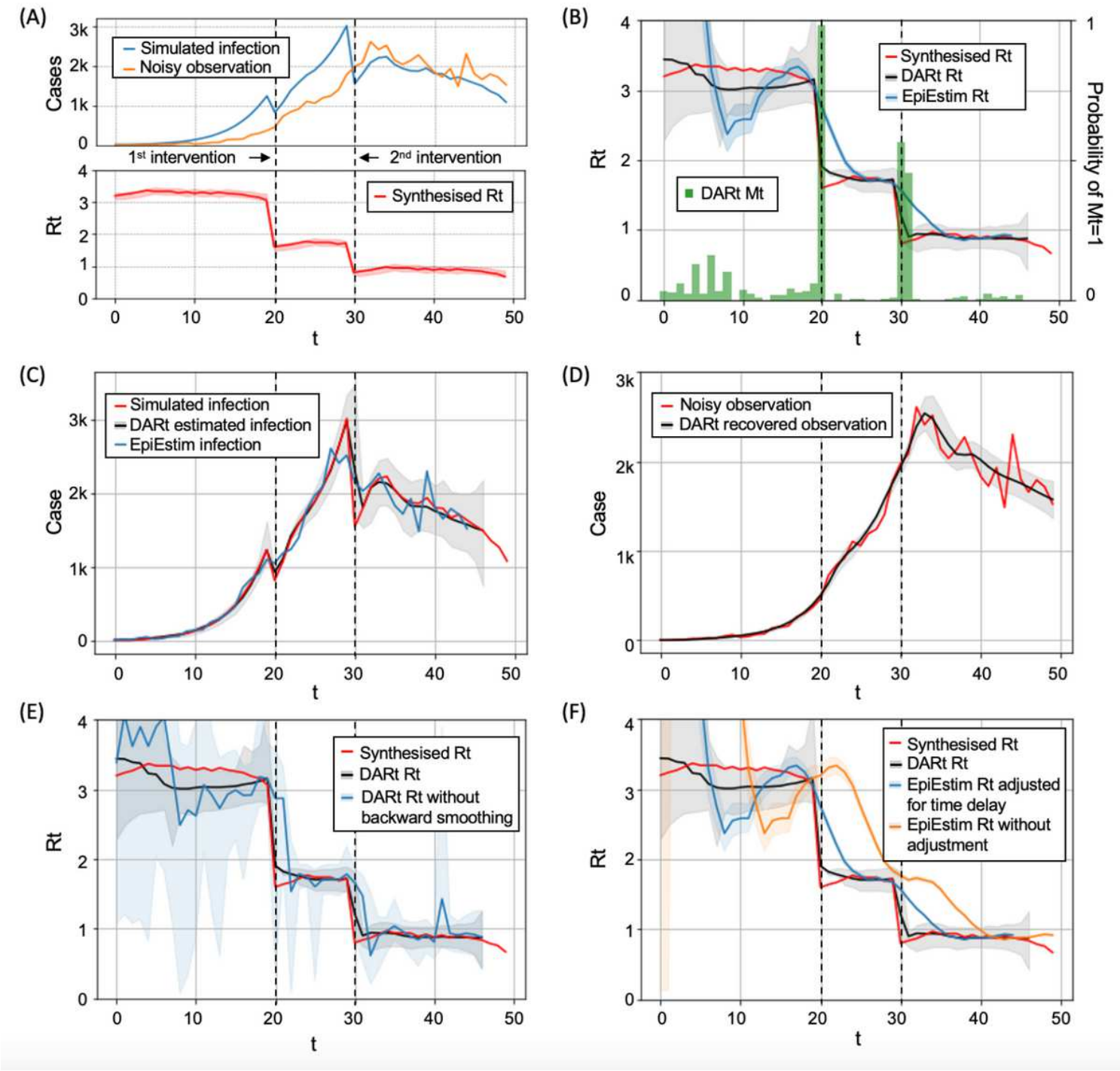


Figure 3

[See manuscript PDF file for full caption] Validation results.

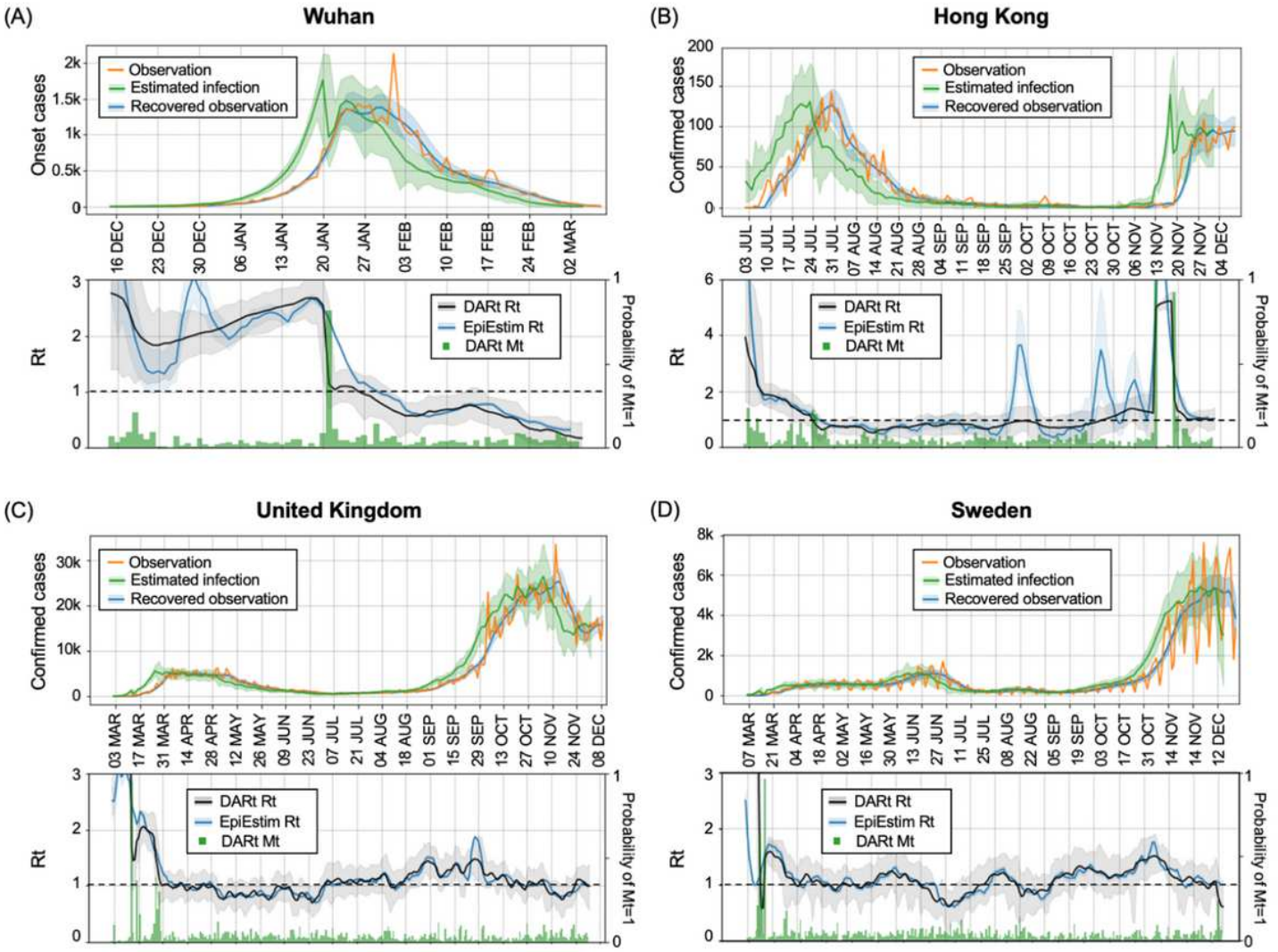


Figure 4

[See manuscript PDF file for full caption] Epidemic dynamics in Wuhan (A), Hong Kong (B), United Kingdom (C) and Sweden (D).

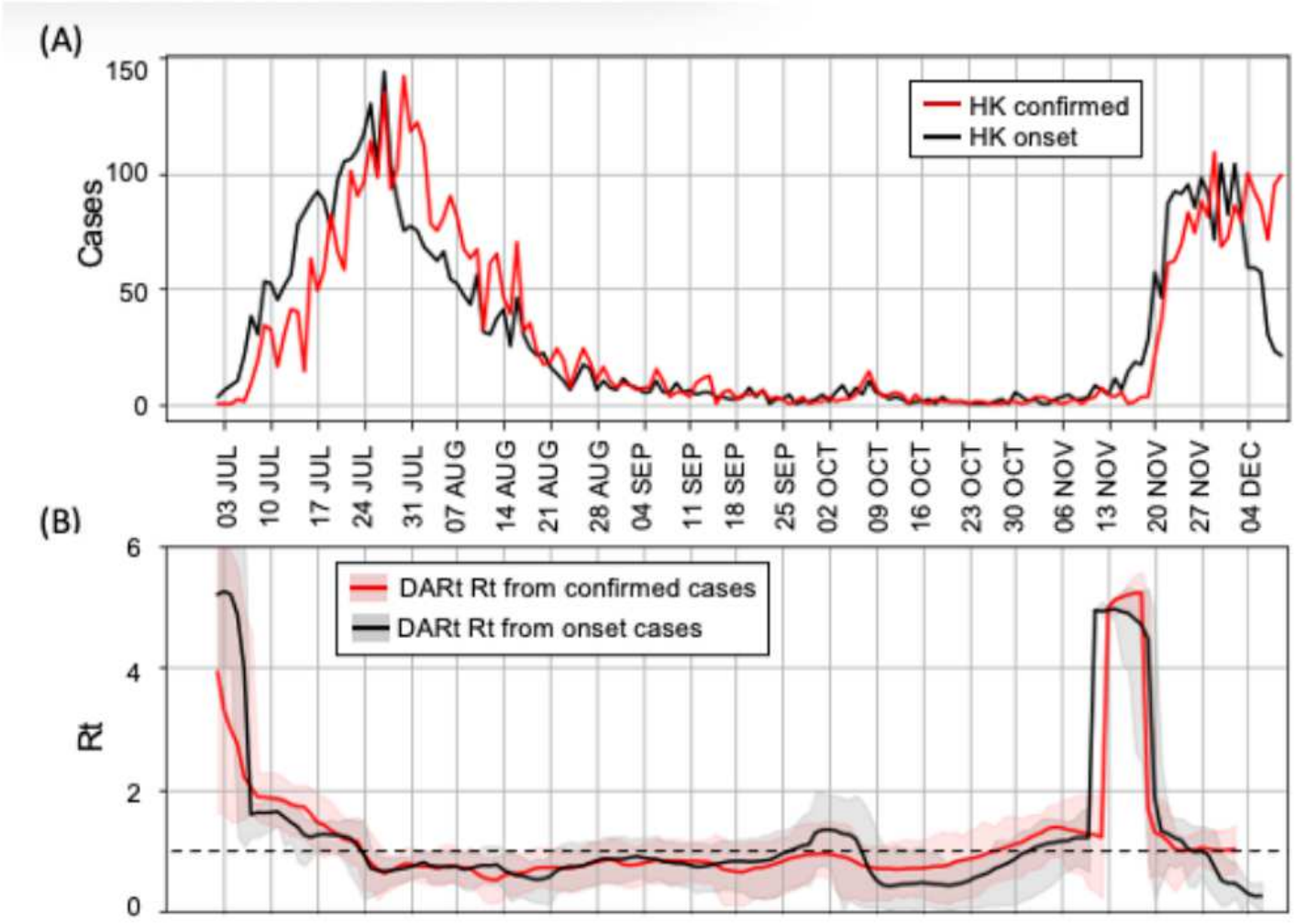


Figure 5

[See manuscript PDF file for full caption] Comparison of estimated □t curves of Hong Kong using different observations.

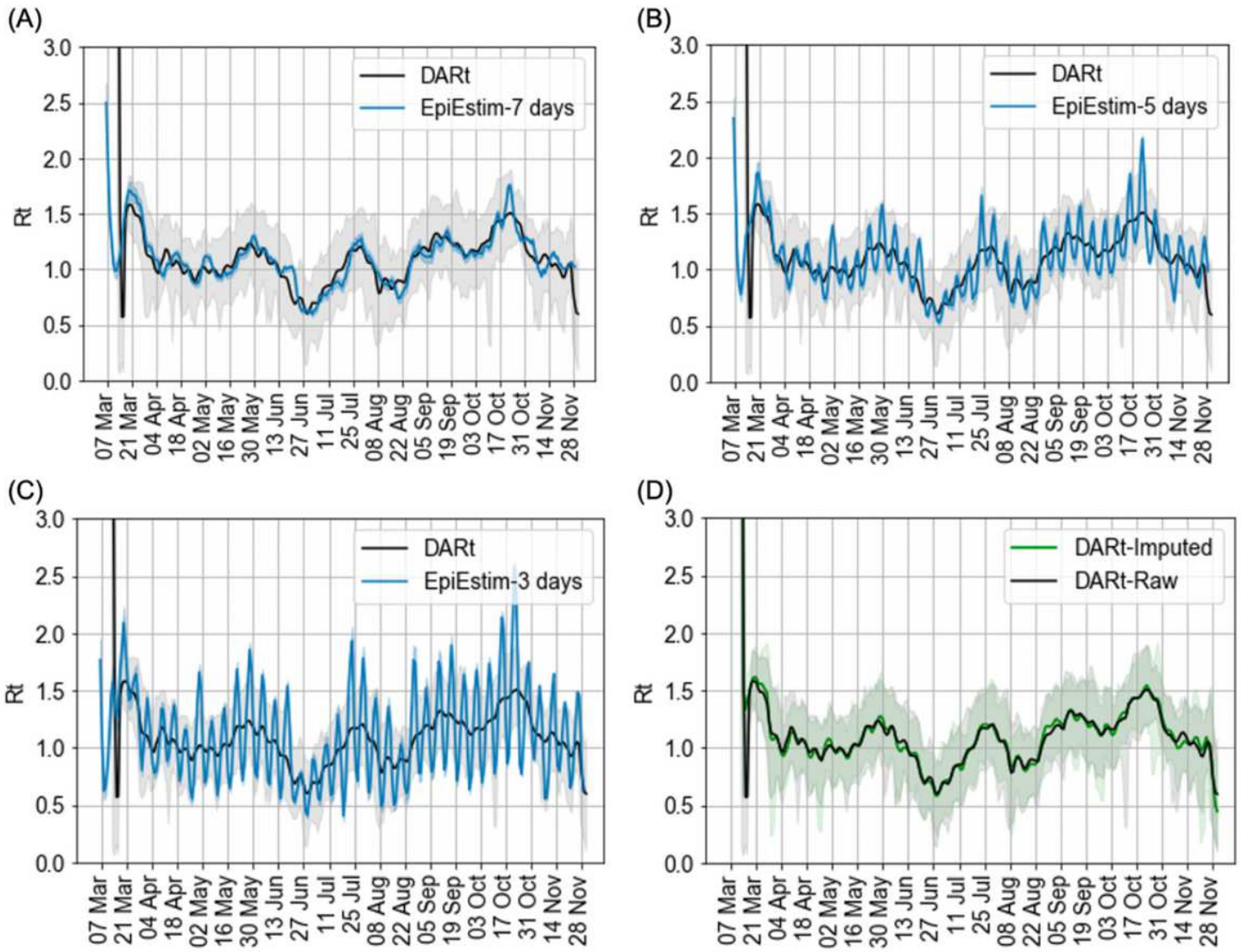


Figure 6

[See manuscript PDF file for full caption] Inference results for Sweden with different settings.

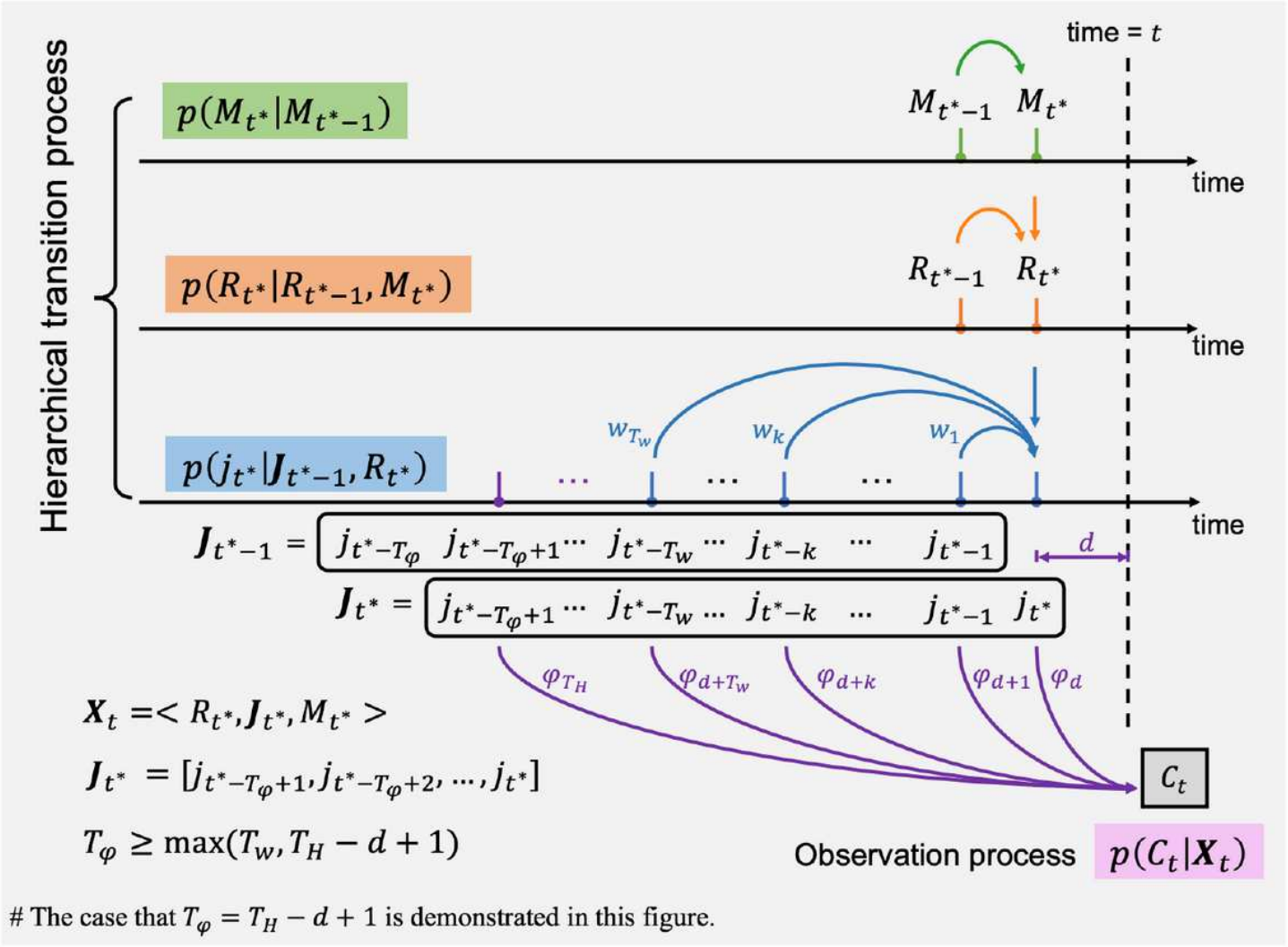


Figure 7

[See manuscript PDF file for full caption] Illustration of the hierarchical transition process and observation process.

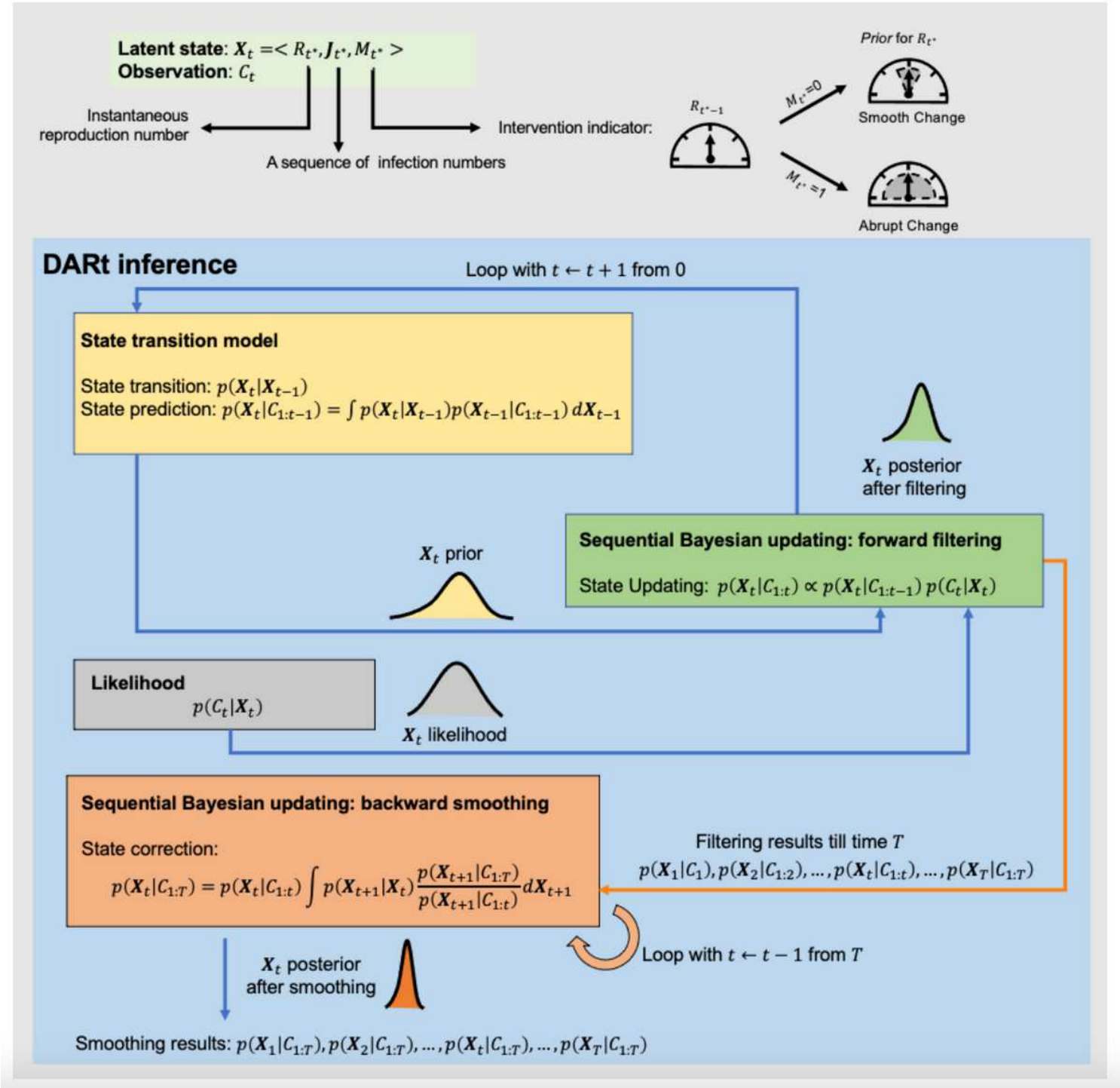


Figure 8

[See manuscript PDF file for full caption] Three components of DARt: state transition model, observation function and sequential Bayesian update module with two phases (forward filtering and backward smoothing).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarysubmission.pdf