medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves

Kris V Parag[†, 1]

[1]MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, W2 1PG, UK
[†]Email: k.parag@imperial.ac.uk

***Abstract*—We construct a recursive Bayesian smoother, termed EpiFilter, for estimating the effective reproduction number, R, from the incidence of an infectious disease in real time and retrospectively. Our approach borrows from Kalman filtering theory, is quick and easy to compute, generalisable, deterministic and unlike many current methods, requires no change-point or window size assumptions. We model R as a flexible, hidden Markov state process and exactly solve forward-backward algorithms, to derive R estimates that incorporate all available incidence information. This unifies and extends two popular methods, EpiEstim, which considers past incidence, and the Wallinga-Teunis method, which looks forward in time. This combination of maximising information and minimising assumptions, makes EpiFilter more statistically robust in periods of low incidence, where existing methods can struggle. As a result, we find EpiFilter to be particularly suited for assessing the risk of second waves of infection, in real time.**

Key-words: Bayesian filters, reproduction numbers, epidemic models, COVID-19, infectious diseases.

INTRODUCTION

During an unfolding epidemic, one of the most commonly available and useful types of surveillance data is the daily (or weekly) number of new infected cases. This time-series of case counts, also known as the incidence curve, not only measures the epidemic size and burden, but also provides information about trends or changes in its transmissibility [1], [2]. These trends are captured by the time-varying effective reproduction number, denoted R_s at time s, which defines how the number of secondary cases generated per primary case varies across the duration of the outbreak [3]. When $R_s > 1$ we can expect and prepare for growing incidence, whereas $R_s < 1$ signifies that the epidemic is being controlled [4].

Inferring changes in R_s given an observed incidence curve is therefore crucial, both to understanding transmissibility and to forecasting upcoming case loads, especially in an ongoing epidemic, where it can help inform policymaking and intervention choices [1], [5]. Real-time and retrospective R_s estimates have been used to characterise rates and patterns of spread in various diseases such as malaria [6] and Ebola virus disease [7] and have already proven valuable across the COVID-19 pandemic, providing updating synopses of global transmission [8] or evidencing the impact of implemented control actions (e.g. lockdowns and social distancing) [9], [10].

Many of the studies that aim to infer R_s either apply the Wallinga-Teunis (WT) method [2] or the Cori *et al* method, known as EpiEstim [11]. Both methods take complementary viewpoints on how incidence data inform on R_s and hence have diverse use-cases. The WT method reconstructs the average number of new cases caused by infectious individuals at s and so requires incidence data beyond time s for its estimate, which is called the case reproduction number. It is better suited for retrospective analyses [12]. Alternatively, EpiEstim infers how past infections propagate to form the incidence observed at s, only requiring data prior to time s. EpiEstim is preferred for real-time investigations and its R_s estimate is termed the instantaneous reproduction number [3].

While both methods provide useful and important estimators of transmission, they are not perfect. Two main limitations exist. First, each suffers from data censoring or edge-effects [3]. Because the WT method is forward-looking i.e. depends on data later than s, its estimates are right censored when s is close to the last observed time point [12]. In contrast, EpiEstim looks backward in time and suffers edge-effects when s is near the first observed time point [11]. Estimates in the vicinity of the start and end of the incidence time-series are therefore unreliable under EpiEstim and the WT method, respectively.

Techniques have been proposed to limit this unreliability [5], [14], but the problem is intrinsic, and inevitable near the actual start and end times of an epidemic, where there is necessarily no or scarce data. This leads into the second key limitation: inference in periods of small incidence. This presents a fundamental challenge for any R_s estimation approach and effectively creates additional edge-effects. When few or no case counts are available to

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

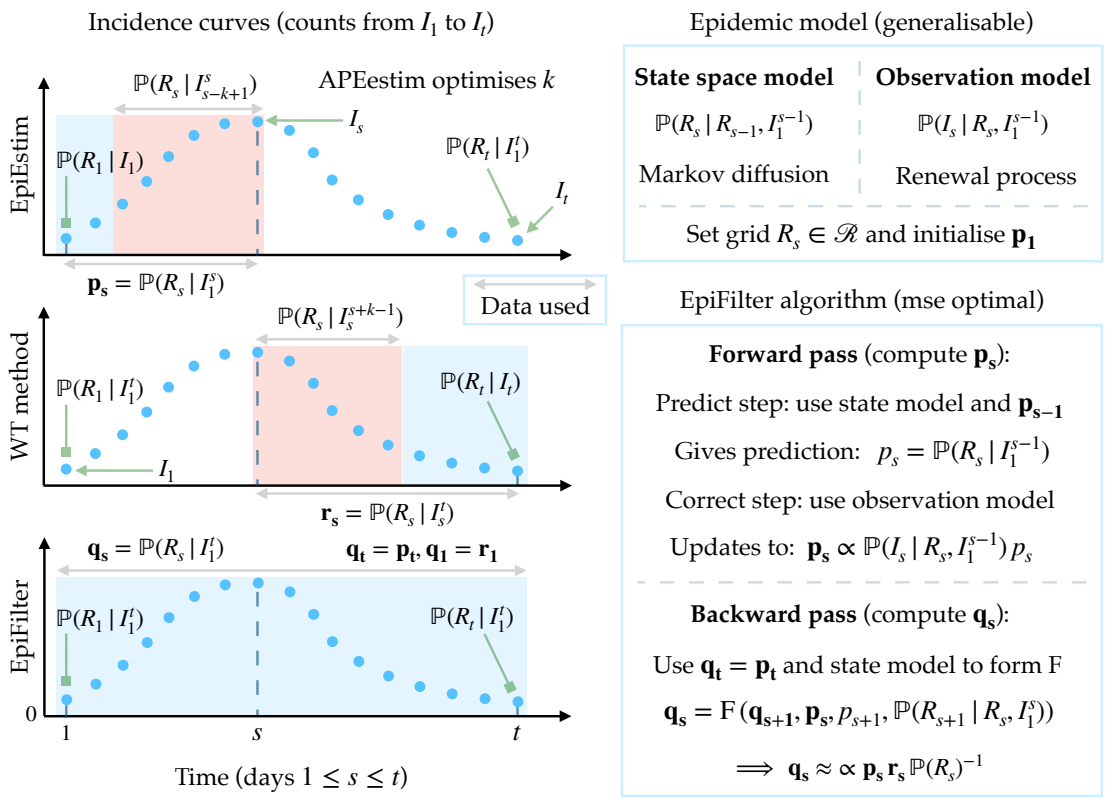

Fig. 1: **EpiFilter algorithm.** We summarise the construction of EpiFilter and compare its use of information to that of EpiEstim and the WT method. Left panels show that EpiEstim and the WT method consider complementary parts of the incidence curve, I^t_1 (blue dots), to be informative about the reproduction number R_s, while EpiFilter makes use of the entire curve. Red windows indicate the portions of the curve that inform on R_s under a window of size k as in [11], while blue windows show the portions involved in constructing the main posterior distributions in this work: $\mathbf{p_s}$, $\mathbf{r_s}$ and $\mathbf{q_s}$. Right panels outline the main assumptions (the model box) and computations (the algorithm box) necessary for realising EpiFilter, which allow us to obtain the most informative (and minimum mse) posterior distribution $\mathbf{q_s}$. See the main text for the specific equations employed in our implementation [13].

constrain inference, methods are largely driven by their inherent prior distributions and assumptions, which can result in misleading and unreliable estimates [11], [15]. Understanding how to best mediate the trade-off between prior assumptions and data when incidence is small is of both statistical and epidemiological importance.

Following a period of low incidence, two outcomes are possible: either the epidemic continues to exhibit small or zero case counts until it goes extinct or a resurgence in infections, also termed a second wave, occurs. Deciphering, in real time, which of these conditions is likely is a key challenge for infectious disease epidemiology given the limited information available at low incidence [16]. Better inference of transmission under these conditions is currently considered central to designing data-informed COVID-19 intervention exit or relaxation strategies [17]. With many countries now facing potential second waves in this pandemic, estimating fluctuations in transmission during suspected epidemic troughs could be essential to achieving sustained control [10].

Here we present and develop a novel method, termed EpiFilter, for reliably estimating R_s in real time, which ameliorates the above limitations. We take an engineering inspired approach and construct an exact, recursive and deterministic (i.e. EpiFilter produces the same output for fixed input data and requires no Monte Carlo steps) inference algorithm that is quick and easy to compute both across an unfolding outbreak and in retrospect. Our method solves what is called the smoothing problem in control engineering [13]. This means we compute R_s estimates that formally integrate both forward and backward looking information. This unifies the WT method and EpiEstim, and nullifies their edge-effect issues.

Further, EpiFilter only makes a minimal Markov assumption for R_s, which allows it to avoid the strong prior window size and change-point assumptions that existing

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

methods may apply to infer shifts in transmission [11], [9]. Using simulated and empirical data, we show that EpiFilter accurately tracks changes in R_s and provides reliable one-step-ahead incidence predictions. Moreover, we find that EpiFilter is appreciably more robust than even optimised versions of EpiEstim [14], if performing real-time inference in periods of low incidence and between epidemic waves. We illustrate the practical utility of EpiFilter on the COVID-19 incidence curve of New Zealand, which exhibits a potential second wave.

Our method, which is outlined in Fig. 1, provides a straightforward yet formally optimal (in mean squared error [18]) solution to real-time and retrospective R_s estimation. Because it couples minimal prior assumptions with maximum information extraction, it more gracefully handles periods with scarce data. Hopefully our approach will serve as a useful inference tool for investigating the risk of second waves of infection in COVID-19 and other epidemics. An implementation of EpiFilter is available at https://github.com/kpzoo/EpiFilter and its mechanics are explored and validated in the Appendix.

METHODS

Renewal models and inference problems

We consider an infectious disease epidemic observed over some period of time $1 \leq s \leq t$. If the incidence or number of newly infected cases at time s is I_s then $I_1^t = \{I_s : 1 \leq s \leq t\}$ is known as the incidence curve of the epidemic. For convenience, we assume that incidence is available on a daily scale so that I_1^t is a vector of t daily counts (weeks or months could be used instead). A common problem in infectious disease is the inference of the transmissibility of the epidemic given this incidence curve. The renewal model [1], [19] presents a general and popular framework for investigating this problem.

This model posits that epidemic transmissibility, summarised by the effective reproduction number, R_s, generates the observed incidence as in Eq. (1). There Pois indicates Poisson observation noise, $\stackrel{d}{=}$ means that both distributions are equal and $\Lambda_s := \sum_{u=1}^{s-1} I_{s-u} w_u$.

$$\mathbb{P}(I_s \,|\, R_s,\, I_1^{s-1}) \stackrel{d}{=} \text{Pois}(\Lambda_s R_s) \tag{1}$$

In Eq. (1) R_s describes the branching of the epidemic i.e. it captures the number of secondary cases generated per effective primary one at s, while Λ_s, which is termed the total infectiousness, defines how many effective past cases are still infectious at s. The generation time distribution of the epidemic, $w_1^{s-1} = \{w_u : 1 \leq u \leq s-1\}$, controls how past incidence influences Λ_s, with w_u as the probability that a primary case takes between $u-1$ and u days to generate a secondary case [1].

We make the standard assumption that w_1^∞ is known for the epidemic of interest and well approximated by its serial interval distribution and focus on inferring the complete set of R_s values i.e. R_1^t given I_1^t with t as the last recorded time of the incidence curve. Estimating R_1^t is important, not only for learning whether the epidemic is still growing (whenever $R_s > 1$) or under control (i.e. if $R_s < 1$) [4], but also for assessing how much existing interventions can be relaxed without risking a second wave of infection [10]. This latter problem can be particularly difficult (as we will show) since, under these conditions, incidence data are scarce.

We define three key inference problems, commonly studied in engineering [18], based on how information in I_1^t is recruited to construct R_1^t estimates. We represent these problems in terms of the posterior distribution they induce over R_1^t. Estimates are then functions of these posteriors. The first is called causal filtering, where we sequentially calculate $\mathbf{p_s} = \mathbb{P}(R_s \,|\, I_1^s)$ [20]. Causal means that estimates of R_s only depend on data up to time s. Solving this problem is fundamental to real-time inference [21]. The second problem is non-causal filtering and is the complement of the first. Here we calculate $\mathbf{r_s} = \mathbb{P}(R_s \,|\, I_s^t)$, which is important for retrospective or backward-looking estimates [22].

The last problem is termed smoothing and is our main interest. To solve it we must extract all the information possible to formulate $\mathbf{q_s} = \mathbb{P}(R_s \,|\, I_1^t)$ [13]. Note that $\mathbf{p_s}$, $\mathbf{r_s}$ and $\mathbf{q_s}$ depend on the choice of state space model, which describes the dynamics of R_s across time, and observation model, which explains how changes in R_s lead to trends in the observed incidence data [18]. These models encode our assumptions about the epidemic of interest and determine how estimates tradeoff assumptions against data. We will find this latter point important for determining performance when incidence data are scarce. We next explain how the most popular approaches for estimating R_1^t, the WT method and EpiEstim fit within this engineering framework.

Inference methods and low incidence

We mostly detail EpiEstim as real-time R_s estimates are the key focus of this work. EpiEstim assumes that all reproduction numbers are the same over a window back in time, $\tau(s) = \{s,\, s-1,\, \ldots,\, s-k+1\}$ of size k. Consequently, only I_{s-k+1}^s provides information about R_s and the causal filtering distribution $\mathbf{p}_{\tau(\mathbf{s})} = \mathbb{P}(R_s \,|\, I_{s-k+1}^s)$ follows as in Eq. (2) with a shape-scale gamma prior distribution applied i.e. $\mathbb{P}(R_s) \stackrel{d}{=} \text{Gam}(a,\, c^{-1})$ [11].

$$\mathbf{p}_{\tau(\mathbf{s})} \stackrel{d}{=} \text{Gam}(a + i_{\tau(s)},\, (c + \lambda_{\tau(s)})^{-1}) \tag{2}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Here $i_{\tau(s)} = \sum_{u\in\tau(s)} I_u$ and $\lambda_{\tau(s)} = \sum_{u\in\tau(s)} \Lambda_u$ are sums over the window of interest.

The resulting posterior mean estimate is then $\tilde{R}_{\tau(s)} = \int \mathbf{p}_{\tau(\mathbf{s})} R_s \, \mathrm{d}R_s = \mathbb{E}[R_s \,|\, I^s_{s-k+1}] = (a + i_{\tau(s)})(c + \lambda_{\tau(s)})^{-1}$ and the Fisher information that I^s_{s-k+1} contains about R_s is proportional to $\lambda_{\tau(s)}$ [23]. The observation model is Eq. (1) but the state space model of EpiEstim is not explicit. However, if $R_{\tau(s)}$ is the assumed average reproduction number in $\tau(s)$ (which is used to estimate R_s) then $\lambda_{\tau(s)} R_{\tau(s)} = \sum_{u\in\tau(s)} \Lambda_u R_u$ [14]. As a result, EpiEstim somewhat incorporates a linear moving average state space model, and assumes that $\mathbf{p_s} \approx \mathbf{p}_{\tau(\mathbf{s})}$ by deeming I_1^{s-k} to be uninformative. Since $\mathbf{p}_{\tau(\mathbf{s})}$ can be computed sequentially across an ongoing epidemic, EpiEstim provides real-time inference.

The WT method takes a complementary approach to EpiEstim and computes an estimate of R_s based on a forward-looking window [2]. It also uses the observation model of Eq. (1) and has an implicit moving average state model $R_{\tau(s)} = \sum_{u\in\tau(s)} R_u w_{u-s+1}$ [3]. We abuse notation and define $\tau(s) = \{s, s+1, \ldots, s+k-1\}$ in this case as a window into the future of size k. The WT method assumes that $\mathbf{r_s} \approx \mathbf{r}_{\tau(\mathbf{s})} = \mathbb{P}(R_s \,|\, I_s^{s-k+1})$ and approximately solves the non-causal filtering problem. We illustrate the information windows employed by the WT method and EpiEstim and the complete causal and non-causal filtering windows in Fig. 1. The goodness of $\mathbf{p}_{\tau(\mathbf{s})}$ and $\mathbf{r}_{\tau(\mathbf{s})}$ as measures of $\mathbf{p_s}$ and $\mathbf{r_s}$ will depend on k and the appropriateness of the state model [14].

While EpiEstim and the WT methods are powerful tools for inferring R_s in real-time and in retrospect, they have two main and related limitations, which necessarily reduce the reliability of their outputs [12]. First, their performance degrades as s gets close to 1 for EpiEstim and t for the WT method [11]. These edge or censoring effects correspond, at the extreme, to $\mathbf{p_1} = \mathbf{p}_{\tau(\mathbf{1})} = \mathbb{P}(R_1 \,|\, I_1)$ and $\mathbf{r_t} = \mathbf{r}_{\tau(\mathbf{t})} = \mathbb{P}(R_t \,|\, I_t)$, which are weakly informed posterior distributions. As a result, at the beginning of the incidence curve EpiEstim can be unreliable (and even unidentifiable). The WT method suffers similarly at the end of that curve [1] (see Fig. 1).

The second limitation occurs in phases of the epidemic where incidence is low and prolonged [23], [17]. In these periods information is scarce and the quality of estimates will depend on how well the method of choice mediates between the little available data and its inherent assumptions [15]. We illustrate this effect using EpiEstim by considering an epidemic with a sequence of n zero incidence days. This sequence is realistic for an epidemic in its tail phase, and can precede either elimination (i.e. the epidemic goes extinct) or resurgence (i.e. a second wave of infection emerges) [24]. As n increases both $i_{\tau(s)}$ and $\lambda_{\tau(s)}$ shrink, though $i_{\tau(s)}$ decays more rapidly. For every $n \geq k$, $i_{\tau(s)} = 0$, meaning that the shape of $\mathbf{p}_{\tau(\mathbf{s})}$ is completely determined by the prior $\mathbb{P}(R_s)$.

As n increases further $\lambda_{\tau(s)}$ becomes negligible and we exactly recover our prior distribution (see Eq. (2)). This problem is acute if the window size, k, is small. Previous studies, which formally optimise k to maximise estimate reliability, found that small k is needed when inferring sharp changes in transmissibility (e.g. due to lockdowns) [14]. Moreover, as $\lambda_{\tau(s)}$ falls, the Fisher information available about R_s from I^s_{s-k+1} shrinks, leading to an inflation in estimate uncertainty and a loss of statistical identifiability [25], [23]. An analogous effect occurs in the WT method if there is low incidence across its forward-looking window [12].

While some estimate degradation is guaranteed for any R_s inference method when faced with either edge-effects or low incidence, robustness can still be improved. Edge-effects can be largely overcome by constructing $\mathbf{q_s}$. Solving the smoothing problem melds the advantages of the opposite looking windows of EpiEstim and the WT method, removing the vulnerability near the ends of I_1^t. This follows as $\mathbf{q_1} = \mathbf{r_1}$ and $\mathbf{q_t} = \mathbf{p_t}$. Further, by maximising the information used for inferring every R_s and by minimising our state model assumptions, we can ameliorate the impact of low incidence. We develop a method, termed EpiFilter, to realise these improvements in the next section.

Bayesian (forward) recursive filtering

We reformulate the causal inference problem of estimating R_s from I_1^s as an optimal Markov state filtering problem. Filtering is the term used in engineering for a general class of estimation problems aimed at optimally (usually in a mean squared error (mse) sense) inferring some hidden state causally and in real time from noisy observations [18], [13]. Given functions f_s and g_s, which describe the state (R_s in our case) space dynamics and the process of generating noisy observations (the I_s here), the filtering problem tries to construct the posterior distribution $\mathbf{p_s}$ [20], which EpiEstim approximates. The conditional mean estimate $\tilde{R}_s = \mathbb{E}[R_s \,|\, I_1^s]$ leads to the causal minimum mse of $\mathbb{E}[(R_s - \tilde{R}_s)^2]$ [20].

The famed Kalman filter [21] was the genesis of these methods. Here we focus on Bayesian recursive filters for models with noisy count observations. These generalise the Kalman filter [20] and have been successfully applied to similar problems in phylodynamics and computational biology [26], [27], [28], [29]. We reconsider our renewal model inference problem within this engineering state-

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

observation framework, as described in Eq. (3) [13].

$$R_s = f_s(R_{s-1},\, \epsilon_{s-1}), \quad I_s = g_s(R_s,\, \nu_s) \tag{3}$$

Here R_s is the hidden Markov state that we wish to infer. It dynamically depends on the previous state R_{s-1} and a noise term ϵ_{s-1} via f_s. Observation I_s is then elicited due to R_s and a noise term ν_s, according to g_s [30].

We formulate our filter under two very mild assumptions. First, we define some closed space, $\mathcal{R}$, over which R_s is valid i.e. for a given resolution m, extrema $R_{\min}$ and $R_{\max}$, and grid size $\delta = m^{-1}(R_{\max} - R_{\min})$ then $\mathcal{R} := \{R_{\min},\, R_{\min}+\delta,\, \ldots,\, R_{\max}\}$. This means R_s must take a discrete value in $\mathcal{R}$, the ith element of which is $\mathcal{R}[i]$. We formalise this in Eq. (4a).

$$\textstyle\sum_{i=1}^{m} \mathbb{P}(R_s = \mathcal{R}[i]) = 1,\; 1 \leq s \leq t \tag{4a}$$

$$R_s = R_{s-1} + (\eta\sqrt{R_{s-1}})\, \epsilon_{s-1} \tag{4b}$$

This is not restrictive since we can compute our filter for large m if needed and usually we are only interested in R_s on a coarse scale (e.g. policymakers may only want to know if $R_s \leq 1$ or not). Note that other approaches, which depend on MCMC or related sampling methods, all implicitly assume some sort of discretisation [27].

Second, we propose a linear model for f_s, as defined in Eq. (4b). There ϵ_{s-1} is a standard white noise term i.e. $\mathbb{P}(\epsilon_{s-1}) \stackrel{d}{=} \mathrm{Norm}(0, 1)$ with Norm signifying a normal distribution and η is a free parameter. We assume that a noisy linear projection of states over consecutive time-points provides a good approximation of the state trajectory. Not only is this assumption standard in engineering [20] and epidemiology [31] but it is also more flexible than the state model inherent to EpiEstim and the WT method. We scale the noise of this projection by a fraction, $\eta < 1$ of the magnitude of R_{s-1}. This allows variation but ensures R_s is *a-priori* non-negative.

Our observation model, g_s, is implicit and leads to the probability law in Eq. (1). As a result, both our observations and state models are discrete and are summarised in Fig. 1. We now define the Bayesian recursive filtering procedure, which is a main contribution of this work, and can be solved exactly, in real-time and with minimal computational effort. We adapt general recursive filtering equations from [30], [13], [22], [18], which are valid for various types of observation and state models, for our renewal model inference problem. The proof of the equations we employ can be found in these works.

Recursive filtering involves two steps: prediction and correction. The first, given in Eq. (5a), constructs a sequential prior predictive distribution, $p_s = \mathbb{P}(R_s \,|\, I_1^{s-1})$, which is informed by past data I_1^{s-1} and the last state R_{s-1}. The second step then corrects or updates this prior into the posterior distribution, $\mathbf{p_s}$, which constrains p_s using the latest observation, I_s, as in Eq. (5b).

$$p_s = \int \mathbb{P}(R_s \,|\, R_{s-1},\, I_1^{s-1})\, \mathbf{p_{s-1}}\, \mathrm{d}R_{s-1} \tag{5a}$$

$$\mathbf{p_s} \propto \mathbb{P}(I_s \,|\, R_s,\, I_1^{s-1})\, p_s \tag{5b}$$

Here $\mathbb{P}(R_s \,|\, R_{s-1},\, I_1^{s-1}) \stackrel{d}{=} \mathrm{Norm}(R_{s-1},\, \eta^2 R_{s-1})$ is the state model from Eq. (4b), $\mathbb{P}(I_s \,|\, R_s,\, I_1^{s-1})$ is the observation model from Eq. (1) and the constant of proportionality for Eq. (5b) is simply a normalising factor.

Solving Eq. (5) iteratively and simultaneously over the grid of $\mathcal{R}$ leads to our novel real-time estimate of the time varying effective reproduction number. We initialise this process with a uniform prior over $\mathcal{R}$ for $\mathbf{p_1}$ and note that p_s and $\mathbf{p_s}$ are m element vectors that sum to 1, with i^{th} term corresponding to when $R_s = \mathcal{R}[i]$. Eq. (5) forms the first half of EpiFilter, is flexible and can be adapted to many related problems [13]. A key difference between EpiFilter and the EpiEstim-type methods [11], [14] is that the latter approximate $\mathbb{P}(R_s \,|\, I_1^s)$ and $\mathbb{P}(R_s \,|\, I_1^{s-1})$ with $\mathbb{P}(R_s \,|\, I_{s-k+1}^s)$ and $\mathbb{P}(R_s)$, respectively. Estimators based on these approximations can be suboptimal, especially when data (i.e. cases) are scarce.

Bayesian (backward) recursive smoothing

While Eq. (5) provides a complete real-time solution to the causal filtering problem, it is necessarily limited at the starting edge of the incidence curve, where past data are scarce or unavailable. Further, because it does not update past estimates as new data accumulate, it cannot provide optimal retrospective estimates. Here we develop the second half of EpiFilter, which involves solving the optimal smoothing problem and hence computing $\mathbf{q_s}$. To our knowledge, smoothing has not yet been explicitly considered in infectious disease epidemiology.

We specialise the general methodology from [13], [22] to obtain the recursive smoother of Eq. (6a).

$$\mathbf{q_s} = \mathbf{p_s} \int \mathbb{P}(R_{s+1} \,|\, R_s,\, I_1^s)\, \mathbf{q_{s+1}}\, p_{s+1}^{-1}\, \mathrm{d}R_{s+1} \tag{6a}$$

$$\mathbf{q_s} \propto \mathbf{r_s}\, \mathbf{p_s}\, \mathbb{P}(R_s)^{-1}, \quad \text{if } \mathbf{r_s} \approx \mathbb{P}(R_s \,|\, I_{s+1}^t) \tag{6b}$$

We can realise Eq. (6a) exactly by taking a forward-backward algorithmic approach with $\mathbf{p_s}$ and p_{s+1} first computed from Eq. (5). Remaining terms emerge from the state space model and by noting that $\mathbf{q_t} = \mathbf{p_t}$ and iterating backwards. This approach neatly links $\mathbf{p_s}$ and $\mathbf{q_s}$. If we assume that $\mathbf{r_s}$ is reasonably approximated by $\mathbb{P}(R_s \,|\, I_{s+1}^t)$ then we can also apply the two-filter smoothing solution of [22] to get Eq. (6b). This shows how smoothing connects EpiEstim and the WT methods,

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

and explains why EpiFilter can be used for both real-time and retrospective inference. We summarise the main ingredients of the EpiFilter algorithm in Fig. 1.

The smoothed posterior $\mathbf{q_s}$ yields the conditional mean estimate $\hat{R}_s = \mathbb{E}[R_s \,|\, I_1^t]$, which is known to significantly improve on the mse of the filtered equivalent $\tilde{R}_s$ [20]. While filtering provides the minimum mse given causal knowledge, smoothing provides the minimum given all knowledge. This relationship is formal, with filtered and smoothed mse values mapping to the amount of mutual information that I_1^t provides about R_1^t [32]. Extracting the maximum information from I_1^t should engender estimates that are more robust to periods of low incidence.

While our main interest is on optimised and rigorous real-time and retrospective estimates of transmissibility, which are completely defined by $\mathbf{q_s}$, it is also important to predict future incidence, for informing epidemic preparedness plans and for validating past R_s estimates [33], [14]. We compute the filtered posterior predictive distribution as in Eq. (7) (integrals are over $\mathcal{R}$ [13].

$$\mathbb{P}(I_{s+1} \,|\, I_1^s) = \int \mathbb{P}(I_{s+1} \,|\, R_s,\, I_1^s)\, \mathbf{p_s}\, \mathrm{d}R_s \qquad (7)$$

We assume, as in [34]: $\mathbb{P}(I_{s+1} \,|\, R_s,\, I_1^s) \stackrel{d}{=} \mathrm{Pois}(\Lambda_{s+1} R_s)$. Replacing $\mathbf{p_s}$ with $\mathbf{q_s}$ yields the smoothed equivalent of Eq. (7). We will use Eq. (7) to compare EpiFilter against APEestim, which is the prediction-optimised version of EpiEstim [14]. Since predictions depend strongly on $\mathbf{p_s}$ or $\mathbf{q_s}$, optimising these distributions can be crucial, for example, to forecasting second waves of infection.

RESULTS

Improved estimation at low incidence

The estimation of time-varying effective reproduction numbers when incidence is low is seen as a key challenge limiting our understanding of transmission [17]. Periods with small counts of new cases contain little information and so present necessary statistical difficulties [23]. Here we compare EpiFilter, which allows exact inference but conditioned on $\mathcal{R}$, with EpiEstim and APEestim, which minimises the prediction error of EpiEstim by optimising its window size, at these data-poor settings [11], [14]. We use EpiEstim with a long window, as this is known to improve robustness at low incidence [24]. The assumptions and choices inherent in estimation methods become important and visible when data are scarce and can bias inference or support spurious predictions [15].

We simulate epidemics using Eq. (1)) under the serial interval distribution of Ebola virus disease available from [35]. We examine several diverse incidence trajectories (e.g. outbreaks with small sizes, large peaks or multiple modes) that all eventually decline to periods of few or no cases. We apply APEestim with optimal window k^*, EpiEstim with a long window k and EpiFilter with state noise η (see Eq. (4b)) to estimate R_s and causally predict I_s for each epidemic. For the first two methods we compute mean predictions ($\tilde{I}_{\tau(s)}$) as in [14] and estimates ($\tilde{R}_{\tau(s)}$) from Eq. (2) with $\tau(s)$ delimiting the window times used. We obtain smoothed EpiFilter estimates ($\hat{R}_s$) and causal predictions ($\tilde{I}_s$) from Eq. (6) and Eq. (7).

Our main results are in Fig. 2, which exposes how these three approaches degrade as data diminish. Each quadrant of this figure examines a different epidemic scenario. In all plots the true R_s and I_s are dashed black and dotted grey respectively, estimates are in red and predictions are blue. The central lines are conditional means and the shaded regions are their 95% confidence intervals. As our main focus is on real-time performance we do not investigate the WT method. See [11], [12] for comparisons of the WT method and EpiEstim. Our reasons for comparing to EpiEstim with a long window and APEestim follow from the Methods.

There we showed that since EpiEstim-type methods group data over a window into the past, they revert to their prior distribution as the total cases in this window becomes small. During low incidence periods, e.g. when the epidemic is dying or in its early growth phase, using long windows makes sense [24]. However, this reduces predictive accuracy as fluctuations in R_s are underfit. This is the use-case for APEestim, which optimises for prediction. The top and middle panels of each quadrant of Fig. 2 illustrate the clear consequences of these trade-offs. APEestim often selects shorter windows to better fit R_s and sequentially predict I_s. Each predicted I_s is a one-step-ahead prediction depending on the past data up to $s-1$ as in Eq. (7) (we have abused notation).

While shorter windows improve estimates and predictions the reversion to the prior distribution (seen as wide confidence intervals in Fig. 2) when I_s is low is acute. Long windows resolve this latter difficulty, showing a slower rise to the prior but at the expense of getting other parts of the epidemic trajectories incorrect. Interestingly, in the bottom panels of the quadrants of Fig. 2 we see that EpiFilter has superior performance for these types of epidemics. Not only does it give good estimates and predictions throughout the trajectory of the epidemic but its rise in uncertainty as the epidemic dies is slower and more controlled. It therefore combines the advantages of APEestim and long-window EpiEstim.

Improved estimation between epidemic waves

Maintaining robust R_s estimation at low incidence is not just statistically important. Two possible outcomes

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

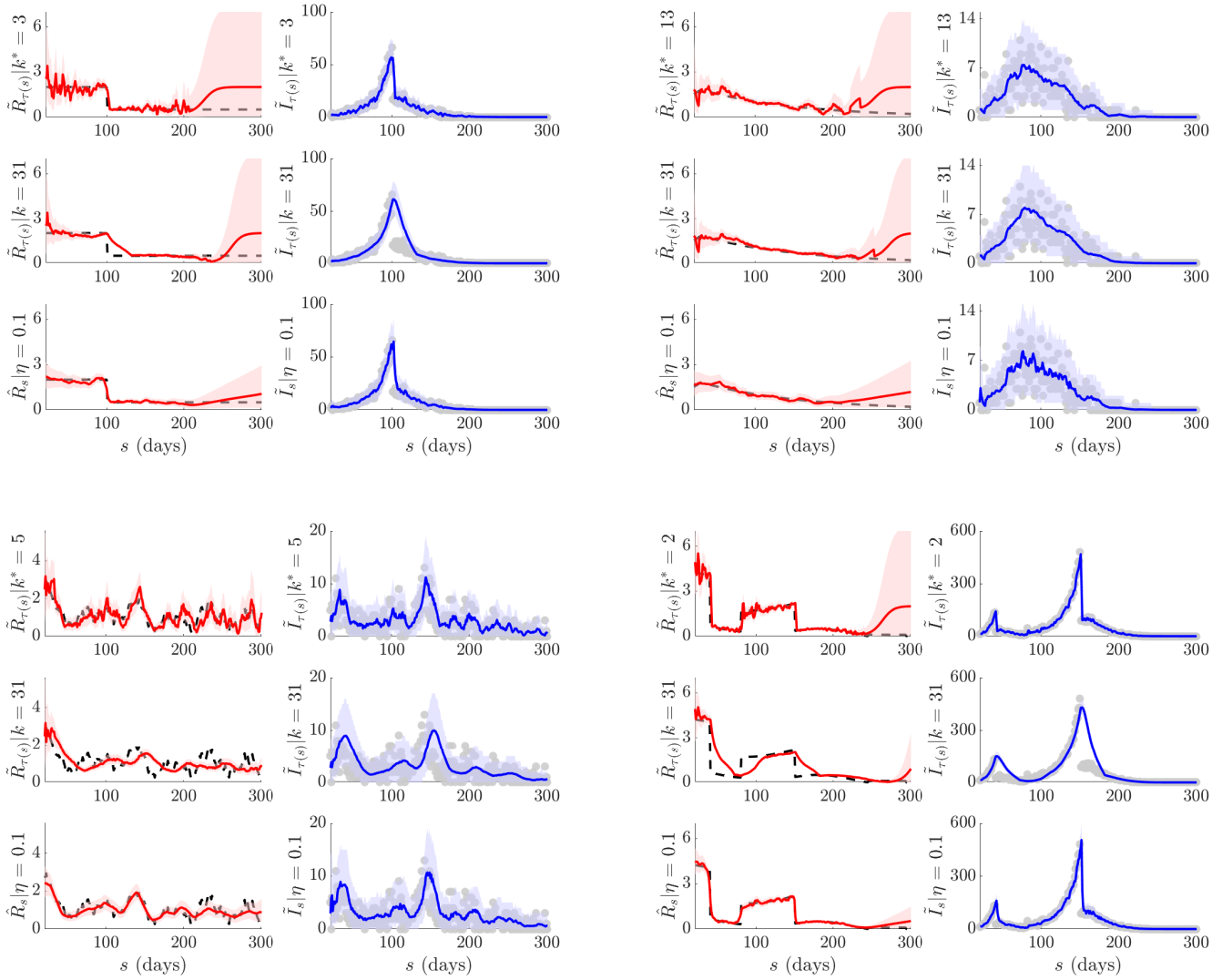

Fig. 2: **Small or dying epidemics.** We compare reproduction number estimates ($\tilde{R}_{\tau(s)}$ or $\hat{R}_s$) and causal one-step-ahead incidence predictions ($\tilde{I}_{\tau(s)}$ or $\tilde{I}_s$) from APEestim with optimised window k^* (top panels), EpiEstim with long window k (middle panels) and EpiFilter with state noise η (bottom panels). We simulate epidemics with low daily case numbers or long tails (long sequences of zero cases) using the standard renewal model (Eq. (1)). The true R_s is in black (dashed, left panels) and its corresponding I_s in grey (dots, right panels). All mean estimates or predictions are in red with 95% confidence intervals. APEestim and EpiEstim use a $\mathrm{Gam}(1, 2)$ prior distribution and EpiFilter a grid with $m = 2000$, $R_{\min} = 0.01$ and $R_{\max} = 10$. We find that EpiFilter is more robust to small incidence (better uncertainty), whereas the other approaches can quickly decay to their prior distribution.

may follow periods of small I_s: either the epidemic goes extinct, or a second wave of infection (also known as resurgence) occurs e.g. due to imports or unmonitored local transmission. Predicting which outcome is likely, rapidly and in real-time, is of major global concern as countries aim to relax interventions during the ongoing COVID-19 pandemic, while also minimising the risks of second waves [17]. As changes in reproduction numbers predate and signal corresponding variations in incidence, reliably identifying and inferring R_s trends in the trough preceding potential new peaks is crucial for preparedness and both timely and effective epidemic control [10].

Reliable estimation of R_s between epidemic waves depends on the prior assumptions of the inference method used and on how that method relies on those assumptions when data are scarce [36], [37]. Here we examine this dependence and exactly investigate this scenario, where after a low incidence period resurgence occurs. As in the

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

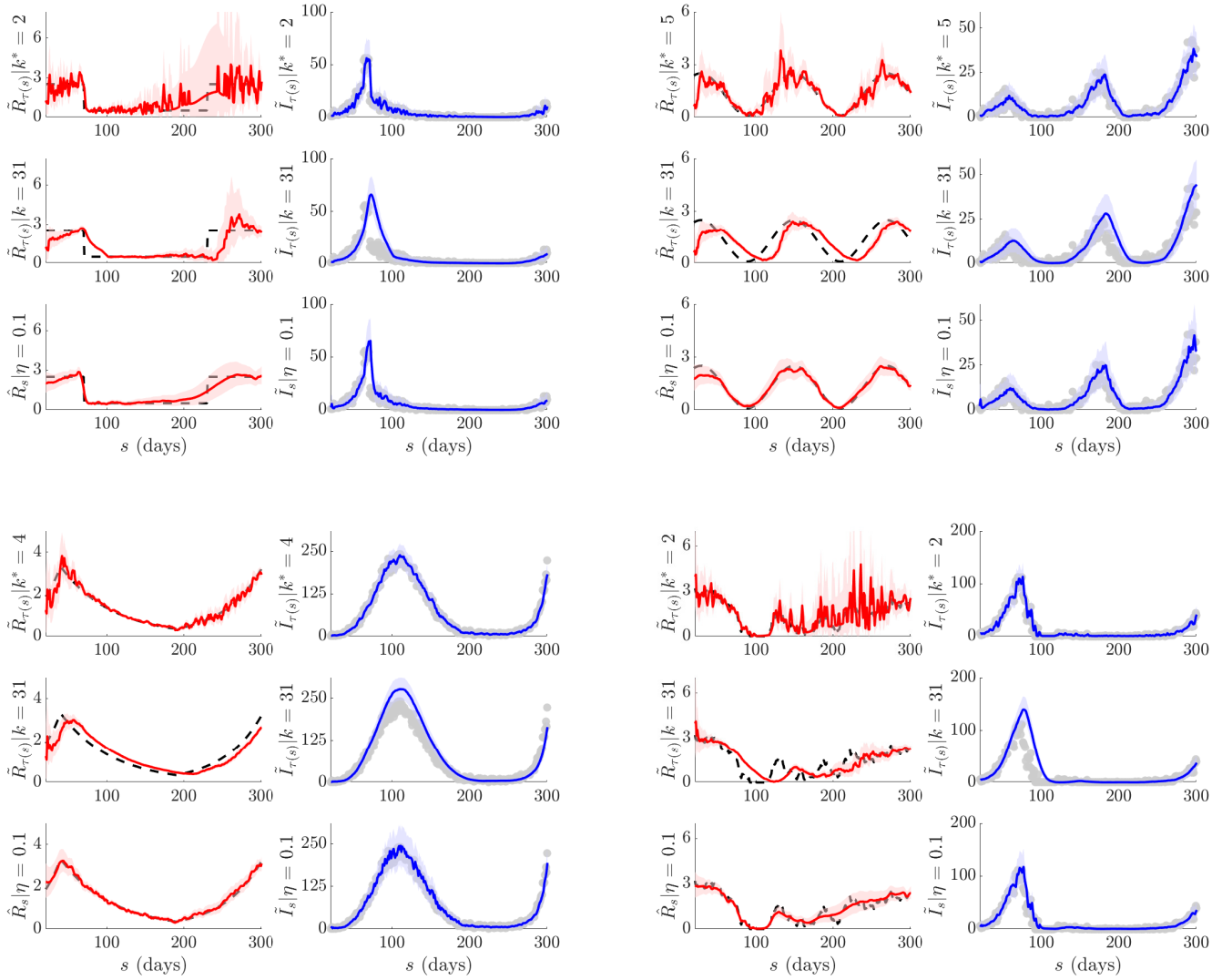

Fig. 3: **Epidemics with multiple waves.** We compare reproduction number estimates ($\tilde{R}_{\tau(s)}$ or $\hat{R}_s$) and causal one-step-ahead incidence predictions ($\tilde{I}_{\tau(s)}$ or $\tilde{I}_s$) from APEestim with optimised window k^* (top panels), EpiEstim with long window k (middle panels) and EpiFilter with state noise η (bottom panels). We simulate epidemics with multiple waves of infection using the standard renewal model (Eq. (1)). The true R_s is in black (dashed, left panels) and its corresponding I_s in grey (dots, right panels). All mean estimates or predictions are in red with 95% confidence intervals. APEestim and EpiEstim use a $\mathrm{Gam}(1, 2)$ prior distribution and EpiFilter a grid with $m = 2000$, $R_{\min} = 0.01$ and $R_{\max} = 10$. We find that due to the improved robustness of EpiFilter to low incidence regions, it is best able to negotiate troughs between epidemic peaks and hence infer re-emerging infectious dynamics.

previous section, we test APEestim with optimal window k^* (top panels), EpiEstim with a long window k (middle) and EpiFilter with state noise η (bottom panels) on four diverse epidemics, which in this case all feature multiple waves or early hints of upcoming resurgence. Simulation parameters are unchanged from the last section.

Our main results are in Fig. 3 with each quadrant specifying a different epidemic trajectory. The true R_s and I_s are in dashed-black and dotted-grey respectively. All estimates ($\tilde{R}_{\tau(s)}$ for EpiEstim and APEestim and $\hat{R}_s$ for Epifilter) are in red with corresponding causal predictions ($\tilde{I}_{\tau(s)}$ and $\tilde{I}_s$) in blue. Plots provide conditional means with 95% confidence intervals. For all second and additional wave epidemic examples we observe a similar tradeoff in performance between APEestim and long-window EpiEstim as in our previous analysis of Fig. 2. APEestim is better able to predict upcoming I_s terms and follow the overall trend in R_s than EpiEstim.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

However, the long window of EpiEstim makes it more stable during the trough between waves, which can be advantageous for understanding transmission there.

EpiFilter once again combines the advantages of the other approaches. For every scenario in Fig. 3 it provides accurate tracking of changes in R_s with stable confidence intervals. Concurrently, its predictive performance rivals that of APEestim. EpiFilter is therefore a powerful tool when dealing with second-wave scenarios. We also find that the $\eta = 0.1$ parameter value seems to be an all-purpose heuristic, meaning that usage of EpiFilter can be somewhat simpler than APEestim and EpiEstim. The superior capability of EpiFilter in these scenarios likely results from its minimalistic assumptions (see Eq. (4)) and its increased information extraction. We next test our method on empirical incidence data.

COVID-19 in New Zealand and H1N1 in the USA

The previous sections confirmed EpiFilter as a powerful inference and prediction tool, especially in data-poor conditions, using simulated epidemics. We now confront our method with empirical data from the 1918 H1N1 influenza pandemic in Baltimore (USA) [38] and the ongoing COVID-19 pandemic in New Zealand (up to 17 Aug 2020) [39]. The H1N1 dataset has been well-studied and so we first use this to benchmark EpiFilter. We clean this data by applying a 5-day moving average filter as recommended in [38]. Previous work [11] analysed this data with EpiEstim and found that sensible R_s estimates result when a weekly window ($k = 7$) is applied.

A recent study, which re-examined this epidemic with APEestim [14], shows that, while EpiEstim with $k = 7$ provides stable estimates for this epidemic, it is a poor causal predictor of the incidence data. Instead, an optimised window of 2 days ($k^* = 2$) yields good predictions but the resulting R_s estimates are noisy. We reproduce the estimates ($\tilde{R}_{\tau(s)}$) and predictions ($\tilde{I}_{\tau(s)}$) from both studies in Fig. 4 and compare them against EpiFilter with $\eta = 0.1$ ($\hat{R}_s$ and $\tilde{I}_s$). The cleaned H1N1 incidence is in dotted grey, the critical $R_s = 1$ line in dashed black and all R_s estimates and I_s predictions (with 95% confidence intervals) are in red and blue respectively.

Top and middle rows of Fig. 4 show the mentioned trade-off between estimate stability and prediction accuracy. The bottom row confirms the power of EpiFilter. Our R_s estimates are of comparable stability to that of EpiEstim at $k = 7$, yet our prediction fidelity matches APEestim. Our improved inference again benefits from using more information (i.e. the backward pass in Fig. 1) and making less restrictive prior assumptions. We see the latter from the R_s confidence intervals over $40 \leq s \leq 60$. There EpiEstim seems overconfident, and this results in a rigid overestimation of incidence. However, EpiFilter mediates its confidence to a level similar to APEestim.

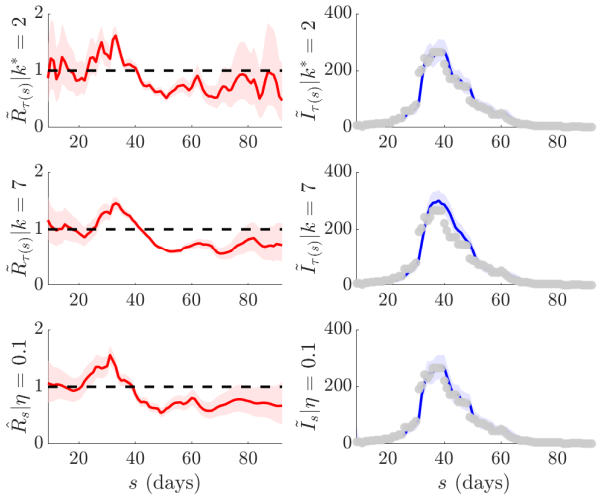

Fig. 4: **H1N1 influenza transmission estimates for Baltimore (1918).** We compare APEestim (top), EpiEstim with recommended weekly window (middle) (both with $\mathrm{Gam}(1, 2)$ prior) and EpiFilter (with $m = 2000$, $\eta = 0.1$, $R_{\min} = 0.01$ and $R_{\max} = 10$) on the H1N1 influenza dataset of [38]. We use a 5-day moving average filter, as in [38], to remove known sampling biases. Estimates of R_s and corresponding 95% confidence intervals are in red. One-step-ahead predictions of I_s (with 95% confidence) are in blue with the actual incidence in grey. We find that EpiFilter combines the benefits of the other methods, achieving good estimates and predictions.

We explore COVID-19 transmission patterns in New Zealand using incidence data up to 17 Aug 2020 from [39]. New Zealand presents an insightful case study because officials combined swift lockdowns with intensive testing to achieve and sustain very low incidence levels that some believed could engender local elimination of COVID-19 [40]. However, an upsurge in cases in early Aug inspired concerns about a second wave (which led to new interventions and is why we do not consider data beyond 17 Aug). Here we investigate time-varying transmission in New Zealand to see if this uptick suggests that the epidemic was resurfacing in mid Aug. We believe smoothing can confer important inferential advantages in exactly these types of low incidence scenarios.

We make the common assumptions that case under-reporting is constant and that reporting delays are negligible [11]. While this makes our investigation somewhat naive, we consider this to be reasonable given the intense surveillance that New Zealand employed [41]. We also use the serial interval distribution from [42] and do not

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

explicitly distinguish imported from local cases in our analysis. The latter could bias our study [24] but our focus is on demonstrating differences between filtering and smoothing on R_s trends instead of specifying the precise value of R_s during this period. We plot the results of our exploration in Fig. 5. More involved analyses can be performed by pre-processing the data for known delays or case ascertainment fractions as in [12], [6].

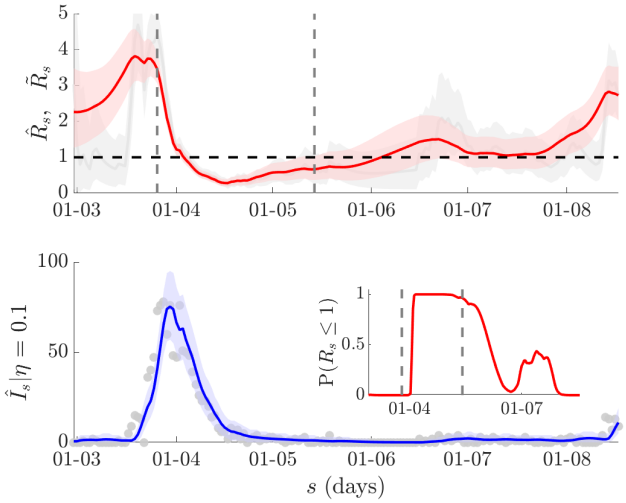

Fig. 5: **COVID-19 transmission in New Zealand.** We compute smoothed and filtered reproduction number estimates, $\hat{R}_s$ and $\tilde{R}_s$ respectively, from the COVID-19 incidence curve for New Zealand (available at [39]). We use EpiFilter with $m = 2000$, $\eta = 0.1$, $R_{\min} = 0.01$ and $R_{\max} = 10$ with a uniform prior over the grid $\mathcal{R}$. The top panel shows conditional mean estimates and 95% confidence intervals for $\hat{R}_s$ (red) and $\tilde{R}_s$ (grey). Vertical lines indicate the start and end of lockdown, a major intervention that was employed to halt transmission. The additional 'future' information used in smoothing has a notable effect. The bottom panel provides smoothed one-step-ahead predictions $\hat{I}_s$ (blue, with 95% confidence) of the actual reported cases I_s (grey). The inset gives the probability of $\hat{R}_s \leq 1$ and lockdown times.

We apply EpiFilter and obtain causally filtered ($\tilde{R}_s$, grey) and smoothed ($\hat{R}_s$, red) conditional mean estimates together with their 95% confidence intervals. These are in the top panel of Fig. 5 and computed from Eq. (5) and Eq. (6) respectively. The times of lockdown and release are the grey vertical dashed lines. Interestingly, we see a notable difference in the quality of inference between $\tilde{R}_s$ and $\hat{R}_s$. The former, as expected, is unreliable at the beginning of the incidence curve and features wider uncertainty and unclear trends. The smoothed estimate, by using both forward and backward-looking data largely overcomes these issues and clarifies transmission trends.

Our smoothed analysis suggests that R_s has resurged and supports the reimplementation of measures around 14 Aug. In the bottom panel of Fig. 5 we provide predictions (which are from the smoothed $\hat{R}_s$ hence the notation $\hat{I}_s$) and their 95% confidence intervals in blue against the reported incidence from [39] in dotted grey. These confirm that $\hat{R}_s$ reasonably describes the data. In the inset we show that the $\mathbb{P}(R_s \leq 1)$ also supports the resurgence hypothesis. Last, we note that transmission decreased considerably (and rise in $\mathbb{P}(R_s \leq 1)$) across the lockdown period.

Discussion

Estimating time-varying trends in effective reproduction number R_s reliably and in real-time is an important and popular problem in infectious disease epidemiology [5]. As the COVID-19 pandemic has continued to unfold, the interest in solving this problem has only elevated [8]. Initially, the focus was on characterising how R_s might respond to interventions such as lockdowns and social distancing [10]. However, as COVID-19 has progressed and countries have entered the controlled phase, concerns have shifted to trying to understand how existing interventions might be relaxed with minimum risk.

The literature on intervention exit strategies is, however, still in development, and several challenges remain to modelling the transmission behaviour of epidemics following prolonged periods of low incidence [17], [24]. While changes in R_s over this period are not the only analytic for assessing this risk, they do provide a key real-time diagnostic since upticks in R_s generally precede corresponding rises in incidence. Unfortunately, current approaches to estimating R_s under these conditions tend to be underpowered or prior-constrained [11], [23].

Here we re-examined existing methodology for inferring R_s from an engineering perspective. We observed that two of the most useful and popular inference approaches, EpiEstim [11] and the WT method [2], only capitalise on a portion of the data available, deeming either upcoming or past incidence informative (see Fig. 1). This informative portion is directly controlled by prior assumptions on the speed of possible R_s changes, which are often characterised by a window of size k. Other methods are also known to apply similarly strong change-point assumptions on R_s [23], [9]. When data are scarce these assumptions can control inference.

In control engineering a common problem, known as filtering, involves optimally (in a mse sense) estimating hidden Markov states, causally in real-time, from noisy and uncertain observations [18]. A related problem termed smoothing provides accompanying and optimal

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

retrospective inference [13]. By reinterpreting R_s as a Markov state (Eq. (4)) observed through a noisy renewal process (Eq. (1)) and defining R_s on a predetermined grid $\mathcal{R}$, we were able to construct exact filtering (Eq. (5)) and smoothing (Eq. (6)) solutions. This led to EpiFilter, which is our central contribution.

Generally, filtering and smoothing can be involved and require sophisticated sequential Monte Carlo techniques [30]. However, because our system is low dimensional and as we make only minimal assumptions about R_s, modelling it as a simple diffusion, we were able to solve these problems exactly and without complex sampling algorithms [27]. Our solutions are computationally simple and deterministic (i.e. precisely reproducible given the same data). While EpiFilter is limited by one free parameter η, our $\eta = 0.1$ heuristic seems statistically justified by its good causal one-step-ahead predictive performance (see Appendix and Fig. 2 and Fig. 3) [14].

Importantly, EpiFilter is able to look both forward and backward through the incidence data, and so maximise the information extracted [32]. This property means it combines the advantages of both EpiEstim and the WT method (see Fig. 1) and largely negates their edge-effect issues [12], offering both real-time and retrospective R_s estimations that are robust. We illustrated the advantages of EpiFilter by comparing it to EpiEstim and APEestim on many simulated examples with diverse periods of low incidence and epidemic resurgences (Fig. 2 and Fig. 3). Interestingly, we found EpiFilter able to track arbitrary changes in transmission without requiring any preset change-points or window sizes or any tuning of η.

Moreover, EpiFilter was notably better at negotiating periods of low incidence, offering a graceful degradation to its prior without sacrificing predictive accuracy. When incidence is low, it is usually beneficial to use long windows with EpiEstim [24]. This keeps R_s estimates reasonably stable. However, it often leads to poor predictions [14]. APEestim, which optimises window size for prediction fidelity, showed that in many of the simulated scenarios short windows are necessary for describing transmission patterns. Consequently, we have a tradeoff between estimate robustness and prediction accuracy.

We found that EpiFilter overcomes this tradeoff, concurrently achieving good estimates and predictions. We confirmed these advantages on empirical data from the H1N1 pandemic of 1918 (see Fig. 4). The ability to integrate the benefits of long window EpiEstim and APEestim make EpiFilter particularly useful for investigating resurgence after a period of low incidence, as it is better able to forewarn of increasing case numbers and can more speedily infer upticks in transmission (see Fig. 3). Capitalising on these properties, we performed a naive exploratory analysis of COVID-19 in New Zealand and found evidence of a potential second wave of infection after a prolonged period of control (see Fig. 5).

Balancing the assumptions inherent to a model against the data it is applied on, to produce reliable inference is a non-trivial problem that is still under active investigation in several fields [15], [36], [37]. EpiFilter, by maximising the information extracted from the incidence data and minimising its state space model assumptions, appears to strike this balance. Consequently, it performs strongly on a wide range of problems, including those involving sparse data, where other methods might struggle. Given its demonstrated advantages, straightforward formulation and theoretical underpinning, we hope that EpiFilter will be useful as a diagnostic tool for reliably warning about second waves of infection as COVID-19 unfolds.

Acknowledgments

This work is jointly funded under grant reference MR/R015600/1 by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP2 programme supported by the European Union.

References

[1] C. Fraser, "Estimating individual and household reproduction numbers in an emerging epidemic," *PLOS One*, vol. 8, p. e758, 2007.

[2] J. Wallinga and P. Teunis, "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures," *Am. J. Epidemiol*, vol. 160, no. 6, pp. 509–16, 2004.

[3] H. Nishiura and G. Chowell, "The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends," in *Mathematical and statistical estimation approaches in epidemiology*, pp. 103–21, Springer, 2009.

[4] R. Anderson and R. May, *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1991.

[5] S. Cauchemez, P. Boelle, G. Thomas, *et al.*, "Estimating in real time the efficacy of measures to control emerging communicable diseases," *Am. J. Epidemiol*, vol. 164, no. 6, pp. 591–7, 2006.

[6] T. Churcher, J. Cohen, N. Ntshalintshali, *et al.*, "Measuring the path toward malaria elimination," *Science*, vol. 344, no. 6189, pp. 1230–32, 2014.

[7] WHO Ebola Response Team, "Ebola virus disease in West Africa – the first 9 months of the epidemic and forward projections," *N. Engl. J. Med*, vol. 371, no. 16, pp. 1481–95, 2014.

[8] S. Bhatia, A. Cori, K. Parag, *et al.*, "Short-term forecasts of COVID-19 deaths in multiple countries.."

[9] S. Flaxman, S. Mishra, A. Gandy, *et al.*, "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe," *Nature*, vol. 584, pp. 257–261, 2020.

[10] X. Hao, S. Cheng, D. Wu, *et al.*, "Reconstruction of the full transmission dynamics of COVID-19 in Wuhan," *Nature*, vol. 584, pp. 420–4, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

[11] A. Cori, N. Ferguson, C. Fraser, *et al.*, "A new framework and software to estimate time-varying reproduction numbers during epidemics," *Am. J. Epidemiol*, vol. 178, no. 9, pp. 1505–12, 2013.

[12] K. Gostic, L. McGough, E. Baskerville, *et al.*, "Practical considerations for measuring the effective reproductive number, rt," *medRxiv*, vol. 2020.06.18.20134858v2, 2020.

[13] S. Sarrka, *Bayesian Filtering and Smoothing*. Cambridge, UK: Cambridge University Press, 2013.

[14] K. Parag and C. Donnelly, "Using information theory to optimise epidemic models for real-time prediction and estimation," *PLOS Comput. Biol*, vol. 16, no. 7, p. e1007990, 2020.

[15] K. Parag, O. Pybus, and C. Wu, "Are skyline plot-based demographic estimates overly dependent on smoothing prior assumptions?," *BioRxiv*, vol. 920215, 2020.

[16] T. Britton, T. House, A. Lloyd, *et al.*, "Five challenges for stochastic epidemic models involving global transmission," *Epidemics*, vol. 10, pp. 54–7, 2015.

[17] R. Thompson, D. Hollingsworth, V. Isham, *et al.*, "Key questions for modelling COVID-19 exit strategies," *Proc. R. Soc. B*, vol. 287, no. 1932, p. 20201405, 2020.

[18] K. Astrom and R. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*. New Jersey: Princeton University Press, 2008.

[19] J. Wallinga and M. Lipsitch, "How generation intervals shape the relationship between growth rates and reproductive numbers," *Proc. R. Soc. B*, vol. 274, pp. 599–604, 2007.

[20] D. Snyder and M. Miller, *Random Point Processes in Time and Space*. Springer-Verlag, 2 ed., 1991.

[21] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng*, vol. 82, pp. 35–45, 1960.

[22] B. Anderson and I. Rhodes, "Smoothing algorithms for nonlinear finite-dimensional systems," *Stochastics*, vol. 9, pp. 139–65, 1983.

[23] K. Parag and C. Donnelly, "Adaptive estimation for epidemic renewal and phylogenetic skyline models," *Syst. Biol*, vol. syaa035, 2020.

[24] K. Parag, C. Donnelly, R. Jha, *et al.*, "An exact method for quantifying the reliability of end-of-epidemic declarations in real time," *medRxiv*, vol. 2020.07.13.20152082, 2020.

[25] T. Rothenberg, "Identification in parametric models," *Econometrica*, vol. 39, no. 3, pp. 577–91, 1971.

[26] K. Parag and O. Pybus, "Optimal point process filtering and estimation of the coalescent process," *J. Theor. Biol*, vol. 421, pp. 153–67, 2017.

[27] K. Parag and O. Pybus, "Exact bayesian inference for phylogenetic birth-death models," *Bioinformatics*, vol. 34, no. 21, pp. 3638–45, 2018.

[28] K. Parag and G. Vinnicombe, "Point Process Analysis of Noise in Early Invertebrate Vision," *PLOS Comput. Biol*, vol. 13, no. 10, p. e1005687, 2017.

[29] C. Zechner and H. Koeppl, "Uncoupled analysis of stochastic reaction networks in fluctuating environments," *PLOS Comput. Biol*, vol. 10, no. 12, p. e1003942, 2014.

[30] Z. Chen, "Bayesian filtering: From Kalman filters to particle filters, and beyond," *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

[31] L. Allen, "A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis," *Infect. Dis. Model*, vol. 2, pp. 128–142, 2017.

[32] D. Guo, D. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Info. Theo*, vol. 51, no. 4, pp. 1261–82, 2005.

[33] S. Funk, A. Camacho, A. Kucharski, *et al.*, "Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the western area region of Sierra Leone, 2014-15," *PLOS Comput. Biol*, vol. 15, no. 2, p. e1006785, 2019.

[34] P. Nouvellet, A. Cori, T. Garske, *et al.*, "A simple approach to measure transmissibility and forecast incidence," *Epidemics*, vol. 22, pp. 29–35, 2018.

[35] M. Van Kerkhove, A. Bento, H. Mills, *et al.*, "A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making," *Sci. Data*, vol. 2, p. 150019, 2015.

[36] E. Volz and X. Didelot, "Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance," *Syst. Biol*, vol. 67, no. 4, pp. 719–28, 2018.

[37] J. Faulkner, A. Magee, B. Shapiro, *et al.*, "Horseshoe-based Bayesian nonparametric estimation of effective population size trajectories," *Biometrics*, pp. 1–14, 2020.

[38] W. Frost and E. Sydenstricker, "Influenza in Maryland: preliminary statistics of certain localities," *Public Health Rep*, vol. 34, pp. 491–504, 1919.

[39] WHO, "WHO coronavirus disease (COVID-19) dashboard," 2020.

[40] S. Cousins, " New Zealand eliminates COVID-19," *Lancet*, vol. 395, p. 1474, 2020.

[41] M. Roser, H. Ritchie, E. Ortiz-Ospina, *et al.*, "Coronavirus pandemic (COVID-19)," 2020.

[42] N. Ferguson, D. Laydon, G. Nedjati-Gilani, *et al.*, "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID- 19 mortality and healthcare demand," tech. rep., Imperial College London, 2020.

APPENDIX

Validating the recursive filter

Having constructed EpiFilter in Eq. (3) – Eq. (6) we now investigate its mechanics and performance. First we illustrate how the choice of the grid resolution m allows our approach to achieve estimates at differing levels of accuracy. Our recursive implementation is also known as a grid smoother, since its estimates are optimal (in mse) but conditioned on the grid $\mathcal{R}$ [27]. Left panels of Fig. A1 show how the smoothed estimator $\hat{R}_s$ (red together with 95% confidence bounds) becomes more accurate with increased resolution. The true R_s is in black. Generally, we find that no visible improvement occurs above $m \sim 10^3$. Wider grid ranges $R_{\max} - R_{\min}$ require larger m for the same resolution.

We also test the behaviour under different choices of the process or state noise parameter η. This parameter is vaguely similar to the window size parameter, k, in EpiEstim approaches, but is easier to select. The right panels of Fig. A1, show how a smaller η might lead to underfitting, while large η (where we would not consider $\eta \geq 1$ given Eq. (4b)) could promote overfitting. However, while the optimal choice of window in EpiEstim varies with the form of the true R_s (e.g. large k is needed for stable R_s time-series and small k for strongly fluctuating ones [14]), we find that a fixed $\eta = 0.1$ works well across many and diverse epidemic scenarios.

We illustrate this further in Fig. A2. There we compare R_s estimates from APEestim (left panels, denoted $\hat{R}_{\tau(s)}$

medRxiv preprint doi: https://doi.org/10.1101/2020.09.14.20194589; this version posted September 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

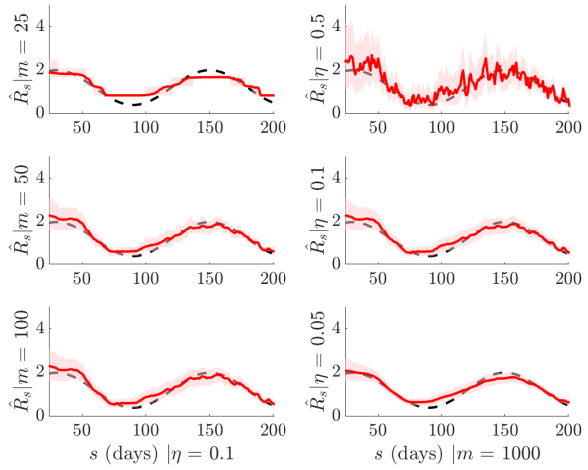

Fig. A1: **Filtering mechanics.** We examine how EpiFilter behaves under various choices of process noise parameter η and grid resolution m. Left panels indicate that increasing m leads to improved estimation, as expected, but that even at coarse settings the estimate behaves sensibly. Right panels show that reducing η sharpens confidence intervals but can potentially underfit. We find that $\eta = 0.1$ serves as a good general heuristic. All simulations are at $R_{\min} = 0.01$ and $R_{\max} = 10$, black is the true R_s (a seasonal epidemic) and red is the EpiFilter mean estimate ($\hat{R}_s$) with 95% confidence intervals.

with optimal window size k) [14], [11] to EpiFilter (right panels, denoted $\hat{R}_s$) at $\eta = 0.1$ and $m = 2000$ for several simulated epidemic scenarios. We generate epidemics according to the standard renewal model of Eq. (1) using the serial interval distribution for Ebola virus disease [35]. The true R_s is in black with respective estimates in red (together with 95% confidence intervals). To keep comparisons fair we set the usual gamma prior distribution over R_s in APEestim to $\mathrm{Gam}(1, 2)$ so that its domain is essentially bounded by $R_{\max}$.

EpiFilter has broadly comparable inference performance to APEestim, despite our using a single η value across the entire range of simulated scenarios. This is in stark contrast to the discordant optimal k values, which are shown in Fig. A2. Moreover, our implementation is fast (it executes in seconds) and unlike other window-less approaches does not require any choices of the timing or number of change-points i.e. at this single η it easily handles unchanging trajectories, seasonal variations and rapidly or gradually fluctuating transmission.

The bottom panels of Fig. A2 illustrate the central advantage of EpiFilter. There the associated incidence values (not shown) are small. While this causes APEestim inference to destabilise (see Methods for explanation), EpiFilter offers more robust and usable estimates. We explore this in detail in the main text. Note that in none of these scenarios does the true R_s have either the state model of Eq. (4b) or the sliding-window relationship in EpiEstim. Consequently, these simulations also indicate robustness of the different inherent state assumptions in both approaches to moderate model mismatch.

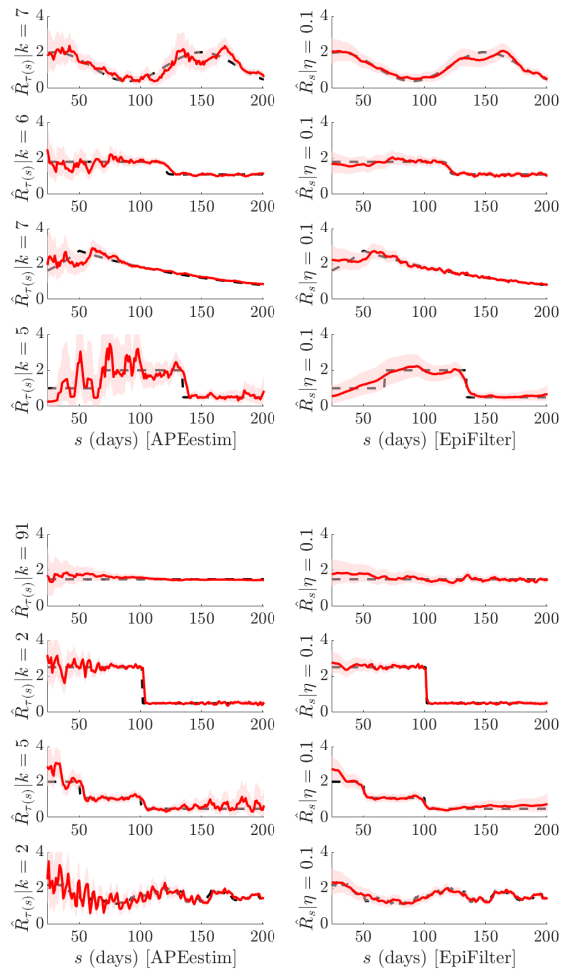

Fig. A2: **Validation against APEestim.** We compare estimates from EpiFilter (right panels) with optimised ones from EpiEstim (called APEestim, left panels [14]) over various epidemic scenarios, simulated using Eq. (1). We use $m = 2000$, $R_{\min} = 0.01$ and $R_{\max} = 10$. To be fair we use a $\mathrm{Gam}(1, 2)$ prior distribution with APEestim, which sums to 1 by $R_{\max}$. Here k is the optimal window choice in APEestim and $\eta = 0.1$ is fixed for EpiFilter inferences. True R_s is in black and conditional mean estimates with 95% confidence intervals are in red. Both approaches offer comparable performance but EpiFilter is more stable especially early in the epidemic when incidence is low.