

# RtEstim: Effective reproduction number estimation with trend filtering

Jiapeng Liu<sup>1\*</sup>, Zhenglun Cai<sup>2</sup>, Paul Gustafson<sup>1</sup>, Daniel J. McDonald<sup>1</sup>

**1** Department of Statistics, The University of British Columbia, Vancouver, British Columbia, Canada

**2** Centre for Health Evaluation and Outcome Sciences, The University of British Columbia, Vancouver, British Columbia, Canada

\* [jiapeng.liu@stat.ubc.ca](mailto:jiapeng.liu@stat.ubc.ca)

## Abstract

To understand the transmissibility and spread of infectious diseases, epidemiologists turn to estimates of the effective reproduction number. While many estimation approaches exist, their utility may be limited. Challenges of surveillance data collection, model assumptions that are unverifiable with data alone, and computationally inefficient frameworks are critical limitations for many existing approaches. We propose a discrete spline-based approach **RtEstim** that solves a convex optimization problem—Poisson trend filtering—using the proximal Newton method. It produces a locally adaptive estimator for effective reproduction number estimation with heterogeneous smoothness. **RtEstim** remains accurate even under some process misspecifications and is computationally efficient, even for large-scale data. The implementation is easily accessible in a lightweight R package [rtestim](#).

## Author summary

Effective reproduction number estimation presents many challenges due to data collection, modelling assumptions, and computational burden. Such limitations hinder the accurate estimation of the effective reproduction number. Our motivation is to

develop a model that produces accurate estimates, is robust to model misspecification, and is straightforward to use and computationally efficient, even for large counts and long time periods. We propose a convex optimization model with an  $\ell_1$  trend filtering penalty. It couples accurate estimation of the effective reproduction number with desired smoothness. We solve the optimization using the proximal Newton method, which converges rapidly and is numerically stable. Our software, conveniently available in the R package `RtEstim`, can produce estimates in seconds for incidence sequences with hundreds of observations. These estimates are produced for a sequence of tuning parameters and can be selected using a built-in cross validation procedure.

## 1 Introduction

The effective reproduction number at time  $t$  is defined to be the expected number of secondary infections produced by a primary infection throughout the course of the entire infection if conditions remain the same at the specific time. The instantaneous reproduction number, specifically, is a type of effective reproduction number focusing on the transmission at a specific timepoint [1]. It is a key quantity for understanding infectious disease dynamics including the potential size of an outbreak and the required stringency of control measures [2,3]. Tracking the time series of this quantity is useful for understanding whether or not future infections are likely to increase or decrease from the current state [4]. Let  $\mathcal{R}(t)$  denote the effective reproduction number at time  $t$ . Practically, as long as  $\mathcal{R}(t) < 1$ , infections will decline gradually, eventually resulting in a disease-free equilibrium, whereas when  $\mathcal{R}(t) > 1$ , infections will continue to increase, resulting in endemic equilibrium. While  $\mathcal{R}(t)$  is fundamentally a continuous time quantity, it can be related to data only at discrete points in time  $t = 1, \dots, n$ . This sequence of effective reproduction numbers over time is not observable, but, nonetheless, is easily interpretable and retrospectively describes the course of an epidemic. Therefore, a number of procedures exist to estimate  $\mathcal{R}_t$  from different types of observed incidence data such as cases, deaths, or hospitalizations, while relying on various domain-specific assumptions, e.g., [5–8]. Importantly, accurate estimation of effective reproduction numbers relies heavily on the quality of the available data, and, due to the limitations of data collection, such as underreporting and lack of standardization, estimation

methodologies rely on various assumptions to compensate. Because model assumptions  
may not be easily verifiable from data alone, it is also critical for any estimation  
procedure to be robust to model misspecification.

Many existing approaches for effective reproduction number estimation are Bayesian:  
they estimate the posterior distribution of  $\mathcal{R}_t$  conditional on the observations. One of  
the first such approaches is the software **EpiEstim** [9], described in [10]. This method is  
prospective, in that it uses only observations available up to time  $t$  in order to estimate  
 $\mathcal{R}_t$  for each  $i = 1, \dots, t$ . An advantage of **EpiEstim** is its straightforward statistical  
model: new incidence data follows the Poisson distribution conditional on past incidence  
combined with the conjugate gamma prior distribution for  $\mathcal{R}_t$  with fixed  
hyperparameters. Additionally, the serial interval distribution, the distribution of the  
period between onsets of primary and secondary infections in a population, is fixed and  
known. For this reason, **EpiEstim** requires little domain expertise for use, and it is  
computationally fast. [11] modified this method to distinguish imported cases from local  
transmission and simultaneously estimate the serial interval distribution. [12] further  
extended **EpiEstim** by using “reconstructed” daily incidence data to handle irregularly  
spaced observations. Recently, [13] proposed a Bayesian latent variable framework,  
**EpiNow2** [14], which leverages incident cases, deaths or other available streams  
simultaneously along with allowing additional delay distributions (incubation period  
and onset to reporting delays) in modelling. [15] proposed an extension that handles  
missing data by imputation followed by a truncation adjustment. These modifications  
are intended to increase accuracy at the most recent (but most uncertain) timepoints,  
to aid policymakers. [16] also proposed a Bayesian approach, **EpiFilter**, based on the  
(discretized) Kalman filter and smoother. **EpiFilter** also estimates the posterior of  $\mathcal{R}_t$   
given a Gamma prior and Poisson distributed incident cases. Compared to **EpiEstim**,  
however, **EpiFilter** estimates  $\mathcal{R}_t$  retrospectively using all available incidence data both  
before and after time  $t$ , with the goal of being more robust in low-incidence periods. [17]  
proposed a Bayesian P-splines approach, **EpiLPS**, that assumes negative Binomial  
distributed observations. [18] also proposed a Bayesian model estimated with particle  
filtering to incorporate spatial structures. Bayesian approaches estimate the posterior  
distribution of the effective reproduction numbers and possess the advantage that  
credible intervals may be easily computed. They incorporate the prior knowledge on

parameters to modelling. Some techniques are used to eliminate the power of prior  
54 parameters on the posterior estimates to make the estimates more plausible, e.g., [11]  
55 assumes a relatively large prior mean of  $\mathcal{R}_t$  (appreciably larger than 1), and if the  
56 estimate is less than 1, researchers will know it is a direct result from data, instead of  
57 the choice of prior parameters. Some Bayesian approaches, however, are  
58 computationally expensive, since they require more intensive computational routines,  
59 especially when observed data sequences are long or hierarchical structures are complex,  
60 e.g., [13]. While, some Bayesian methods with more efficient structures, especially the  
61 ones with only conjugate priors, can be computationally efficient, e.g., [10]. Below, we  
62 compare our method to the aforementioned Bayesian models, **EpiEstim**, **EpiLPS**,  
63 **EpiFilter**, and **EpiNow2**.  
64

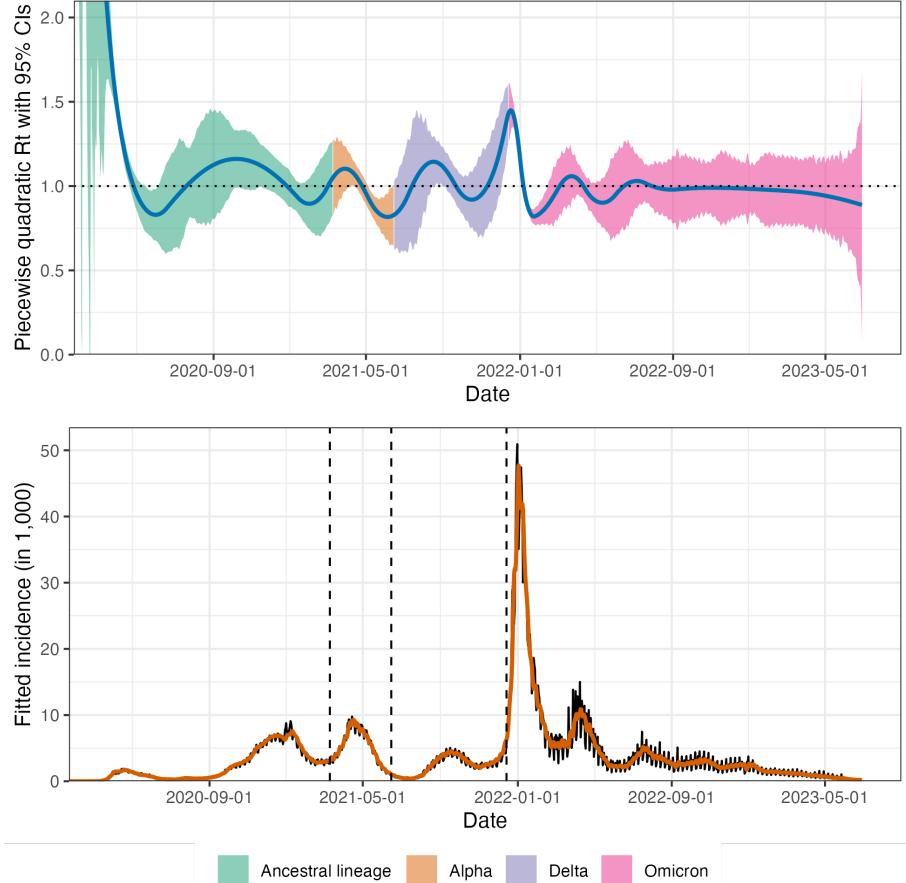
There are also frequentist approaches for  $\mathcal{R}_t$  estimation. [19] proposed regularizing  
65 the smoothness of  $\mathcal{R}_t$  through penalized regression with second-order temporal  
66 regularization, additional spatial penalties, and with Poisson loss. [20] extended this  
67 procedure by introducing another penalty on outliers. [21] proposed a spline-based  
68 model relying on the assumption of exponential-family distributed incidence. [22]  
69 estimates  $\mathcal{R}_t$  while monitoring the time-varying level of overdispersion. There are other  
70 spline-based approaches such as [23, 24], autoregressive models with random effects [25]  
71 that are robust to low incidence, and generalized autoregressive moving average  
72 (GARMA) models [26] that are robust to measurement errors in incidence data.  
73

We propose an effective (instantaneous, specifically) reproduction number estimator,  
74 called **RtEstim** that requires only incidence data. Our model makes the conditional  
75 Poisson assumption, similar to much of the prior work described above, but is  
76 empirically more robust to misspecification. This estimator is defined by a convex  
77 optimization problem with Poisson loss and  $\ell_1$  penalty on the temporal evolution of  
78  $\log(\mathcal{R}_t)$  to impose smoothness over time. As a result, **RtEstim** generates discrete  
79 splines, and the estimated curves (in logarithmic space) appear to be piecewise  
80 polynomials of an order selected by the user. Importantly, the estimates are locally  
81 adaptive, meaning that different time ranges may possess heterogeneous smoothness.  
82 Because we penalize the logarithm of  $\mathcal{R}_t$ , we naturally accommodate the positivity  
83 requirement, in contrast to related methods, can handle large or small incidence  
84 measurements, and are automatically (reasonably) robust to outliers without additional  
85

constraints. A small illustration using three years of Covid-19 case data in Canada is  
86 shown in Fig 1 [27]. We get the viral evolution and spread information from the  
87 duotang project provided by [28] and compute the probabilities of having each variant  
88 at each timepoint using multinomial logistic regression. The variant with the highest  
89 probabilities is deemed to dominate the specific timepoint. There are four dominant  
90 variants throughout the epidemic, Ancestral lineage, Alpha, Delta, and Omicron over  
91 time. We use the estimated serial interval distributions of the four variants by Xu et  
92 al. [29], specifically Ancestral lineage (mean= 5.1, sd= 4.0), Alpha (mean= 3.5,  
93 sd= 4.5), Delta (mean= 3.5, sd= 2.9), and Omicron (mean= 3.0, sd= 2.1) variants.  
94

While our approach is straightforward and requires little domain knowledge for  
95 implementation, we also implement a number of refinements. We use a proximal  
96 Newton method to solve the convex optimization problem along with warm starts to  
97 produce estimates efficiently, typically in a matter of seconds, even for long sequences of  
98 data. In a number of simulation experiments, we show empirically that our approach is  
99 more accurate than existing methods at estimating the true effective reproduction  
100 numbers and robust under multiple settings of the misspecification of incidence  
101 distribution, serial interval distribution, and the order of graphical curvature.  
102

The manuscript proceeds as follows. We first introduce the methodology of RtEstim  
103 including the renewal equation and the development of Poisson trend filtering estimator.  
104 We explain how this method could be interpreted from the Bayesian perspective,  
105 connecting it to previous work in this context. We provide illustrative experiments  
106 comparing our estimator to other Bayesian alternatives. We then apply our RtEstim on  
107 the Covid-19 pandemic incidence in Canada and the 1918 influenza pandemic incidence  
108 in the United States. Finally, we conclude with a discussion of the advantages and  
109 limitations of our approach and describe practical considerations for effective  
110 reproduction number estimation.  
111



**Fig 1.** A demonstration of effective reproduction number estimation by RtEstim and the corresponding predicted incident cases for the Covid-19 epidemic in Canada during the period from January 23, 2020 to June 28, 2023. In the top panel, the blue curve is the estimated piecewise quadratic  $\mathcal{R}_t$  and the colorful ribbon is the corresponding 95% confidence band. The ribbon is dyed by four colors representing the variants whose serial interval distributions are used to estimate  $\mathcal{R}_t$ . The y-axis is truncated for a better illustration; the estimated  $\mathcal{R}_t$  is larger than 2 at the first 54 timepoints, and the largest one, which is at the first timepoint, is 10.77. The black curve in the bottom panel is the observed Covid-19 daily confirmed cases, and the orange curve is the predicted incident cases corresponding to the estimated  $\mathcal{R}_t$ . The three vertical dashed lines represent the beginning of a new dominant variant.

## 2 Methods

### 2.1 Renewal model for incidence data

The effective reproduction number  $\mathcal{R}(t)$  is defined to be the expected number of secondary infections at time  $t$  produced by a primary infection sometime in the past. To make this precise, denote the number of new infections at time  $t$  as  $y(t)$ . Then the total primary infectiousness can be written as  $\eta(t) := \int_0^\infty p(i)y(t-i)di$ , where  $p(i)$  is the

probability that a new secondary infection is the result of a primary infection that  
 occurred  $i$  time units in the past. The effective reproduction number is then given as  
 the value that equates

$$\mathbb{E}[y(t) | y(j), j < t] = \mathcal{R}(t)\eta(t) = \mathcal{R}(t) \int_0^\infty p(i)y(t-i)di, \quad (1)$$

otherwise known as the renewal equation. The period between primary and secondary infections is exactly the generation time of the disease, but given real data, observed at discrete times (say, daily), this delay distribution must be discretized into contiguous time intervals, say,  $(0, 1], (1, 2], \dots$ . It results in the sequence  $\{p_i\}_0^\infty$  corresponding to observations  $y_t$  and yields the discretized version of Eq (1),

$$\mathbb{E}[y_t | y_1, \dots, y_{t-1}] = \mathcal{R}_t\eta_t = \mathcal{R}_t \sum_{i=1}^{\infty} p_i y_{t-i}. \quad (2)$$

Many approaches to estimating  $\mathcal{R}_t$  rely on Eq (2) as motivation for their procedures, among them, **EpiEstim** [10] and **EpiFilter** [16].

In most cases, it is safe to assume that infectiousness disappears beyond  $\tau$  timepoints ( $p(i) = 0$  for  $i > \tau$ ), resulting in the truncated integral of the generation interval distribution  $\int_0^\tau p(i)di = 1$ . Generation time, however, is usually unobservable and tricky to estimate, so common practice is to approximate it by the serial interval: the period between the symptom onsets of primary and secondary infections. If the infectiousness profile after symptom onset is independent of the incubation period (the period from the time of infection to the time of symptom onset), then this approximation is justifiable: the serial interval distribution and the generation interval distribution share the same mean. However, other properties may not be similarly shared, and, in general, the generation interval distribution is a convolution of the serial interval distribution with the distribution of the difference between independent draws from the delay distribution from infection to symptom onset. See, for example, [1] for a fuller discussion of the dangers of this approximation. Nonetheless, treating these as interchangeable is common [10, 30] and doing otherwise is beyond the scope of this work. Additionally, we assume that the generation interval (and, therefore, the serial interval), is constant over time  $t$ . That is, the probability  $p(i)$  depends only on the gap between

primary and secondary infections and not on the time  $t$  when the secondary infection 144  
occurs. For our methods, we will assume that the serial interval can be accurately 145  
estimated from auxiliary data (say by contact tracing, or previous epidemics) and we 146  
will take it as fixed, as is common in existing studies, e.g., [10, 19, 20]. 147

The renewal equation in Eq (2) relates observable data streams (incident cases) 148  
occurring at different timepoints to the effective reproduction number given the serial 149  
interval. The fact that it depends only on the observed incident counts makes it 150  
reasonable to estimate  $\mathcal{R}_t$ . However, data collection idiosyncrasies can obscure this 151  
relationship. Diagnostic testing targets symptomatic individuals, omitting 152  
asymptomatic primary infections which can lead to future secondary infections. Testing 153  
practices, availability, and uptake can vary across space and time [31, 32]. Finally, 154  
incident cases as reported to public health are subject to delays due to laboratory 155  
confirmation, test turnaround times, and eventual submission to public health [33]. For 156  
these reasons, reported cases are lagging indicators of the course of the pandemic. 157  
Furthermore, they do not represent the actual number of new infections that occur on a 158  
given day, as indicated by exposure to the pathogen. The assumptions described above 159  
(constant serial interval distribution, homogenous mixing, similar susceptibility and 160  
social behaviours, etc.) are therefore consequential. That said, Eq (2) also provides some 161  
comfort about deviations from these assumptions. If  $y_t$  is scaled by a constant (in time) 162  
describing the reporting ratio, then it will cancel from both sides. Similar arguments 163  
mean that even if such a scaling varies in time, as long as it varies slowly relative to the 164  
set of  $p_i$  that are larger than 0, Eq (2) will be a reasonably accurate approximation, so 165  
that  $\mathcal{R}_t$  can still be estimated well from reported incidence data. Finally, even a sudden 166  
change in reporting ratio, say from  $c_1$  for  $i = 1, \dots, t_1$  to  $c_2$  for  $i > t_1$  would only result 167  
in large errors for  $t$  in the neighbourhood of  $t_1$  (where the size of this neighbourhood is 168  
again determined by the effective support of  $\{p_i\}$ ). This robustness to certain types of 169  
data reporting issues partially justifies using Eq (2) to calculate  $\mathcal{R}_t$ . 170

## 2.2 Poisson trend filtering estimator 171

We use the daily confirmed incident cases  $y_t$  on day  $t$  to estimate the observed infectious 172  
cases under the model that  $y_t$ , given previous incident cases  $y_{t-1}, \dots, y_1$  and a constant 173

serial interval distribution, follows a Poisson distribution with mean  $\Lambda_t$ . That is,

174

$$y_t \mid y_1, \dots, y_{t-1} \sim \text{Poisson}(\Lambda_t), \text{ where } \Lambda_t = \mathcal{R}_t \sum_{i=1}^{t-1} p_i y_{t-i} = \mathcal{R}_t \eta_t. \quad (3)$$

Given a history of  $n$  confirmed incident counts  $\mathbf{y} = (y_1, \dots, y_n)^\top$ , our goal is to estimate  $\mathcal{R}_t$ . A natural approach is to maximize the likelihood, producing the maximum likelihood estimator (MLE):

$$\begin{aligned} \hat{\mathcal{R}} &= \underset{\mathcal{R} \in \mathbb{R}_+^n}{\operatorname{argmax}} \mathbb{P}(\mathcal{R} \mid \mathbf{y}, \mathbf{p}) = \underset{\mathcal{R} \in \mathbb{R}_+^n}{\operatorname{argmax}} \prod_{t=1, \dots, n} \frac{(\mathcal{R}_t \eta_t)^{y_t} \exp\{-\mathcal{R}_t \eta_t\}}{y_t!} \\ &= \underset{\mathcal{R} \in \mathbb{R}_+^n}{\operatorname{argmin}} \sum_{t=1}^n \mathcal{R}_t \eta_t - y_t \log(\mathcal{R}_t \eta_t). \end{aligned} \quad (4)$$

This optimization problem, however, is easily seen to yield a one-to-one correspondence between the observations and the estimated effective reproduction number, i.e.,  $\hat{\mathcal{R}}_t = y_t / \eta_t$ , so that the estimated sequence  $\hat{\mathcal{R}}$  will have no significant smoothness.

178

179

180

The MLE is an unbiased estimator of the true parameter  $\mathcal{R}_t$ , but unfortunately has high variance: changes in  $y_t$  result in proportional changes in  $\hat{\mathcal{R}}_t$ . To avoid this behaviour, and to match the intuition that  $\mathcal{R}_t \approx \mathcal{R}_{t-1}$ , we advocate enforcing smoothness of the effective reproduction numbers. This constraint will decrease the estimation variance, and hopefully lead to more accurate estimation of  $\mathcal{R}$ , as long as the smoothness assumption is reasonable. Smoothness assumptions are common (see e.g., [16] or [1]), but the type of smoothness assumed is critical. [9] imposes smoothness indirectly by estimating  $\mathcal{R}_t$  with moving windows of past observations. The Kalman filter procedure of [16] would enforce in  $\ell_2$ -smoothness ( $\int_0^n (\hat{\mathcal{R}}''(t))^2 dt < C$  for some  $C$ ), although the computational implementation results in  $\hat{\mathcal{R}}$  taking values over a discrete grid. [20] produces piecewise linear  $\hat{\mathcal{R}}_t$ , which turns out to be closely related to a special case of our methodology. Smoother estimated curves will provide high-level information about the entire epidemic, obscuring small local changes in  $\mathcal{R}(t)$ , but may also remove the ability to detect large sudden changes, such as those resulting from lockdowns or other major containment policies.

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

To enforce smoothness of  $\hat{\mathcal{R}}_t$ , we add a trend filtering penalty to Eq (5) [34–37].

196

Because  $\mathcal{R}_t > 0$ , we explicitly penalize the divided differences (discrete derivatives) of

197

neighbouring values of  $\log(\mathcal{R}_t)$ . Let  $\theta := \log(\mathcal{R}) \in \mathbb{R}^n$ , so that  $\Lambda_t = \eta_t \exp(\theta_t)$ , and 198  
 $\log(\eta_t \mathcal{R}_t) = \log(\eta_t) + \theta_t$ . For evenly spaced incident case, we write our estimator as the 199  
solution to the optimization problem 200

$$\hat{\mathcal{R}} = \exp(\hat{\theta}) \quad \text{where} \quad \hat{\theta} = \underset{\theta \in \mathbb{R}^n}{\operatorname{argmin}} \eta^\top \exp(\theta) - \mathbf{y}^\top \theta + \lambda \|D^{(k+1)}\theta\|_1, \quad (5)$$

where  $\exp(\cdot)$  applies elementwise and  $\|\mathbf{a}\|_1 := \sum_{i=1}^n |a_i|$  is the  $\ell_1$  norm. Here, 201  
 $D^{(k+1)} \in \mathbb{Z}^{(n-k-1) \times n}$  is the  $(k+1)^{\text{th}}$  order divided difference matrix for any 202  
 $k \in \{0, \dots, n-1\}$ .  $D^{(1)} \in \{-1, 0, 1\}^{(n-1) \times n}$  is the divided difference matrix for  $k=0$ . 203  
It is a sparse matrix with diagonal band of the form: 204

$$D^{(1)} = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix}. \quad (6)$$

$D^{(k+1)}$  for  $k \geq 1$  is defined recursively as  $D^{(k+1)} := D^{(1)}D^{(k)}$ , where 205  
 $D^{(1)} \in \{-1, 0, 1\}^{(n-k-1) \times (n-k)}$  takes the form defined in Eq (6). More description on 206  
the recursive definition of divided difference matrix for trend filtering can be found 207  
in [35, 36]. 208

The tuning parameter (hyperparameter)  $\lambda$  balances data fidelity with desired 209  
smoothness. When  $\lambda = 0$ , the problem in Eq (5) reduces to the MLE in Eq (4). Larger 210  
tuning parameters privilege the regularization term and yield smoother estimates. 211  
Finally, there exists  $\lambda_{\max}$  such that any  $\lambda \geq \lambda_{\max}$  will result in  $D^{(k+1)}\hat{\theta} = 0$  and  $\hat{\theta}$  will 212  
be the Kullback-Leibler projection of  $\mathbf{y}$  onto the null space of  $D^{(k+1)}$  (see 213  
subsection 2.3). 214

The solution to Eq (5) will result in piecewise polynomials, specifically called 215  
discrete splines. For example, 0<sup>th</sup>-degree discrete splines are piecewise constant, 216  
1<sup>st</sup>-degree curves are piecewise linear, and 2<sup>nd</sup>-degree curves are piecewise quadratic. 217  
For  $k \geq 1$ ,  $k^{\text{th}}$ -degree discrete splines are continuous and have continuous discrete 218  
differences up to degree  $k-1$  at the knots. This penalty results in more flexibility 219  
compared to the homogeneous smoothness that is created by the squared  $\ell_2$  norm. 220  
Using different orders of divided differences result in estimated effective reproduction 221

numbers with different smoothness properties.

For unevenly-spaced data, the spacing between neighbouring parameters varies with the time between observations, and thus, the divided differences must be adjusted by the times that the observations occur. Given observation times  $\mathbf{x} = (x_1, \dots, x_n)^\top$ , for  $k \geq 1$ , define a  $k^{\text{th}}$ -order diagonal matrix

$$X^{(k)} = \text{diag} \left( \frac{k}{x_{k+1} - x_1}, \frac{k}{x_{k+2} - x_2}, \dots, \frac{k}{x_n - x_{n-k}} \right). \quad (7)$$

Letting  $D^{(\mathbf{x},1)} := D^{(1)}$ , then for  $k \geq 1$ , the  $(k+1)^{\text{th}}$ -order divided difference matrix for unevenly spaced data can be created recursively by  $D^{(\mathbf{x},k+1)} := D^{(1)} X^{(k)} D^{(\mathbf{x},k)}$ . No adjustment is required for  $k = 0$ .

Due to the penalty structure, this estimator is locally adaptive, meaning that it can potentially capture local changes such as the initiation of control measures. [19, 20] considered only the 2<sup>nd</sup>-order divided difference of  $\mathcal{R}_t$  rather than its logarithm. In comparison to their work, our estimator (i) allows for arbitrary degrees of temporal smoothness and (ii) avoids the potential numerical issues of penalizing/estimating positive real values. Furthermore, as we will describe below, our procedure is computationally efficient for estimation over an entire sequence of penalty strengths  $\lambda$  and provides methods for choosing how smooth the final estimate should be.

### 2.3 Solving over a sequence of tuning parameters

We can solve the Poisson trend filtering estimator over an arbitrary sequence of  $\lambda$  that produces different levels of smoothness in the estimated curves. We consider a candidate set of  $M$   $\lambda$ -values,  $\boldsymbol{\lambda} = \{\lambda_m\}_{m=1}^M$ , that is strictly decreasing.

Let  $D := D^{(k+1)}$  for simplicity in the remainder of this section. As  $\lambda \rightarrow \infty$ , the penalty term  $\lambda \|D\theta\|_1$  dominants the Poisson objective, so that minimizing the objective is asymptotically equivalent to minimizing the penalty term, which results in  $\|D\theta\|_1 = 0$ . In this case, the divided differences of  $\theta$  with order  $k+1$  is always 0, and thus,  $\theta$  must lie in the null space of  $D$ , that is,  $\theta \in \mathcal{N}(D)$ . The same happens for any  $\lambda$  beyond this threshold, so define  $\lambda_{\max}$  to be the smallest  $\lambda$  that produces  $\theta \in \mathcal{N}(D)$ . It turns out that this value can be written explicitly as  $\lambda_{\max} = \|(D^\dagger)^\top (\eta - y)\|_\infty$ , where  $D^\dagger$  is the (left) generalized inverse of  $D$  satisfying  $D^\dagger D = I$  and  $\|a\|_\infty := \max_{i=1}^n \{|a_i|\}$  is the

infinity norm. Therefore, we use  $\lambda_1 = \lambda_{\max}$  and then choose the minimum  $\lambda_M$  to be  
 $r\lambda_{\max}$  for some  $r \in (0, 1)$  (typically  $r = 10^{-5}$ ). Given any  $M \geq 3$ , we generate a  
sequence of  $\lambda$  values to be equally spaced on the log-scale between  $\lambda_1$  and  $\lambda_M$ .  
250  
251  
252

To compute the sequence efficiently, the model is estimated sequentially by visiting  
each component of  $\boldsymbol{\lambda}$  in order. The estimates produced for a larger  $\lambda$  are used as the  
initial values (warm starts) for the next smaller  $\lambda$ . By solving through the entire  
sequence of tuning parameters, we have a better chance to achieve a better trade-off  
between bias and variance, and accordingly, improved accuracy relative to procedures  
examining one fixed value of  $\lambda$ .  
253  
254  
255  
256  
257  
258

## 2.4 Choosing a final $\lambda$

259

We estimate model accuracy over the candidate set through  $K$ -fold cross validation  
(CV) to choose the best tuning parameter. Specifically, we divide  $\mathbf{y}$  (except the first and  
last observations) roughly evenly and randomly into  $K$  folds, estimate  $\mathcal{R}_t$  for all  $\boldsymbol{\lambda}$   
leaving one fold out, and then predict the held-out observations. Model accuracy can be  
measured by multiple metrics such as mean squared error  $\text{MSE}(\hat{y}, y) = n^{-1}\|\hat{y} - y\|_2^2$  or  
mean absolute error  $\text{MAE}(\hat{y}, y) = n^{-1}\|\hat{y} - y\|_1$ , but we prefer to use the (average)  
deviance, to mimic the likelihood in Eq (4):  
260  
261  
262  
263  
264  
265  
266

$D(y, \hat{y}) = n^{-1} \sum_{i=1}^n 2(y_i \log(y_i) - y_i \log(\hat{y}_i) - y_i + \hat{y}_i)$ , with the convention that  
 $0 \log(0) = 0$ . Note that for any  $K$  and any  $M$ , we will end up estimating the model  
 $(K + 1)M$  times rather than once.  
267  
268  
269

## 2.5 Approximate confidence bands

270

We also provide empirical confidence bands of the estimators with approximate  
coverage. Consider the related estimator  $\tilde{\mathcal{R}}_t$  defined as  
271  
272

$$\tilde{\mathcal{R}} = \exp(\tilde{\theta}) \quad \text{where} \quad \tilde{\theta} = \underset{\theta \in \mathbb{R}^n}{\operatorname{argmin}} \eta^\top \exp(\theta) - \mathbf{y}^\top \theta + \lambda \|D\theta\|_2^2. \quad (8)$$

Let  $\tilde{\mathbf{y}} = \eta \circ \tilde{\mathcal{R}}$ , and then it can be shown (for example, Theorem 2 in [38]) that an  
estimator for  $\text{Var}(\tilde{\mathbf{y}})$  is given by  $(\text{diag}(\tilde{\mathbf{y}}^{-2}) + \lambda D^\top D)^\dagger$ . Finally, an application of the  
delta method shows that  $\text{Var}(\tilde{\mathbf{y}}_t)/\eta_t^2$  is an estimator for  $\text{Var}(\tilde{\mathcal{R}}_t)$  for each  $t = 1, \dots, n$ .  
We therefore use  $(\text{diag}(\hat{\mathbf{y}}^{-2}) + \lambda D^\top D)_t^\dagger/\eta_t^2$  as an estimator for  $\text{Var}(\hat{\mathcal{R}}_t)$ . An  
273  
274  
275  
276

approximate  $(1 - \alpha)\%$  confidence interval then can be written as  $\widehat{\mathcal{R}}_t \pm s_t \times T_{\alpha/2, n-\text{df}}$ ,  
where  $s_t$  is the square-root of  $\text{Var}(\widehat{\mathcal{R}}_t)$  for each  $t = 1, \dots, n$  and  $\text{df}$  is the number of  
changepoints in  $\widehat{\theta}$  plus  $k + 1$ .  
277  
278  
279

## 2.6 Bayesian perspective

  
280

Unlike many other methods for  $\mathcal{R}_t$  estimation, our approach is frequentist rather than  
Bayesian. Nonetheless, it has a corresponding Bayesian interpretation: as a state-space  
model with Poisson observational noise, autoregressive transition equation of degree  
 $k \geq 0$ , e.g.,  $\theta_{t+1} = 2\theta_t - \theta_{t-1} + \varepsilon_{t+1}$  for  $k = 1$ , and Laplace transition noise  
 $\varepsilon_{t+1} \sim \text{Laplace}(0, 1/\lambda)$ . Compared to **EpiFilter** [16], we share the same observational  
assumptions, but our approach has a different transition noise. **EpiFilter** estimates  
the posterior distribution of  $\mathcal{R}_t$ , and thus it can provide credible interval estimates as  
well. Our approach produces the maximum *a posteriori* estimate via an efficient convex  
optimization, obviating the need for MCMC sampling. But the associated confidence  
bands are created differently.  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290

# 3 Results

  
291

Implementation of our approach is provided in the R package **rtestim**. All  
computational experiments are conducted on the Cedar cluster provided by Compute  
Canada with R 4.3.1. The R packages used for simulation and real-data application are  
**EpiEstim** 2.2-4, **EpiLPS** 1.2.0, and **rtestim** 0.0.4.  
292  
293  
294  
295

## 3.1 Synthetic experiments

  
296

### 3.1.1 Problem design

  
297

We simulate four scenarios of the time-varying effective reproduction number, intended  
to mimic different epidemics. The first two scenarios are rapidly controlled by  
intervention, where the  $\mathcal{R}(t)$  consists of one discontinuity and two segments. Scenario 1  
has constant  $\mathcal{R}(t)$  before and after an intervention, while Scenario 2 grows exponentially,  
then decays. The other two scenarios are more complicated, where more waves are  
involved. Scenario 3 has four linear segments with three discontinuities, which reflect  
298  
299  
300  
301  
302  
303

the effect of an intervention, resurgence to rapid transmission, and finally suppression of  
 the epidemic. Scenario 4 involves sinusoidal waves throughout the epidemic. The first  
 three scenarios and the last scenario are motivated by [16] and [17] respectively. We  
 name the four scenarios as (1) *piecewise constant*, (2) *piecewise exponential*, (3)  
*piecewise linear*, and (4) *periodic* lines or curves respectively.  
304  
305  
306  
307  
308

In all cases, the times of observation are regular, and epidemics are of length  
 $n = 300$ . Specifically, in Scenario 1,  $\mathcal{R}_t = 2, 0.8$  before and after  $t = 120$ . In Scenario 2,  
 $\mathcal{R}_t$  increases and decreases exponentially with rates 0.01, 0.005 pre and post  $t = 100$ . In  
Scenario 3,  $\mathcal{R}_t$  is piecewise linear with four discontinuous segments following  
309  
310  
311  
312

$$\begin{aligned}\mathcal{R}(t) = & \left(2.5 - \frac{0.5}{74}(t-1)\right) \mathbf{1}_{[1,76)}(t) + \left(0.8 - \frac{0.2}{74}(t-76)\right) \mathbf{1}_{[76,151)}(t) \\ & + \left(1.7 + \frac{0.3}{74}(t-151)\right) \mathbf{1}_{[151,226)}(t) + \left(0.9 - \frac{0.4}{74}(t-226)\right) \mathbf{1}_{[226,300]}(t),\end{aligned}\quad (9)$$

where  $\mathbf{1}_A(t) = 1$ , if  $t \in A$ , and  $\mathbf{1}_A(t) = 0$  otherwise. In Scenario 4,  $\mathcal{R}_t$  is realization of  
the continuous, periodic curve generated by the function  
313  
314

$$\mathcal{R}(t) = 0.2((\sin(\pi t/12) + 1) + (2\sin(5\pi t/12) + 2) + (3\sin(5\pi t/6) + 3)), \quad (10)$$

evaluated at equally spaced points  $t \in [0, 10]$ . These  $\mathcal{R}_t$  scenarios are illustrated in Fig 2.  
We compute the expected incidence  $\Lambda_t$  using the renewal equation, and generate the  
incident infections from the Poisson distribution  $y_t \sim \text{Pois}(\Lambda_t)$ . To verify the  
performance of our model under the violation of the distributional assumption, we also  
generate incident cases using the negative Binomial distribution with dispersion size  
 $\rho = 5$ , i.e.,  $y_t \sim \text{NB}(\mu = \Lambda_t, \rho = 5)$ . The probability mass function can be written as  
315  
316  
317  
318  
319  
320

$$p(y_t | \Lambda_t, \rho) = \frac{\Gamma(y_t + \rho)}{\Gamma(\rho)y_t!} \left(\frac{\rho}{\rho + \Lambda_t}\right)^\rho \left(\frac{\Lambda_t}{\rho + \Lambda_t}\right)^{y_t}, \quad (11)$$

where  $y_t = 0, 1, 2, \dots, t = 1, \dots, n$  and the variance is  $\sigma_t^2 = \Lambda_t + \frac{\Lambda_t^2}{\rho}$ , which follows the  
definition in [17]. The variance equals the mean appended by another term of the  
squared mean over a relatively small dispersion size, so the squared mean can dominate  
the term and result in sufficient overdispersion level in the generated samples. We  
visualize the dispersion level (using the ratio of standard deviation over mean) of  
321  
322  
323  
324  
325

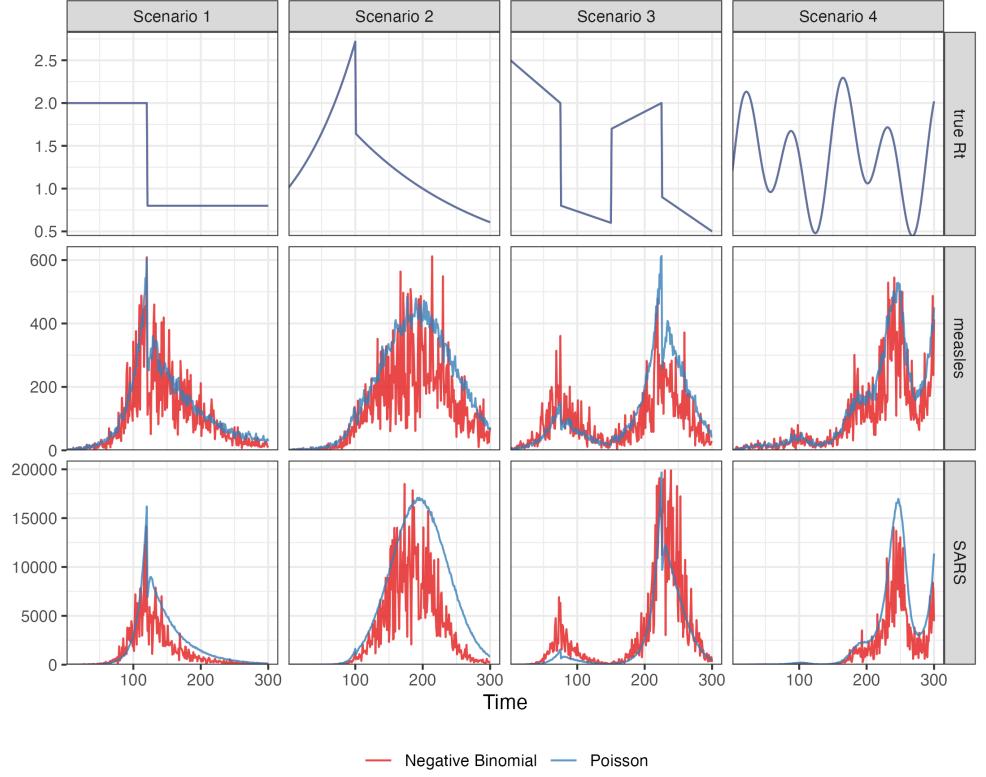
synthetic epidemics in Figure A.2.1 in Appendix for Poisson and negative Binomial incidence separately to have a better idea on the overdispersion level of the negative Binomial distributions (with size 5) compared the Poisson distributions with the same mean.

We use serial interval (SI) distributions of measles (with mean 14.9 and standard deviation 3.9) at Hagelloch, Germany in 1861 [39] and SARS (with mean 8.4 and standard deviation 3.8) at Hong Kong in 2003 [40], inspired by [10], to generate synthetic epidemics. We initialize all epidemics with  $y_1 = 2$  cases and generate for  $t = 2, \dots, 300$ . The synthetic measles epidemics have smaller incident cases in general, and the SARS epidemics have larger incidence. The intuition behind this is a smaller mean of serial interval with a similar standard deviation leads to an averaged shorter period of the onsets of symptoms between the primary and secondary infected individuals, which results in a greater growth of incidence within the same period of time. We also consider shorter flu epidemics with 50 timepoints with piecewise linear  $\mathcal{R}_t$  considering both incidence distributional assumptions. The motivation is to compare our method and other aforementioned alternatives with EpiNow2 which takes much longer time to converge for long epidemics (taking almost 2 hours to converge for a measles epidemic with 300 timepoints) than other methods. Besides using the correct SI distributions to estimate  $\mathcal{R}_t$ , we also consider the scenarios where SI is mildly or majorly misspecified. More details on experimental settings and results for shorter epidemics and misspecification of SI distributions will be given in Sections A.2.1 and A.3 in Appendix respectively.

For each problem setting (including a SI distribution, an  $\mathcal{R}_t$  scenario, and an incidence distribution), we generate 50 random samples, resulting in 800 total synthetic epidemics. An example of measles and SARS epidemics for each effective reproduction number scenario with an incidence distribution is displayed in Fig 2.

### 3.1.2 Algorithm design

We compare `RtEstim` to `EpiEstim`, `EpiLPS`, and `EpiFilter`. `EpiEstim` estimates the posterior distribution of the effective reproduction number given a Gamma prior and Poisson distributed observations over a trailing window, under the assumption that the effective reproduction number is constant during that window. A larger window



**Fig 2.** The effective reproduction numbers for four  $\mathcal{R}_t$  scenarios (*in the top row*). The sample measles (*in the middle row*) and SARS (*in the bottom row*) incident cases drawn from Poisson (*in blue curves*) or negative Binomial (*in red curves*) distribution across 4  $\mathcal{R}_t$  scenarios (*in four columns respectively*).

averages out more fluctuations, leading to smoother estimates, whereas, a shorter sliding window is more responsive to sudden spikes or declines. We tried the weekly sliding window, as well as a monthly window. However, since neither considerably outperforms the other across all scenarios, we defer the monthly results to the supplementary document. **EpiLPS** is another Bayesian approach that estimates P-splines based on the Laplace approximation to the conditional posterior with negative Binomial likelihood. **EpiFilter** is also a Bayesian approach that smooths  $\mathcal{R}_t$  at each timepoint given all observed incidence, improved upon the filtering methods that filter  $\mathcal{R}_t$  given the observations prior to and on time  $t$ .

We apply **RtEstim** with four degrees, piecewise constant  $k = 0$ , piecewise linear  $k = 1$ , piecewise quadratic  $k = 2$ , and piecewise cubic  $k = 3$  polynomials, to solve all settings. We run 10-fold cross validation (CV) to choose the best tuning parameter from the candidate set of size 50, i.e.,  $\lambda = \{\lambda_1, \dots, \lambda_{50}\}$ , for long epidemics, and 5-fold CV

for short epidemics. Specifically, we divide all samples (except the first and last entries) 370  
into, e.g., 10 folds evenly and randomly, and build models on each subset of samples by 371  
leaving a fold out using each choice of the tuning parameter. We select the tuning 372  
parameter that gives the lowest averaged deviance between the estimated incidence and 373  
the observed samples averaged over all folds. 374

We specify the parameters of the alternative methods, using the ones which were 375  
applied to their own experimental settings and so are deemed as the “best” tuned ones, 376  
in our experiments. Due to the limitations of the software implementations that some 377  
necessary hyperparameters are not allowed to choose, we can only specify the supported 378  
ones. This limitation does not only restrict the choices on tuning parameters which 379  
impacts the model fitting, but also those for optimization (as mentioned above). For 380  
example, in EpiLPS, one must specify the number of basis functions as well as 5 prior 381  
parameters. EpiFilter needs a grid of possible  $\mathcal{R}_t$  values, along with a fixed value for 382  
the diffusion noise, and a prior value for  $\mathcal{R}_t$ . EpiNow2 has many prior parameters 383  
(depending on the particular model used) and does not compute a Bayes Factor or other 384  
model selection criterion. Had we attempted to tune some or all of these parameters, we 385  
would need to implement cross validation from scratch for each, as none of these provide 386  
ways to choose them. Here are the hyperparameters used in modelling for each 387  
alternative method. We consider both weekly and monthly sliding windows in 388  
EpiEstim, 40 basis functions in EpiLPS with the NelderMead method to maximize the 389  
hyperparameter posterior distribution. We input 2000 grid size in EpiFilter with 0.1 390  
diffusion noise and uniform prior on  $\mathcal{R}_t$  with mean 1/2000, and use the smoothed  $\mathcal{R}_t$  391  
given all observed incidence as the final estimates. 392

For the  $\mathcal{R}_t$  estimation using all models for each problem, we use the same serial 393  
interval distribution, that was used to generate synthetic data. Taking different 394  
hyperparameters into consideration, we solve each problem using 8 methods including 395  
EpiEstim with weekly or monthly sliding windows, EpiLPS, EpiFilter, and RtEstim 396  
with piecewise constant, linear, quadratic, or cubic curves. Throughout the four  $\mathcal{R}_t$  397  
scenarios, the degrees of RtEstim can be correctly or wrongly specified. Our method 398  
can take the advantage of a correctly specified degree of piecewise polynomials 399  
compared to other methods, while the competitors only consider one fixed degree of 400  
smoothness which may not coincide with the “true” (assumed) degree of  $\mathcal{R}_t$ . Meanwhile, 401

by using different degrees to solve the same problem, we will illustrate that a wrongly 402  
specified degree can still result in accurate  $\mathcal{R}_t$  estimation in our experiments. 403

**RtEstim** estimates can appear to be piecewise polynomials with a selected degree  $k$ . 404  
When faced with real data, the choice of  $k$  should be done either (1) based on the 405  
analyst's preference for the result (e.g., "I want to find large jumps, so  $k = 0$ ") or (2) in 406  
a data-driven manner, as a component of the estimation process. Our software enables 407  
both cases, and the second case can be implemented by simply fitting for different  $k$  and 408  
choosing the set  $k, \lambda$  that has smallest CV score. Thus, all necessary choices can be 409  
accomplished based solely on the data. Our software is a departure from existing 410  
methods in that we *allow* this choice and provide simple data-driven methods to 411  
accomplish it. **EpiEstim** has no such facility (although, implicitly, one must somehow 412  
choose the size of the rolling window). **EpiFilter** effectively uses cubic splines (similar 413  
to  $k = 3$ , but simply continuous rather than piecewise continuous). Similarly, **EpiLPS** 414  
specifically chooses the cubic B-spline basis, which is similar to degree  $k = 3$ . **EpiNow2** 415  
allows various choices through Gaussian Process kernels, and while one can put priors 416  
on the parameters of the kernel, the choice of kernel is required. In our experience, 417  
using  $k > 3$  is nearly indistinguishable from  $k = 3$ , though it is allowed. So if the 418  
analyst somehow imagines " $\mathcal{R}_t$  is best described by a 10<sup>th</sup> order piecewise polynomial" 419  
then the software can easily accommodate this desire. 420

### 3.1.3 Accuracy measurement

 421

To measure estimation accuracy, we compare the estimated  $\widehat{\mathcal{R}}$  to "true"  $\mathcal{R}$  using the 422  
Kullback-Leibler (KL) divergence. The KL divergence for the Poisson distribution 423  
(summed over across all  $t$ ) which measures the accuracy of the  $\mathcal{R}_t$  estimates is defined as 424

$$D_{KL}(\mathcal{R} \parallel \widehat{\mathcal{R}}) = \sum_{t=1}^N \eta_t \left( \mathcal{R}_t \log \left( \frac{\mathcal{R}_t}{\widehat{\mathcal{R}}_t} \right) + \widehat{\mathcal{R}}_t - \mathcal{R}_t \right), \quad (12)$$

where  $\mathcal{R} = \{\mathcal{R}_t\}_{t=1}^N$  and  $\eta_t$  is the total infectiousness. We use the scaled (mean) KL 425  
divergence:  $\overline{D_{KL}}(\mathcal{R} \parallel \widehat{\mathcal{R}}) := D_{KL}(\mathcal{R} \parallel \widehat{\mathcal{R}})/N$ , where  $N$  is the length of the estimated  $\widehat{\mathcal{R}}$  426  
sequence. To fairly compare across methods, we drop the estimates during the first 427  
week because estimates from **EpiEstim** do not begin until  $t = 8$  (using a weekly 428  
window). The details on the derivation of the KL divergence in Eq (12) is deferred to 429

Section A.1 in Appendix.

430

KL divergence is more appropriate for measuring accuracy because it connects directly to the Poisson likelihood used to generate the data, whereas standard measures like the mean-squared error correspond to Gaussian likelihood. Using Poisson likelihood has the effect of increasing the relative cost of mistakes when  $\Lambda_t$  is small. Other details of the experimental settings are deferred to the supplementary document.

431

432

433

434

435

### 3.2 Results for synthetic data

436

`RtEstim` overall outperforms the other competitors in the experimental study. Fig 3 and Fig 4 visualize the KL divergence across the seven methods. For low incidence in measles epidemics, `RtEstim` is the most accurate for all  $\mathcal{R}_t$  scenarios given both Poisson and negative Binomial incidence. The best performance of `RtEstim` has the lowest median and has low or no overlap with other methods. For Scenario 1, `EpiFilter` is a competitive alternative given Poisson incidence, which has similar median to the best performance of our `RtEstim` and with a small variation. While given negative Binomial incidence, `EpiFilter` loses its advantage and even has the largest medians in Scenarios 1 and 2. The large incidence in SARS epidemics imposes more difficulty of  $\mathcal{R}_t$  estimation for all methods. The best performance of our method is quite robust in the scale of incidence given Poisson data, since the KL values are of the similar scale for two types of epidemics. Given negative Binomial incidence, `EpiLPS` shows robustness in the scale of incident cases in Scenarios 2 and 4. Our `RtEstim` has similar KL divergence values as `EpiLPS`, where the counterpart boxes overlap to a large degree. We will examine a single realization of each experiment to investigate these global conclusions in more detail.

451

452

453

454

455

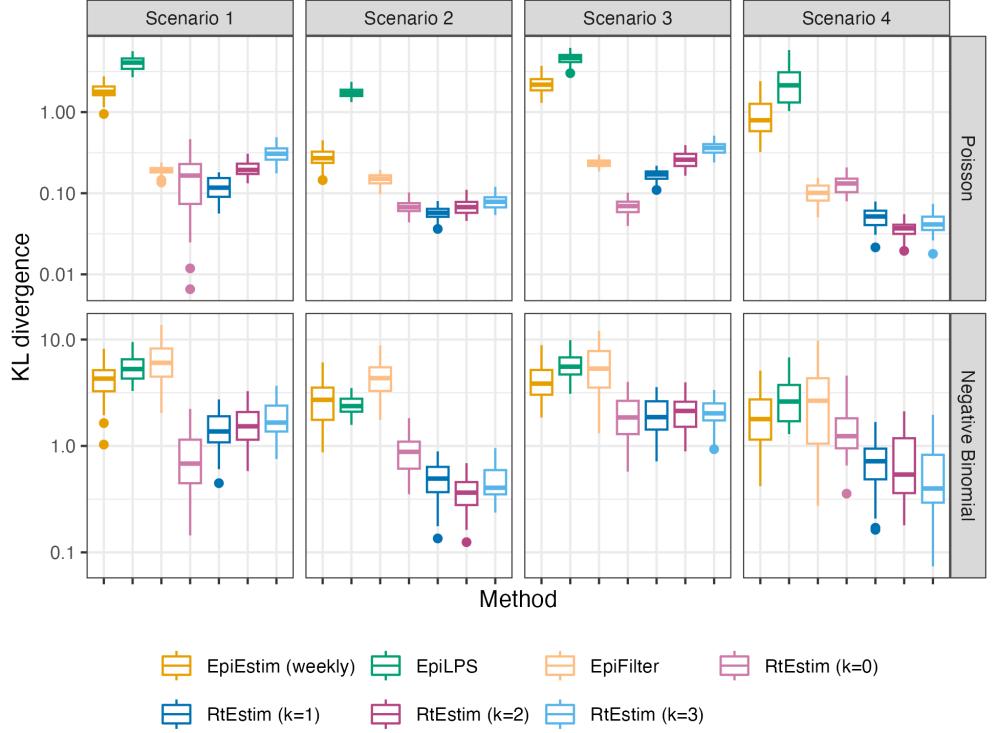
456

457

458

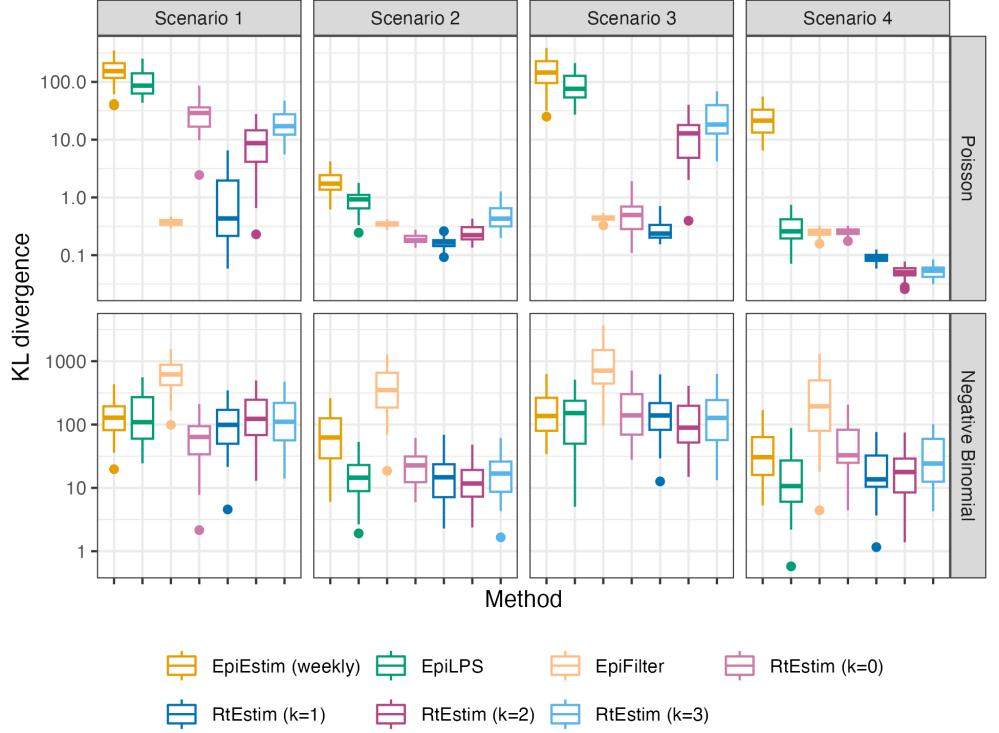
459

Fig 5 shows one realization for the estimated effective reproduction number under the Poisson generative model in measles synthetic epidemics for all four scenarios. An expanded visualization with each estimated  $\mathcal{R}_t$  curve displayed in a separate panel is provided in Figure A.6.1 in Appendix. Compared to `EpiEstim` and `EpiLPS`, which have rather severe difficulties at the beginning of the time series, `RtEstim` and `EpiFilter` estimates are more accurate without suffering from the initialization problem. Scenario 1 is the simplest case with only one knot and two constant segments. Besides the edge problem, `EpiEstim` and `EpiLPS` produce “smooth” estimated curves that are continuous



**Fig 3.** Boxplot of mean KL divergence between the estimated  $\hat{\mathcal{R}}_t$  and the true  $\mathcal{R}_t$  across 50 synthetic measles epidemics for each approach given Poisson incidence (*in top panels*) and negative Binomial incidence (*in bottom panels*) respectively. The mean KL divergence ignores the first weeks in all experiments, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Outliers beyond  $1.5 \times \text{IQR}$  of each box are excluded, and full illustration in provided in the Figure A.3.1 in Appendix.

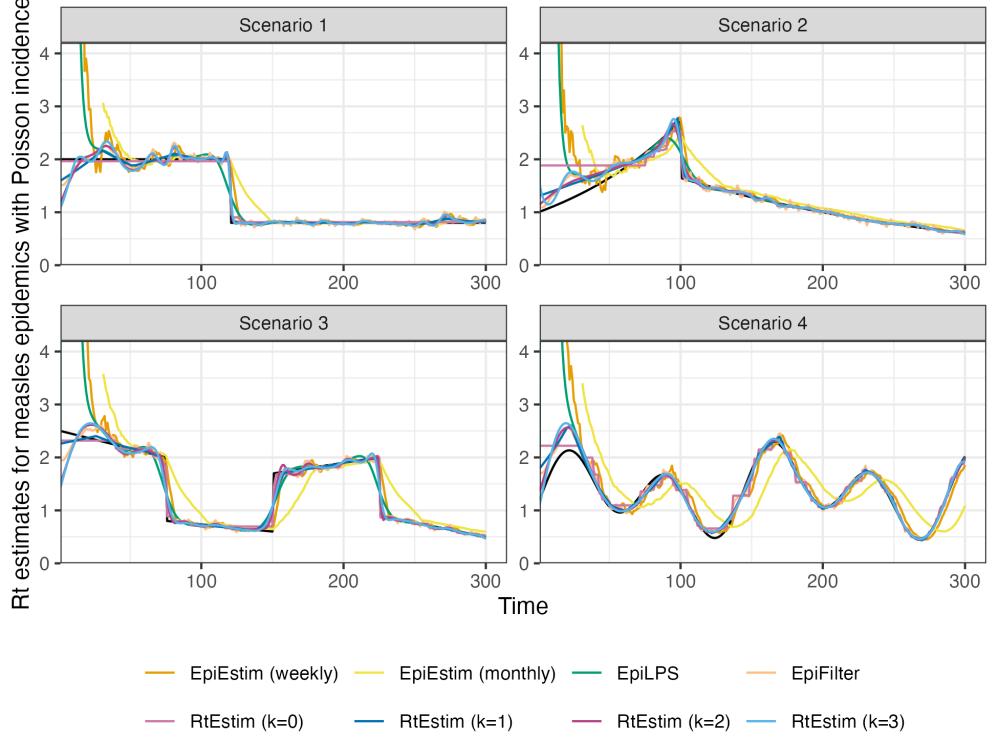
at the changepoint, which results in large mistakes in that neighbourhood. Since the piecewise constant **RtEstim** estimator does not force any smoothness in  $\mathcal{R}_t$ , it easily captures the sharp change. It nearly overlaps with the true values. Scenario 2 is relatively easy for all methods, except at the changepoint occurring at the end of the exponential growth. It is a common problem for all methods to estimate the first stage of exponential growth. **EpiEstim** and **EpiLPS** tend to start with unreasonably large estimates and decreases rapidly. **RtEstim** also suffers from the difficulty to estimate the first few timepoints, especially in the piecewise constant case. All methods are more or less inaccurate at the beginning of the epidemic in both Scenarios 3 and 4, and recover the underlying curves quite well, except **EpiEstim** with monthly sliding windows, which tends to have a delay in the changes. Fig 6 shows a realization of the estimated  $\mathcal{R}_t$  given negative Binomial incidence in SARS epidemics for each setting. All methods



**Fig 4.** Boxplot of mean KL divergence between the estimated  $\hat{R}_t$  and the true  $R_t$  across 50 synthetic SARS epidemics for each approach given Poisson incidence (*in top panels*) and negative Binomial incidence (*in bottom panels*) respectively. The mean KL divergence ignores the first weeks in all experiments, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Outliers beyond  $1.5 \times \text{IQR}$  of each box are excluded, and full illustration in provided in the Figure A.3.1 in Appendix.

more or less performs worse in this setting, due to larger incidence and overdispersed data. The challenges to recover the start of epidemics are harder in this setting for all methods. **EpiFilter** has much more wiggly estimates than the estimates of other methods. Our **RtEstim** estimates are close to the best performance in the first three  $R_t$  scenarios, while face the challenge to recover the curve in the periodic scenario.

Finally, it is important to provide a brief comparison of the running times of all three models across the 8 experimental settings. We find that almost all models across all experiments complete within 10 seconds. **RtEstim** generally takes the longest, due to a relatively large number of estimates—50 values of  $\lambda$  and 10 folds of cross validation require 550 estimates—while other models run only a single time for a fixed setting of hyperparameters per experiment. Additional results on timing comparisons are deferred to the supplementary document.

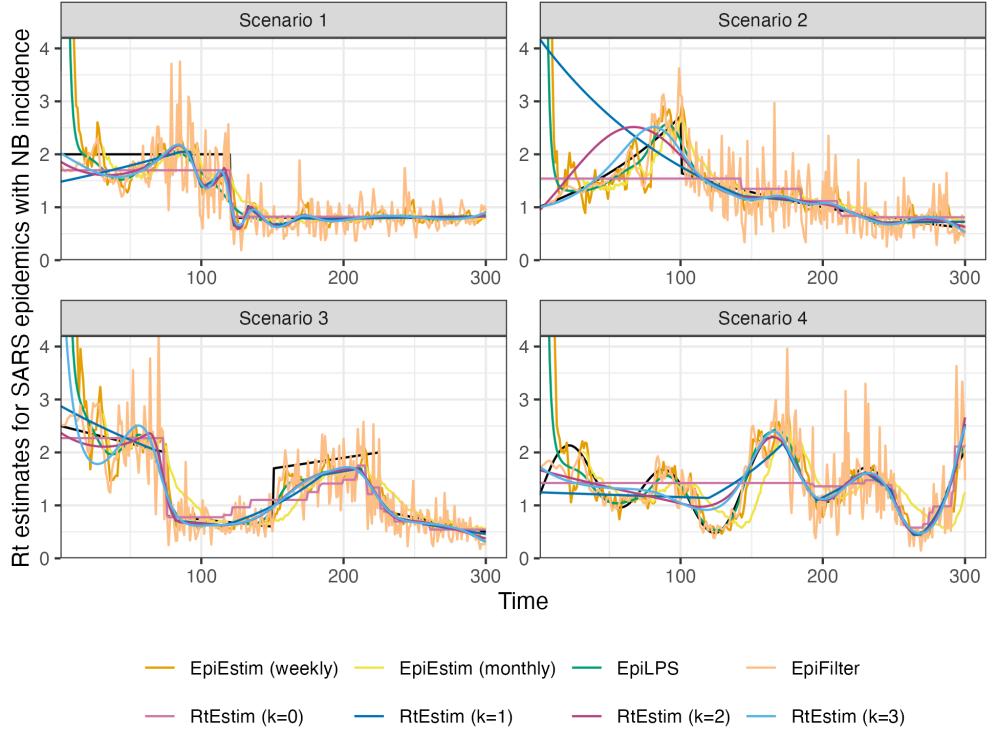


**Fig 5.** Example of effective reproduction number estimation for measles epidemics with Poisson observations. An expanded visualization with each estimated  $\mathcal{R}_t$  curve displayed in a separate panel is provided in Figure A.6.1 in Appendix.

### 3.3 Real-data results: Covid-19 incident cases in Canada

We implement `RtEstim` on Covid-19 confirmed incident cases in Canada (visualized in Fig 1). We use the weighted probabilities of serial interval distributions of four dominated variants used in Fig 1 as the serial interval distribution here for the comparison with other methods, which cannot incorporate time-varying serial interval distributions. We compute their percentages of dominated days throughout the pandemic as the weights in computation, specifically Ancestral lineage (32.6%), Alpha (8.5%), Delta (16.0%), and Omicron(42.9%). The estimates of our method is displayed in Fig 7, and the estimates of all competitors are deferred to Figures A.8.1 and A.8.2 in Appendix.

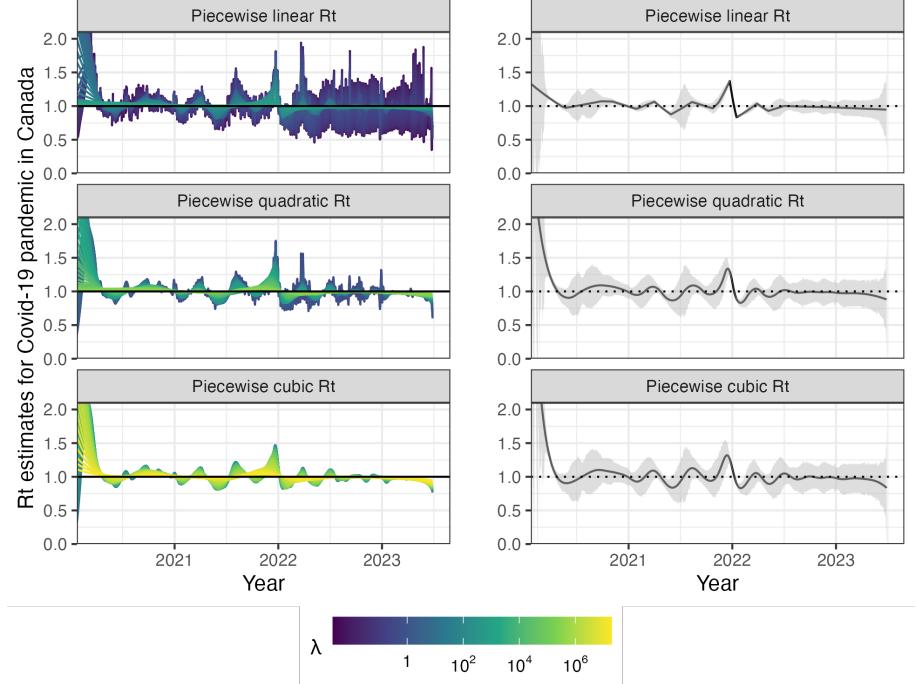
Considering the first, second, and third polynomial degrees,  $\hat{\mathcal{R}}_t$  for Covid-19 in Canada is always less than 2 except at the very early stage, which means that one distinct infected individuals on average infects less than three other individuals in the population. Examining three different settings for  $k$ , the temporal evolution of  $\hat{\mathcal{R}}$



**Fig 6.** Example of effective reproduction number estimation for SARS epidemics with negative Binomial observations.

(across all regularization levels  $\lambda$ ) are similar near the highest peak around the end of 498  
2021 before dropping shortly thereafter. Throughout the estimated curves, the peaks 499  
and troughs of the effective reproduction numbers precede the growth and decay cycles 500  
of confirmed cases, as expected. We also visualize 95% confidence bands for the point 501  
estimates with  $\lambda$  chosen by minimizing cross-validated KL divergence in Fig 7. 502

The estimated effective reproduction numbers are relatively unstable before April, 503  
2022. The highest peak coincides with the emergence and global spread of the Omicron 504  
variant. The estimated effective reproduction numbers fall below 1 during two time 505  
periods—roughly from April, 2021 to July, 2021 and from January, 2022 to April, 2022. 506  
The first trough coincides with the introduction of Covid-19 vaccines in Canada. The 507  
second trough, shortly after the largest peak may be due to variety of factors resulting 508  
in the depletion of the susceptible population such as increased self-isolation in response 509  
to media coverage of the peak or immunity incurred via recent infection. Since April, 510  
2022, the estimated effective reproduction number has remained relatively stable 511  
(fluctuating around one) corresponding to low reported cases, though reporting 512



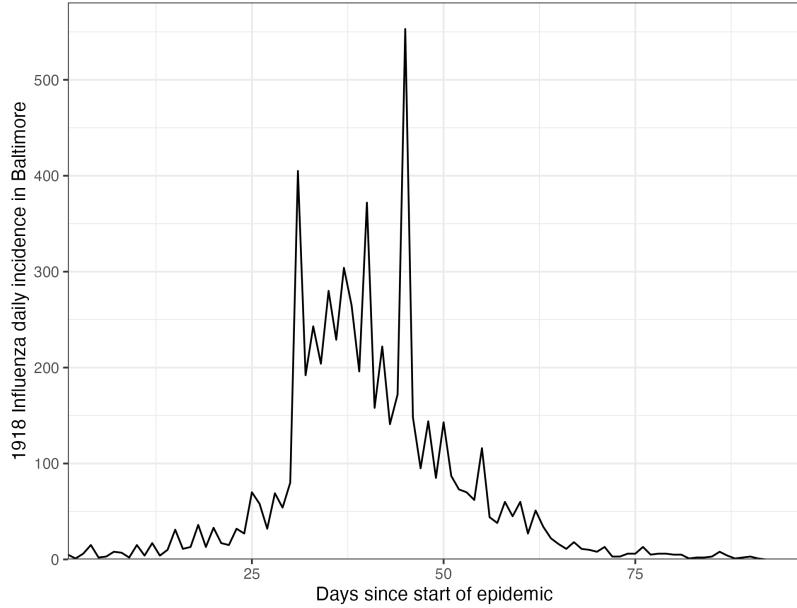
**Fig 7.** Estimated effective reproduction number based on Covid-19 daily confirmed incident cases between January 23rd, 2020 and June 28th, 2023 in Canada. The left panels show estimates corresponding to 50 tuning parameters. The right panels show the CV-tuned estimate along with approximate 95% confidence bands. The top, middle and bottom panels show the estimated  $\mathcal{R}_t$  using the Poisson trend filtering in Eq (5) with degrees  $k = 1, 2, 3$  respectively. All estimates are fitted using a constant serial interval distribution, which is the weighted sum of probabilities of the 4 dominant variants per timepoint used in Fig 1. All panels are truncated in y-axes for better illustration. The CV-tuned  $\mathcal{R}_t$  estimates rapidly decreases at the early stage from 3.37, 5.16 for  $k = 2, 3$  respectively.

behaviours also changed significantly since the Omicron wave.

### 3.4 Real-data results: influenza in Baltimore, Maryland, 1918

We also apply `RtEstim` to daily reported influenza cases in Baltimore, Maryland occurring during the world-wide pandemic of 1918 from September to November [41]. The data, shown in Fig 8, is included in the `EpiEstim` R package. The 1918 influenza outbreak, caused by the H1N1 influenza A virus, was unprecedentedly deadly with case fatality rate over 2.5%, infecting almost one-third of the population across the world [42]. The CV-tuned piecewise cubic estimates in Fig 9 better capture the growth at the beginning of the pandemic in Fig 8. The estimated  $\mathcal{R}_t$  curve suggests that the transmissibility of the pandemic grew rapidly over the first 30 days before declining

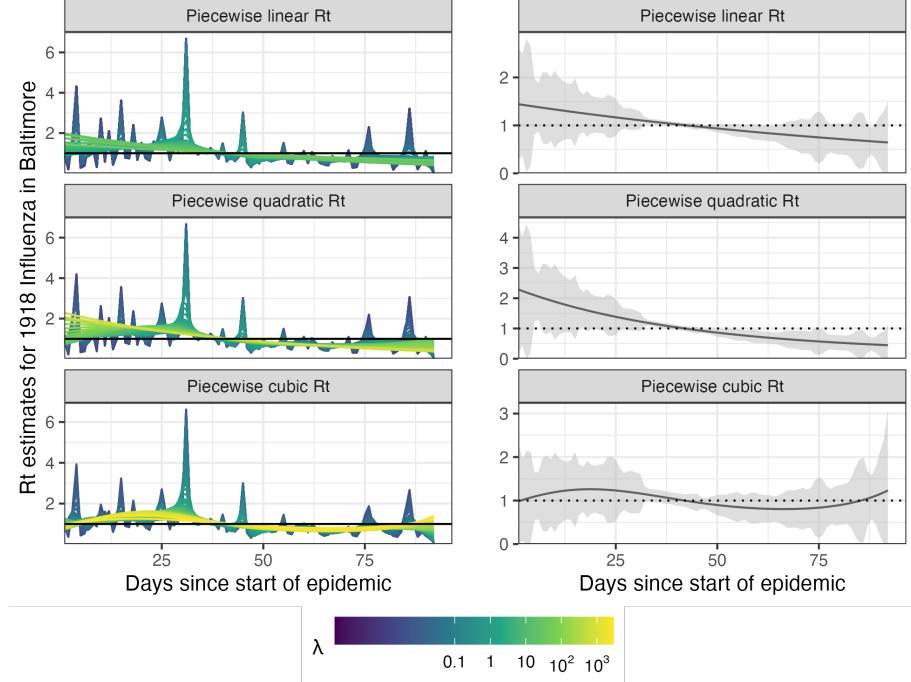
below one after 50 days. However, it also suggests an increase in infectiousness toward the end of the period. With this data, it is difficult to determine if there is a second wave or a steady decline ahead. The CV-tuned piecewise constant and linear estimates in Fig 9 both suggest a steady decline. This conclusion is supported by the fact that incident cases decline to zero at the end of the period and matches  $\mathcal{R}$  estimates in [10], which are all lower than one.



**Fig 8.** Daily incident influenza cases in Baltimore, Maryland between September and November in 1918.

## 4 Discussion

The **RtEstim** methodology provides a locally adaptive estimator using Poisson trend filtering on univariate data. It captures the heterogeneous smoothness of effective reproduction numbers given observed incidence data rather than resulting in global smoothness. This is a nonparametric regression model which can be written as a convex optimization (minimization) problem. Minimizing the distance (KL divergence across all coordinates) between the estimators and (functions of) observations guarantees data fidelity while the penalty on divided differences between pairs of neighbouring parameters imposes smoothness. The  $\ell_1$ -regularization results in sparsity of the divided differences, which leads to heterogeneous smoothness across time.



**Fig 9.** Estimated effective reproduction numbers for influenza in Baltimore, Maryland in 1918. The left panels show estimates for a set of 50 tuning parameters. The right column displays the CV-tuned estimates with approximate 95% confidence bands. The rows (top to bottom) show estimated effective reproduction numbers ( $\mathcal{R}_t$ ) using the Poisson trend filtering in Eq (5) with  $k = 1, 2, 3$  respectively.

The property of local adaptivity (heterogenous smoothness) is useful to automatically distinguish, for example, seasonal outbreaks from outbreaks driven by other factors (behavioural changes, foreign introduction, etc.). Given a well-chosen polynomial degree, the growth rates can be quickly detected, potentially advising public health authorities to implement policy changes. The effective reproduction numbers can be estimated retrospectively to examine the efficacy of such policies, whether they result in  $\mathcal{R}_t$  falling below 1 or the speed of their effects. The smoothness of  $\mathcal{R}_t$  curves (including the polynomial degrees and tuning parameters) should be chosen based on the purpose of the study in practice.

Our method `RtEstim` provides a natural way to deal with missing data, for example, on weekends and holidays or due to changes in reporting frequency. While solving the convex optimization problem, our method can easily handle uneven spacing or irregular reporting. Computing the total primary infectiousness is also easily generalized to irregular reporting by modifying the discretization of the serial interval distribution.

There are many other aspects to be considered in choosing the delay distribution to  
make a more accurate estimation [43]. Imported cases can be distinguished from the  
local cases to avoid the biasness in effective reproduction number estimation, for  
example, [11] assumed the total past infectiousness of combined previous cases scaled by  
effective reproduction number to be the mean of the local incidence and illustrated that  
failure of distinguishing the local cases from the imported cases may cause the  
overestimation of  $\mathcal{R}$  using the MERS data in Saudi Arabia from August 2014 to  
December 2015. Additionally, because the  $\ell_1$  penalty introduces sparsity (operating like  
a median rather than a mean), this procedure is relatively insensitive to outliers  
compared to  $\ell_2$  regularization.

There are a number of limitations that may influence the quality of  $\mathcal{R}_t$  estimation.  
While our model is generic for incidence data rather than tailored to any specific  
disease, it does assume that the generation interval is short relative to the period of  
data collection. More specialized methodologies would be required for diseases with long  
incubation periods such as HIV or Hepatitis. Our approach, does not explicitly model  
imported cases, nor distinguish between subpopulations that may have different mixing  
behaviour. While the Poisson assumption is common, it does not handle overdispersion  
(observation variance larger than the mean). The negative binomial distribution is a  
good alternative, but more difficult to estimate in this context. As described in  
[section 1](#), the expression for  $\mathcal{R}$  assumes that a relatively constant proportion of true  
infections is reported. However, if this proportion varies with time (say, due to changes  
in surveillance practices or testing recommendations), the estimates may be biased over  
this window. A good example is in early January 2022, during the height of the  
Omicron wave, Canada moved from testing all symptomatic individuals to testing only  
those in at-risk groups. The result was a sudden change that would render  $\mathcal{R}_t$  estimates  
on either side of this timepoint incommensurable.

Our `RtEstim` implementation can take a fixed serial interval throughout the period of  
study (as implemented in simulation and in the real epidemics) or varying serial interval  
distributions at different timepoints (as implemented in Fig [Fig 1](#) for Covid-19 data in  
Canada). In reality, the serial interval may vary due to changes in the factors such as  
population immunity [12]. An issue regarding the serial interval distribution relates to  
equating serial and generation intervals (also mentioned above). The serial interval

distribution is generally wider than that of the generation interval, because the serial 585  
interval involves the convolution of two distributions, and is unlikely to actually follow a 586  
named distribution like Gamma, though it may be reasonably well approximated by one. 587  
Our implementation allows for an arbitrary distribution to be used, but requires the 588  
user to specify the discretization explicitly, requiring more nuanced knowledge than is 589  
typically available. Pushing this analysis further, to accommodate other types of 590  
incidence data (hospitalizations or deaths), a modified generation interval distribution 591  
would be necessary, and further assumptions would be required as well. Or else, one 592  
would first need to deconvolve deaths to infection onset before using our software. 593

Nonetheless, our methodology is implemented in a lightweight R package **rtestim** 594  
and computed efficiently, especially for large-scale data, with a proximal Newton solver 595  
coded in C++. Given available incident case data, prespecified serial interval distribution, 596  
and a choice of degree  $k$ , **RtEstim** is able to produce accurate estimates of effective 597  
reproduction number and provide efficient tuning parameter selection via cross 598  
validation. 599

## Acknowledgments

This research was enabled in part by support provided by BC DRI group who manages 600  
Cedar cloud (<https://docs.alliancecan.ca/wiki/Cedar>) and the Digital Research Alliance 601  
of Canada ([alliancecan.ca](http://alliancecan.ca)). 602  
603

## References

1. Gostic KM, McGough L, Baskerville EB, Abbott S, Joshi K, Tedijanto C, et al. Practical considerations for measuring the effective reproductive number,  $R_t$ . PLoS Computational Biology. 2020;16(12):e1008409.
2. Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. Mathematical and statistical estimation approaches in epidemiology. 2009; p. 103–121.
3. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PloS one. 2007;2(8):e758.

4. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. Oxford university press; 1991.
5. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*. 2004;160(6):509–516.
6. Hao X, Cheng S, Wu D, Wu T, Lin X, Wang C. Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature*. 2020;584(7821):420–424.
7. Goldstein IH, Parker DM, Jiang S, Minin VM. Semiparametric inference of effective reproduction number dynamics from wastewater pathogen surveillance data. *arXiv preprint arXiv:230815770*. 2023;.
8. Goldstein IH, Wakefield J, Minin VN. Incorporating testing volume into estimation of effective reproduction number dynamics. *arXiv preprint arXiv:220804418*. 2022;.
9. Cori A, Cauchemez S, Ferguson NM, Fraser C, Dahlqwist E, Demarsh PA, et al.. EpiEstim: estimate time varying reproduction numbers from epidemic curves; 2020.
10. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*. 2013;178(9):1505–1512.
11. Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*. 2019;29:100356.
12. Nash RK, Bhatt S, Cori A, Nouvellet P. Estimating the epidemic reproduction number from temporally aggregated incidence data: A statistical modelling approach and software tool. *PLoS Computational Biology*. 2023;19(8):e1011439.
13. Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research*. 2020;5:112.

14. Abbott S, Funk S, Hickson J, Badr HS, Monticone P, Ellis P, et al.. *epiforecasts/EpiNow2*: 1.4.0 release; 2023.
15. Lison A, Abbott S, Huisman J, Stadler T. Generative Bayesian modeling to nowcast the effective reproduction number from line list data with missing symptom onset dates. arXiv preprint arXiv:230813262. 2023;.
16. Parag KV. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Computational Biology*. 2021;17(9):e1009347.
17. Gressani O, Wallinga J, Althaus CL, Hens N, Faes C. EpiLPS: A fast and flexible Bayesian tool for estimation of the time-varying reproduction number. *PLoS Computational Biology*. 2022;18(10):e1010618.
18. Trevisin C, Bertuzzo E, Pasetto D, Mari L, Miccoli S, Casagrandi R, et al. Spatially explicit effective reproduction numbers from incidence and mobility data. *Proceedings of the National Academy of Sciences*. 2023;120(20):e2219816120.
19. Abry P, Pustelnik N, Roux S, Jensen P, Flandrin P, Gribonval R, et al. Spatial and temporal regularization to estimate COVID-19 reproduction number  $R(t)$ : Promoting piecewise smoothness via convex optimization. *PLoS ONE*. 2020;15(8):e0237901.
20. Pascal B, Abry P, Pustelnik N, Roux S, Gribonval R, Flandrin P. Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data. *IEEE Transactions on Signal Processing*. 2022;70:2859–2868.
21. Pircalabelu E. A spline-based time-varying reproduction number for modelling epidemiological outbreaks. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 2023;72(3):688–702.
22. Ho F, Parag KV, Adam DC, Lau EH, Cowling BJ, Tsang TK. Accounting for the Potential of Overdispersion in Estimation of the Time-varying Reproduction Number. *Epidemiology*. 2023;34(2):201–205.

23. Azmon A, Faes C, Hens N. On the estimation of the reproduction number based on misreported epidemic data. *Statistics in Medicine*. 2014;33(7):1176–1192.
24. Gressani O, Faes C, Hens N. An approximate Bayesian approach for estimation of the instantaneous reproduction number under misreported epidemic data. *Biometrical Journal*. 2022;65(6):2200024.
25. Jin S, Dickens BL, Lim JT, Cook AR. EpiMix: A novel method to estimate effective reproduction number. *Infectious Disease Modelling*. 2023;8(3):704–716.
26. Hettinger G, Rubin D, Huang J. Estimating the instantaneous reproduction number with imperfect data: a method to account for case-reporting variation and serial interval uncertainty. arXiv preprint arXiv:230212078. 2023;.
27. Berry I, O'Neill M, Sturrock SL, Wright JE, Acharya K, Brankston G, et al. A sub-national real-time epidemiological and vaccination database for the COVID-19 pandemic in Canada. *Scientific Data*. 2021;8(1). doi:10.1038/s41597-021-00955-2.
28. Coronavirus Variants Rapid Response Network Computational Analysis M, Group EOC. CoVaRR-NET/duotang: Release for Zenodo Archive; 2023. Available from: <https://doi.org/10.5281/zenodo.10367461>.
29. Xu X, Wu Y, Kummer AG, Zhao Y, Hu Z, Wang Y, et al. Assessing changes in incubation period, serial interval, and generation time of SARS-CoV-2 variants of concern: a systematic review and meta-analysis. *BMC medicine*. 2023;21(1):374.
30. Park SW, Sun K, Champredon D, Li M, Bolker BM, Earn DJ, et al. Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. *Proceedings of the National Academy of Sciences*. 2021;118(2):e2011548118.
31. Pitzer VE, Chitwood M, Havumaki J, Menzies NA, Perniciaro S, Warren JL, et al. The impact of changes in diagnostic testing practices on estimates of COVID-19 transmission in the United States. *American Journal of Epidemiology*. 2021;190(9):1908–1917.

32. Hitchings MD, Dean NE, García-Carreras B, Hladish TJ, Huang AT, Yang B, et al. The usefulness of the test-positive proportion of severe acute respiratory syndrome coronavirus 2 as a surveillance tool. *American Journal of Epidemiology*. 2021;190(7):1396–1405.
33. Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LH, Fearon E, et al. Challenges in control of COVID-19: short doubling time and long delay to effect of interventions. *Philosophical Transactions of the Royal Society B*. 2021;376(1829):20200264.
34. Kim SJ, Koh K, Boyd S, Gorinevsky D.  $\ell_1$  trend filtering. *SIAM Review*. 2009;51(2):339–360.
35. Tibshirani RJ. Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*. 2014;42(1):285–323.
36. Tibshirani RJ. Divided differences, falling factorials, and discrete splines: Another look at trend filtering and related problems. *Foundations and Trends® in Machine Learning*. 2022;15(6):694–846.
37. Sadhanala V, Bassett R, Sharpnack J, McDonald DJ. Exponential Family Trend Filtering on Lattices. *arXiv preprint arXiv:220909175*. 2022;
38. Vaiter S, Deledalle C, Fadili J, Peyré G, Dossal C. The degrees of freedom of partly smooth regularizers. *Annals of the Institute of Statistical Mathematics*. 2017;69:791–832.
39. Groendyke C, Welch D, Hunter DR. Bayesian inference for contact networks given epidemic data. *Scandinavian Journal of Statistics*. 2011;38(3):600–616.
40. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, et al. Transmission dynamics and control of severe acute respiratory syndrome. *science*. 2003;300(5627):1966–1970.
41. Frost WH, Sydenstricker E. Influenza in Maryland: preliminary statistics of certain localities. *Public Health Reports (1896-1970)*. 1919; p. 491–504.

42. Taubenberger JK, Morens DM. 1918 Influenza: the mother of all pandemics. *Emerging Infectious Diseases*. 2006;17(1):69–79.
43. Park SW, Akhmetzhanov AR, Charniga K, Cori A, Davies NG, Dushoff J, et al. Estimating epidemiological delay distributions for infectious diseases. *medRxiv*. 2024; p. 2024–01.

# Supplementary details on experiments of effective reproduction number estimation with trend filtering

Jiapeng Liu, Zhenglun Cai, Paul Gustafson, and Daniel J. McDonald

## A.1 Derivation of Kullback Leibler divergence for accuracy comparison

We provide the detailed derivation of the Kullback Leibler (KL) divergence in (11) that is used to compare the accuracy of the estimated effective reproduction number with the true ones. Given the total infectiousness  $\eta$ , we compare the distance between the Poisson distributions  $f_1(y; \eta, \hat{\mathcal{R}}) = Pois(\eta\hat{\mathcal{R}})$  and  $f_0(y; \eta, \mathcal{R}) = Pois(\eta\mathcal{R})$ , where  $y, \mathcal{R} \in \mathbb{N}_0^n$  are natural numbers including 0,  $\eta \in \mathbb{R}^n$ ,  $f_0(y) = \prod_{t=1}^n \frac{(\eta_t \mathcal{R}_t)^{y_t} e^{-\eta_t \mathcal{R}_t}}{y_t!}$ ,  $f_1(y) = \prod_{t=1}^n \frac{(\eta_t \hat{\mathcal{R}}_t)^{y_t} e^{-\eta_t \hat{\mathcal{R}}_t}}{y_t!}$ ,  $y_t \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$ . Then, the KL divergence between them is defined as

$$\begin{aligned}
D_{KL}(\mathcal{R} || \hat{\mathcal{R}}) &= D_{KL}(f_0(y) || f_1(y)) \\
&= \sum_{y \in \mathbb{N}_0^n} f_0(y) \log \frac{f_0(y)}{f_1(y)} \\
&= \sum_{y \in \mathbb{N}_0^n} \prod_{t=1}^n \frac{(\eta_t \mathcal{R}_t)^{y_t} e^{-\eta_t \mathcal{R}_t}}{y_t!} \log \prod_{t=1}^n \frac{\mathcal{R}_t^{y_t} e^{-\eta_t \mathcal{R}_t}}{\hat{\mathcal{R}}_t^{y_t} e^{-\eta_t \hat{\mathcal{R}}_t}} \\
&= \sum_{y_n=0}^{\infty} \dots \sum_{y_1=0}^{\infty} \prod_{t=1}^n \frac{(\eta_t \mathcal{R}_t)^{y_t} e^{-\eta_t \mathcal{R}_t}}{y_t!} \sum_{t=1}^n \left( y_t \log \frac{\mathcal{R}_t}{\hat{\mathcal{R}}_t} - \eta_t (\mathcal{R}_t - \hat{\mathcal{R}}_t) \right) \text{ for independent } y_t, t = 1, \dots, n \\
&= \sum_{y_n=0}^{\infty} \frac{(\eta_n \mathcal{R}_n)^{y_n} e^{-\eta_n \mathcal{R}_n}}{y_n!} \dots \sum_{y_1=0}^{\infty} \frac{(\eta_1 \mathcal{R}_1)^{y_1} e^{-\eta_1 \mathcal{R}_1}}{y_1!} \left( y_1 \log \frac{\mathcal{R}_1}{\hat{\mathcal{R}}_1} + \sum_{t=2}^n y_t \log \frac{\mathcal{R}_t}{\hat{\mathcal{R}}_t} - \sum_{t=1}^n \eta_t (\mathcal{R}_t - \hat{\mathcal{R}}_t) \right) \\
&= \sum_{y_n=0}^{\infty} \frac{(\eta_n \mathcal{R}_n)^{y_n} e^{-\eta_n \mathcal{R}_n}}{y_n!} \dots \sum_{y_2=0}^{\infty} \frac{(\eta_2 \mathcal{R}_2)^{y_2} e^{-\eta_2 \mathcal{R}_2}}{y_2!} \left( \eta_1 \mathcal{R}_1 \log \frac{\mathcal{R}_1}{\hat{\mathcal{R}}_1} + \sum_{t=2}^n y_t \log \frac{\mathcal{R}_t}{\hat{\mathcal{R}}_t} - \sum_{t=1}^n \eta_t (\mathcal{R}_t - \hat{\mathcal{R}}_t) \right) \\
&= \sum_{y_n=0}^{\infty} \frac{(\eta_n \mathcal{R}_n)^{y_n} e^{-\eta_n \mathcal{R}_n}}{y_n!} \left( \sum_{t=1}^{n-1} \eta_t \mathcal{R}_t \log \frac{\mathcal{R}_t}{\hat{\mathcal{R}}_t} + y_n \log \frac{\mathcal{R}_n}{\hat{\mathcal{R}}_n} - \sum_{t=1}^n \eta_t (\mathcal{R}_t - \hat{\mathcal{R}}_t) \right) \\
&= \sum_{t=1}^n \eta_t \left( \mathcal{R}_t \log \frac{\mathcal{R}_t}{\hat{\mathcal{R}}_t} + \hat{\mathcal{R}}_t - \mathcal{R}_t \right),
\end{aligned}$$

where

$$\begin{aligned}
& \sum_{y_1=0}^{\infty} \frac{(\eta_1 \mathcal{R}_1)^{y_1} e^{-\eta_1 \mathcal{R}_1}}{y_1!} \left( y_1 \log \frac{\mathcal{R}_1}{\hat{\mathcal{R}}_1} + \sum_{t=2}^n y_t \log \frac{\mathcal{R}_t}{\hat{\mathcal{R}}_t} - \sum_{t=1}^n \eta_t (\mathcal{R}_t - \hat{\mathcal{R}}_t) \right) \\
&= \left( \sum_{y_1=0}^{\infty} \frac{(\eta_1 \mathcal{R}_1)^{y_1-1} e^{-\eta_1 \mathcal{R}_1}}{(y_1-1)!} \eta_1 \mathcal{R}_1 \log \frac{\mathcal{R}_1}{\hat{\mathcal{R}}_1} \right) + \sum_{t=2}^n y_t \log \frac{\mathcal{R}_t}{\hat{\mathcal{R}}_t} - \sum_{t=1}^n \eta_t (\mathcal{R}_t - \hat{\mathcal{R}}_t) \\
&= \eta_1 \mathcal{R}_1 \log \frac{\mathcal{R}_1}{\hat{\mathcal{R}}_1} + \sum_{t=2}^n y_t \log \frac{\mathcal{R}_t}{\hat{\mathcal{R}}_t} - \sum_{t=1}^n \eta_t (\mathcal{R}_t - \hat{\mathcal{R}}_t).
\end{aligned}$$

We use mean KL divergence (denoted,  $\overline{D_{KL}}(\mathcal{R}||\hat{\mathcal{R}}) := D_{KL}(\mathcal{R}||\hat{\mathcal{R}})/n$ , which is the KL divergence divided by the sequence length) in experiments for accuracy comparison.

## A.2 Supplementary details on experimental settings

We compare the accuracy of the estimated effective reproduction numbers using the mean Kullback Leibler (KL) divergence (with Poisson distributional assumption on incidence) in (10) across our **RtEstim** and several alternative methods including **EpiEstim** with weekly and monthly sliding windows, **EpiLPS**, **EpiFilter**, **EpiNow2**, and **RtEstim** with degrees  $k=0,1,2,3$ , which yields 9 methods in total. We consider two lengths of epidemics with  $n = 50$  or  $n = 300$  timepoints respectively. Since **EpiNow2** runs too long (specifically, for a long **measles** epidemic, it takes almost 2 hours (115 minutes computed on Cedar cluster provided by Compute Canada), we only compare it with other methods for short epidemics.

We consider serial interval (SI) distributions of **measles** and **SARS** to generate long synthetic epidemics, and **flu** for short epidemics, inspired by Cori et al. (2013). The means and standard deviations of SI distributions are estimated by existing literatures; specifically, (14.9, 3.9) for **measles** (Groendyke, Welch, and Hunter (2011)), (8.4, 3.8) for **SARS** (Lipsitch et al. (2003)), and (2.6, 1.5) for **flu** (Ferguson et al. (2005), Boëlle et al. (2011)). Incident cases in synthetic **measles** epidemics are relatively low (within 1000 at the peak overall), and **SARS** incident cases are relatively large (between 15000 and 20000 at the peak overall). We consider a reasonably large overdispersion level of negative Binomial incidence with size 5. Figure A.2.1 displays the ratio of standard deviation over mean (called, sigma to mean ratio) of incidence across different settings using the same set of sample epidemics in Fig 5, Fig 6, and all figures in section A.6.1. Compared to the counterpart of Poisson incidence (which decreases quickly to 0 and remains to be under 0.25) per  $\mathcal{R}_t$  scenario for each epidemic, the negative Binomial incidence appears to have an apparently larger sigma to mean ratio (staying at around 0.5 or above), which implies a distinguishable overdispersion level.

In model fitting, we use both true and misspecified serial interval (SI) distributions to test the robustness of our method, compared to other alternatives. The misspecification of serial interval distributions are either mild or major, where, in the major misspecification, we use a completely different pair of SI parameters, e.g., we use SI of **SARS** to solve measles epidemics, and SI of measles to solve short **flu** epidemics. While, in the mild SI misspecification, we consider slightly adjusted parameters for both **measles** and **flu** epidemics, where the mean is decreased by 2 for **measles** and increased by 2 for **flu** and the standard deviation is increased by 10, denoted as **adj\_flu** and **adj\_measles** respectively. These settings result in 7 pairs of SI distributions (for epidemic generating, model fitting), i.e., (**measles**, **measles**), (**SARS**, **SARS**), (**measles**, **adj\_measles**), (**measles**, **SARS**) for long epidemics and (**flu**, **flu**), (**flu**, **adj\_flu**), (**flu**, **measles**) for short

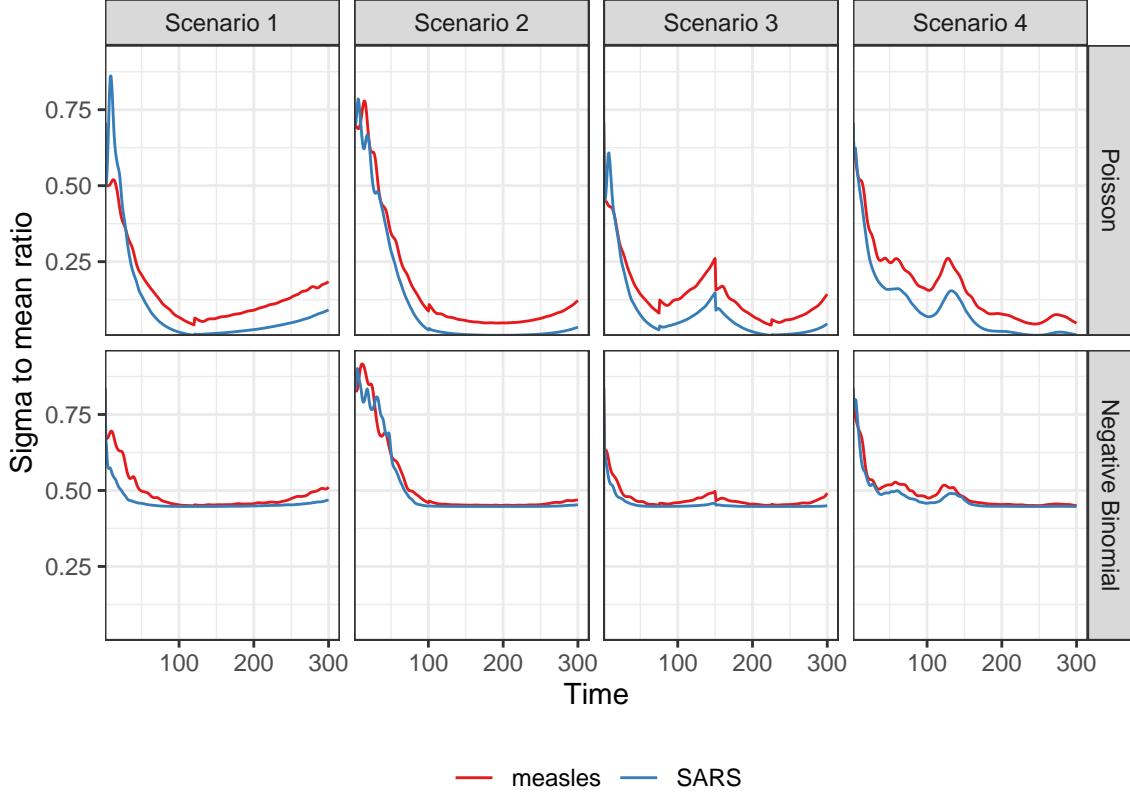


Figure A.2.1: Dispersion level of incidence of sample epidemics

Table 1: Summary of experimental setting on accuracy comparison

Length	SI	Rt scenario	Incidence	SI for modelling	Method
300	measles	1-4	Poisson, NB	measles, adj_measles, SARS	8 methods
300	SARS	1-4	Poisson, NB	SARS	8 methods
50	flu	3	Poisson, NB	flu, adj_flu, measles	9 methods

epidemics. Figure A.2.2 displays all SI distributions (`measles`, `adj_measles`, `SARS`, `flu`, and `adj_flu`) used in the experiments.

Table 1 summarizes the aforementioned experimental setting for accuracy comparison. Poisson and negative Binomial (NB) distributions for incidence and 4  $\mathcal{R}_t$  scenarios are used for all long epidemics. We only consider one  $\mathcal{R}_t$  scenario for short epidemics. Each experimental setting is replicated for 50 times, which yields 12800 experiments for long epidemics and 2700 for short epidemics.

We visualize the selected key results of the accuracy comparison using long synthetic epidemics in Section 3.2. Other main experimental results are displayed in Section A.3.

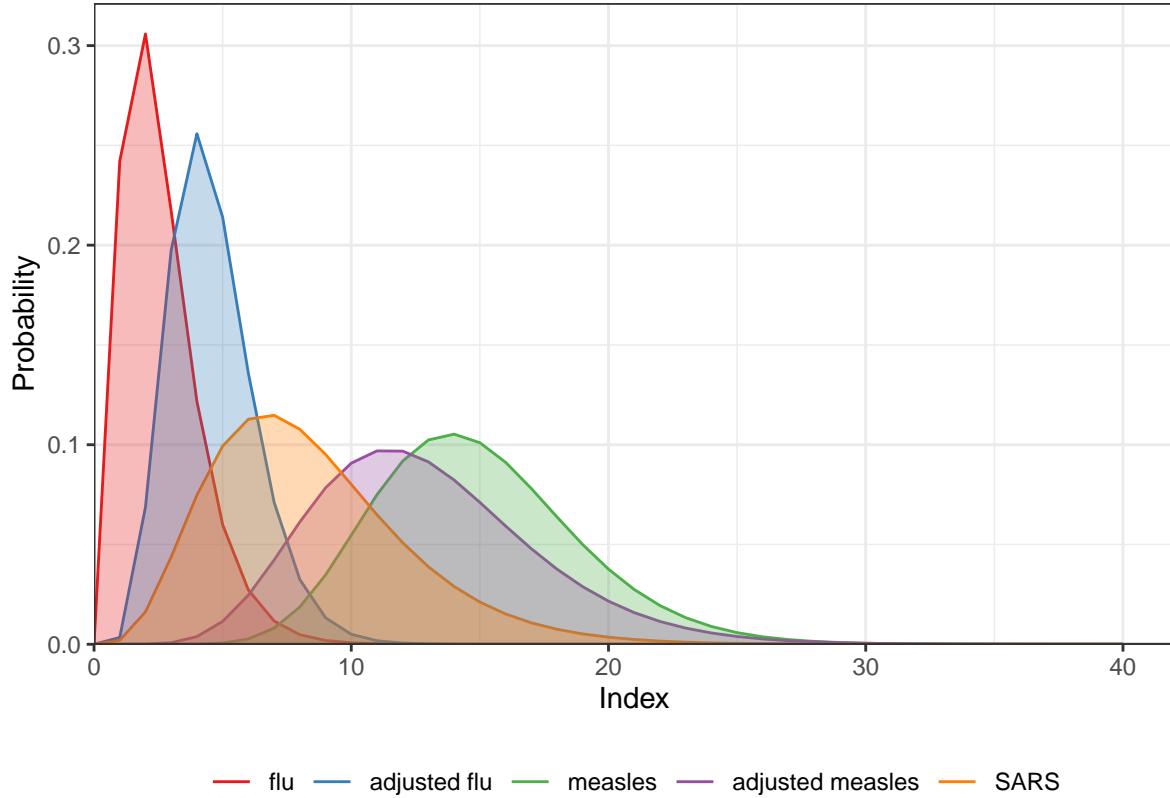


Figure A.2.2: Density curves of serial interval distributions used in the experiments

## A.3 Supplementary experimental results on accuracy comparison

### A.3.1 Long epidemics

We display the accuracy of all methods (where **EpiEstim** uses weekly sliding window) for measles and SARS sample epidemics (by excluding the first weeks in computing KL divergence) in Fig 3 and Fig 4, where we exclude the outliers. A full visualization is in A.3.1.

Figure A.3.2 compares **EpiEstim** with *monthly* sliding windows with other methods. We average the KL divergence per coordinate excluding the timepoints in the first months for all approaches, since **EpiEstim** estimates with the monthly sliding windows are not available until the second months. The *y*-axis is displayed on a logarithmic scale for a better visualization, since a few values are much larger than others.

The relative performance of **EpiEstim** with monthly sliding windows, in general, is not as good as its weekly sliding window based on the relative positions of its boxes and the counterparts of the other methods, except for the Scenario 2 with negative Binomial incidence. It can be explained by **EpiEstim** with longer sliding windows assume similarity of neighbouring  $\mathcal{R}_t$  across longer periods, and thus, is smoother and less accurate compared to the one with shorter sliding windows.

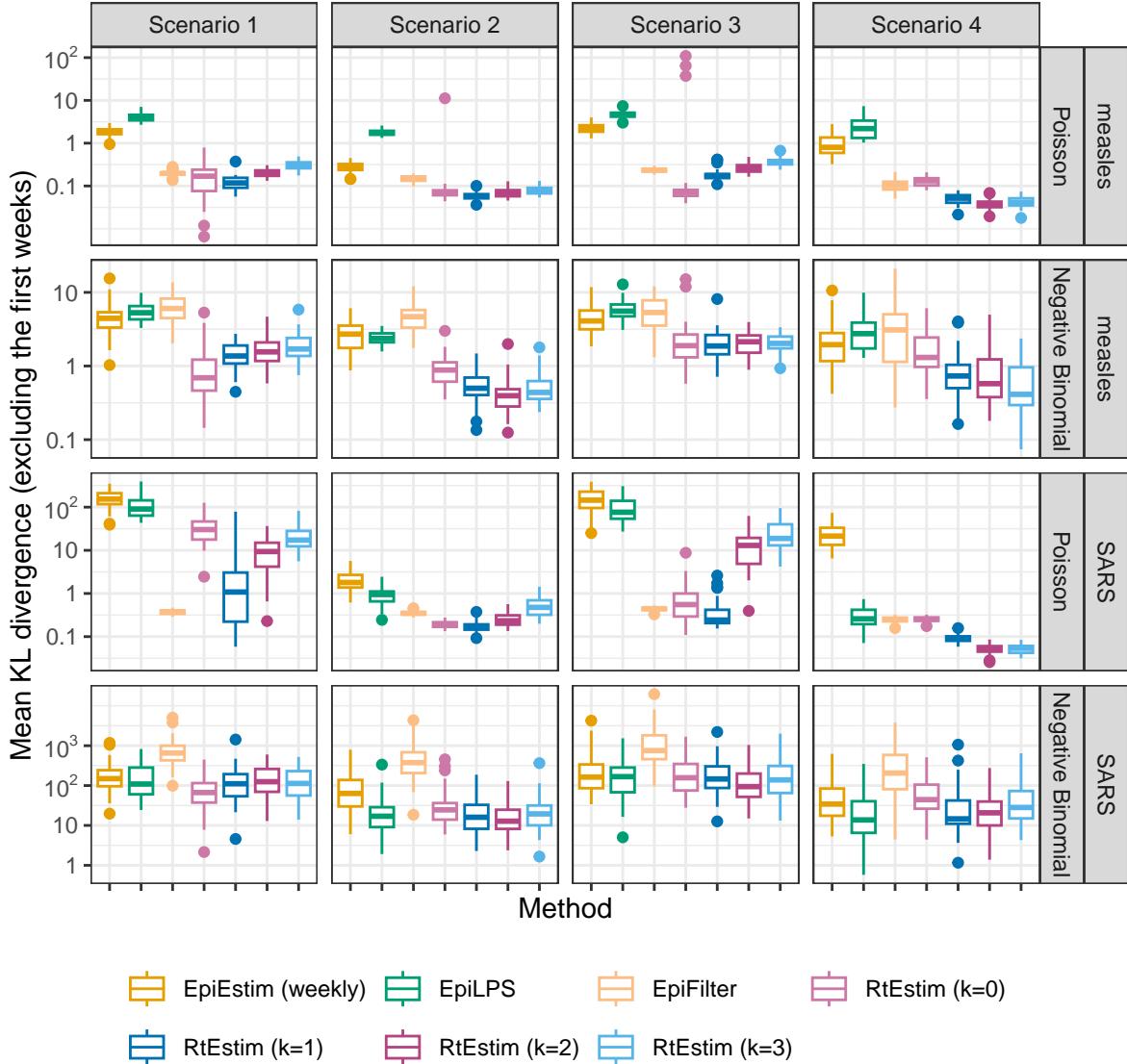


Figure A.3.1: Mean KL divergence excluding the first weeks for measles and SARS epidemics, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Y-axis is on a logarithmic scale.

### A.3.2 Short epidemics

Figures A.3.3 and A.3.4 display the KL divergence for short epidemics aggregated over per coordinate excluding the first weeks and months respectively.

## A.4 Experimental results on accuracy under misspecification of serial interval distributions

### A.4.1 SI misspecification for long epidemics

Figures A.4.1 and A.4.2 display KL divergence (excluding first weeks and months respectively) for all 8 methods with mild misspecification (shaped and scaled **measles** SI parameters) and major misspecification

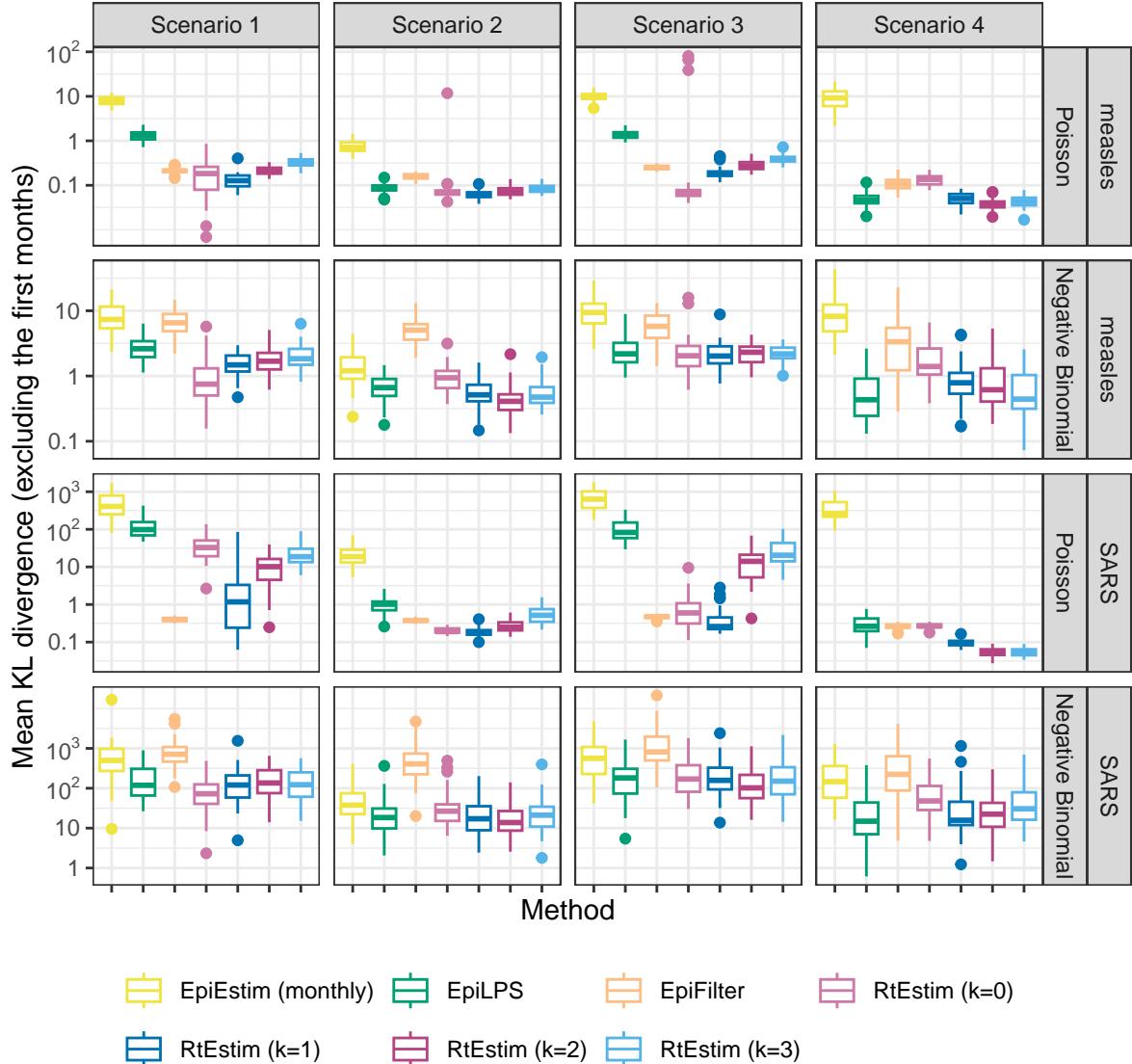


Figure A.3.2: Mean KL divergence excluding the first months for measles and SARS epidemics, since EpiEstim with the monthly sliding window does not provide estimates for the first month. Y-axis is on a logarithmic scale.

(SARS SI parameters) for long `measles` epidemics across all settings.

#### A.4.2 SI misspecification for short epidemics

Figures A.4.3 and A.4.4 display KL divergence (excluding first weeks and months respectively) for all 9 methods with minor misspecification (shaped and scaled `flu` SI parameters) and major misspecification (measles parameters) for short `flu` epidemics across all settings.

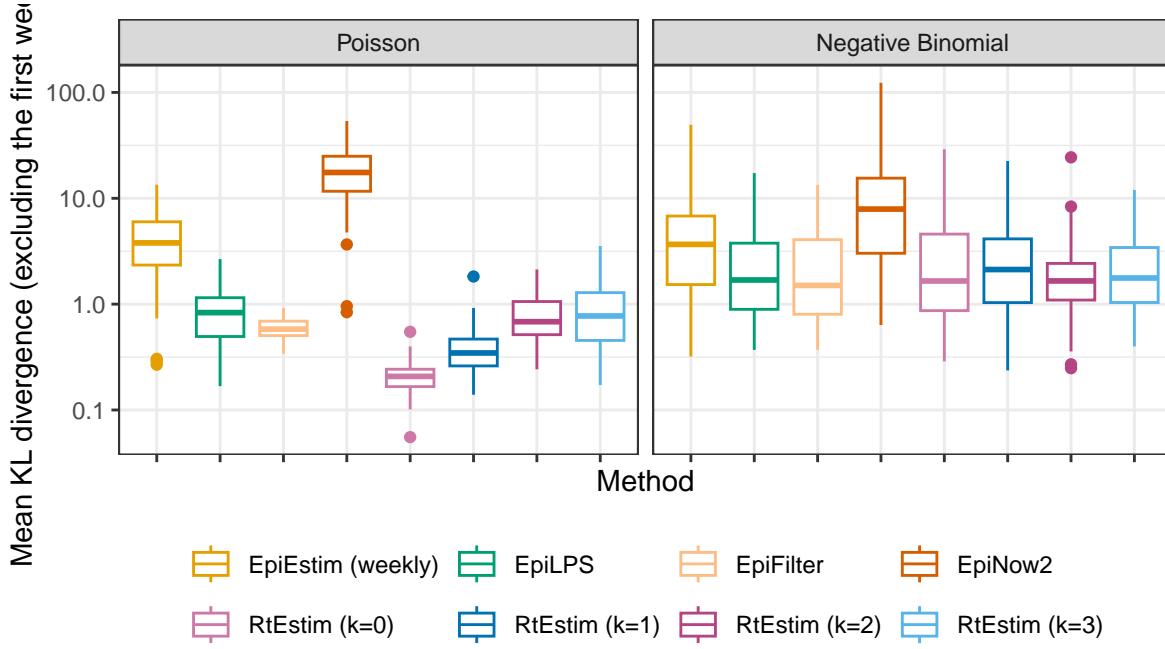


Figure A.3.3: Mean KL divergence excluding the first weeks for flu epidemics, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Y-axis is on a logarithmic scale.

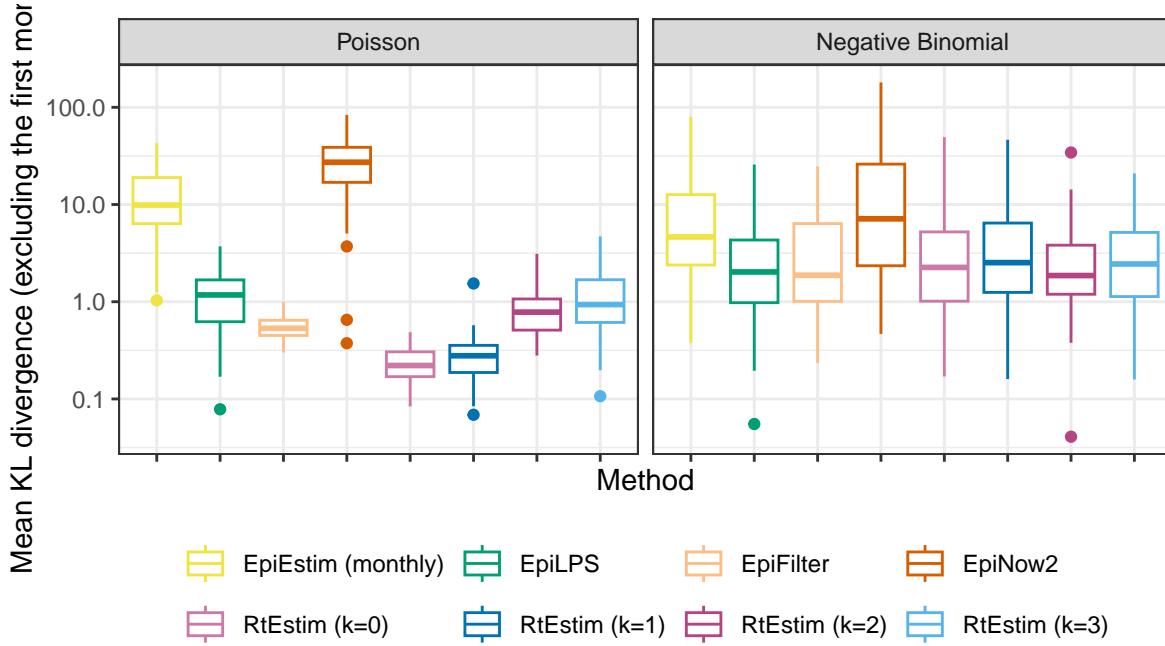


Figure A.3.4: Mean KL divergence excluding the first months for flu epidemics, since EpiEstim with the monthly sliding window does not provide estimates for the first month. Y-axis is on a logarithmic scale.

## A.5 Time comparisons of all methods

Figures A.5.1 show the time comparisons across all methods for long (measles and SARS) epidemics. **EpiEstim** with both weekly and monthly sliding windows are very fast and converge in less than 0.1 seconds. Piecewise

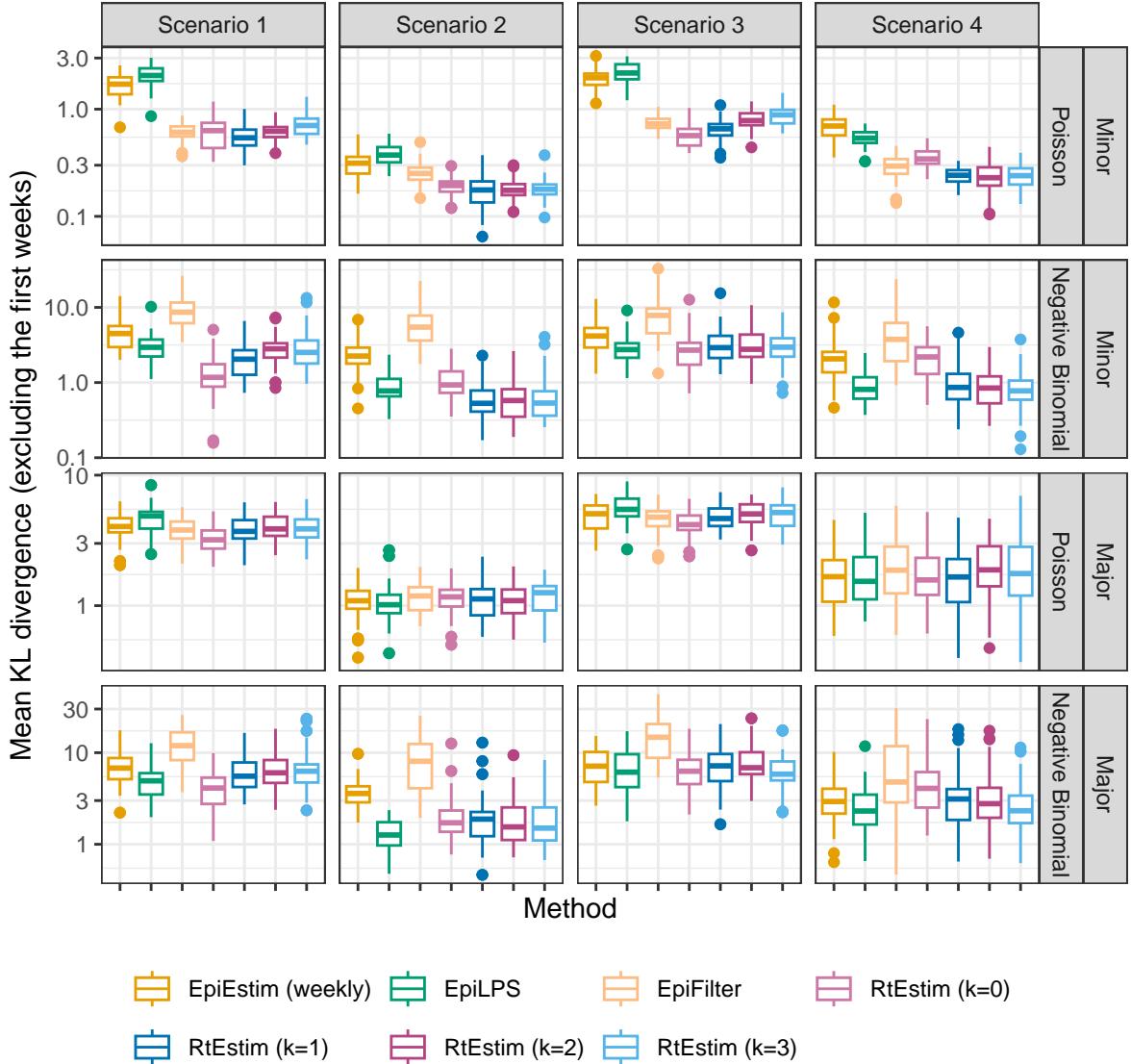


Figure A.4.1: Mean KL divergence excluding the first weeks for measles epidemics with SI misspecification, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Y-axis is on a logarithmic scale.

constant **RtEstim** (with  $k=0$ ) estimates can be generated within 0.1 seconds as well. **EpiLPS** is slightly slower, but still very fast and within 1 second for all experiments. Piecewise linear and cubic **RtEstim** (with  $k = 1$  and  $k = 3$  respectively) are slower, but mostly within 10 seconds.

It is remarkable that our **RtEstim** computes 50 lambda values with 10-fold CV for each experiment, which results in 550 times of modelling per experiment (including modelling for all folds). The running times are no more than 10 seconds for most of the experiments, which means the running time for each time of modelling is very fast, and on average can be less than 0.02 seconds. The other two methods only run once for a fixed set of hyperparameters for each experiment.

We visualize the running times of each case in separate panels in Figures A.5.2 and A.5.3 for measles and SARS epidemics respectively. Similar results as in Figure A.5.1 are found in each setting.

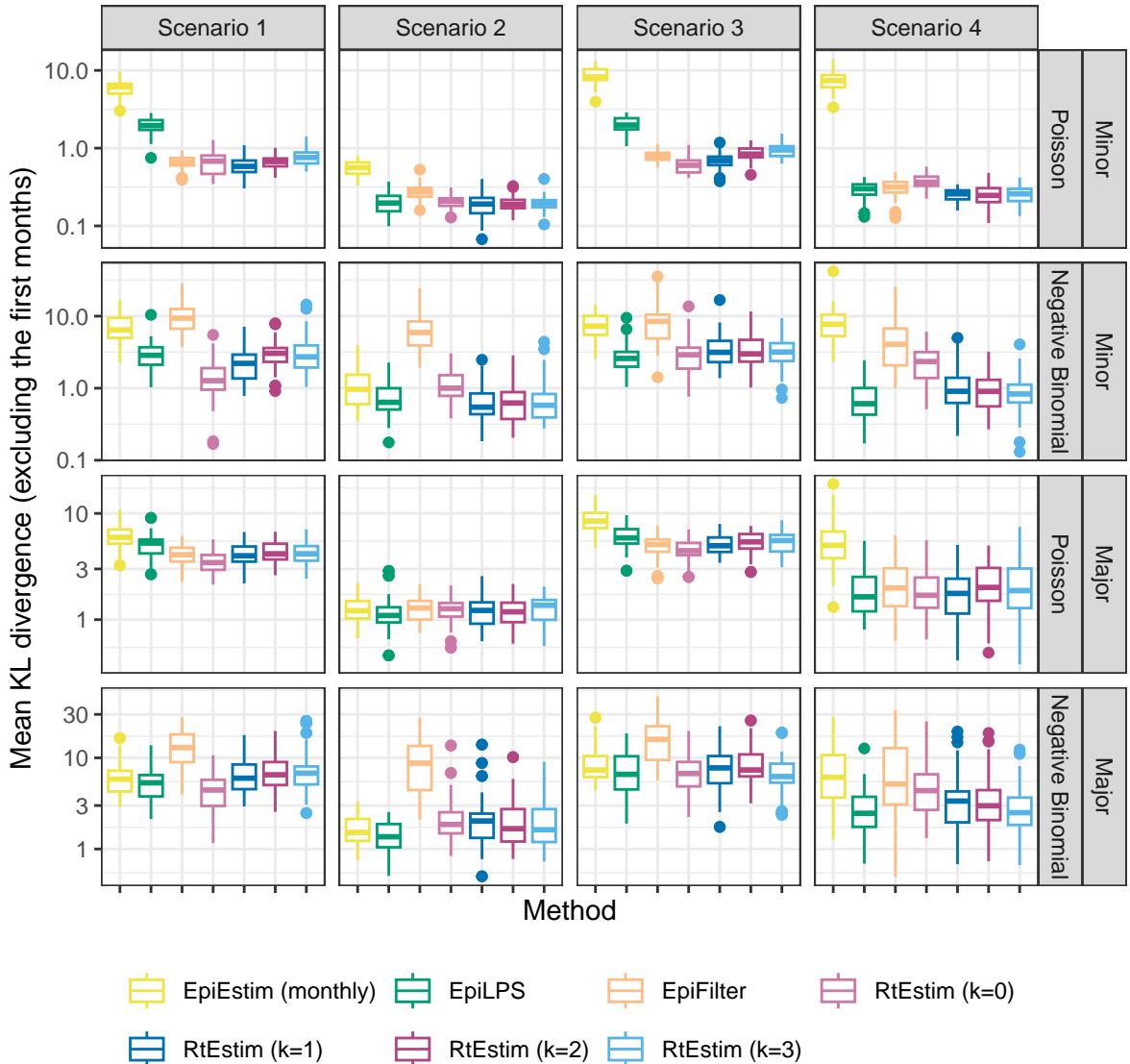


Figure A.4.2: Mean KL divergence excluding the first months for measles epidemics with SI misspecification, since EpiEstim with the monthly sliding window does not provide estimates for the first month. Y-axis is on a logarithmic scale.

Figure A.5.4 displays the running times of all methods for short (flu) epidemics. Figure A.5.5 displays the running times for each setting separately, and finds similar results as in the overall running time comparison.

## A.6 Confidence interval coverage

### A.6.1 Display estimates and confidence intervals for sample epidemics

Let's take a clearer view of the estimated  $\mathcal{R}_t$  with 95% confidence intervals for the sample long epidemics by all methods in Fig 5 and Fig 6 in Figures A.6.1 and A.6.4 respectively. The full view of other example epidemics are visualized in Figures A.6.2 and A.6.3 as follows.

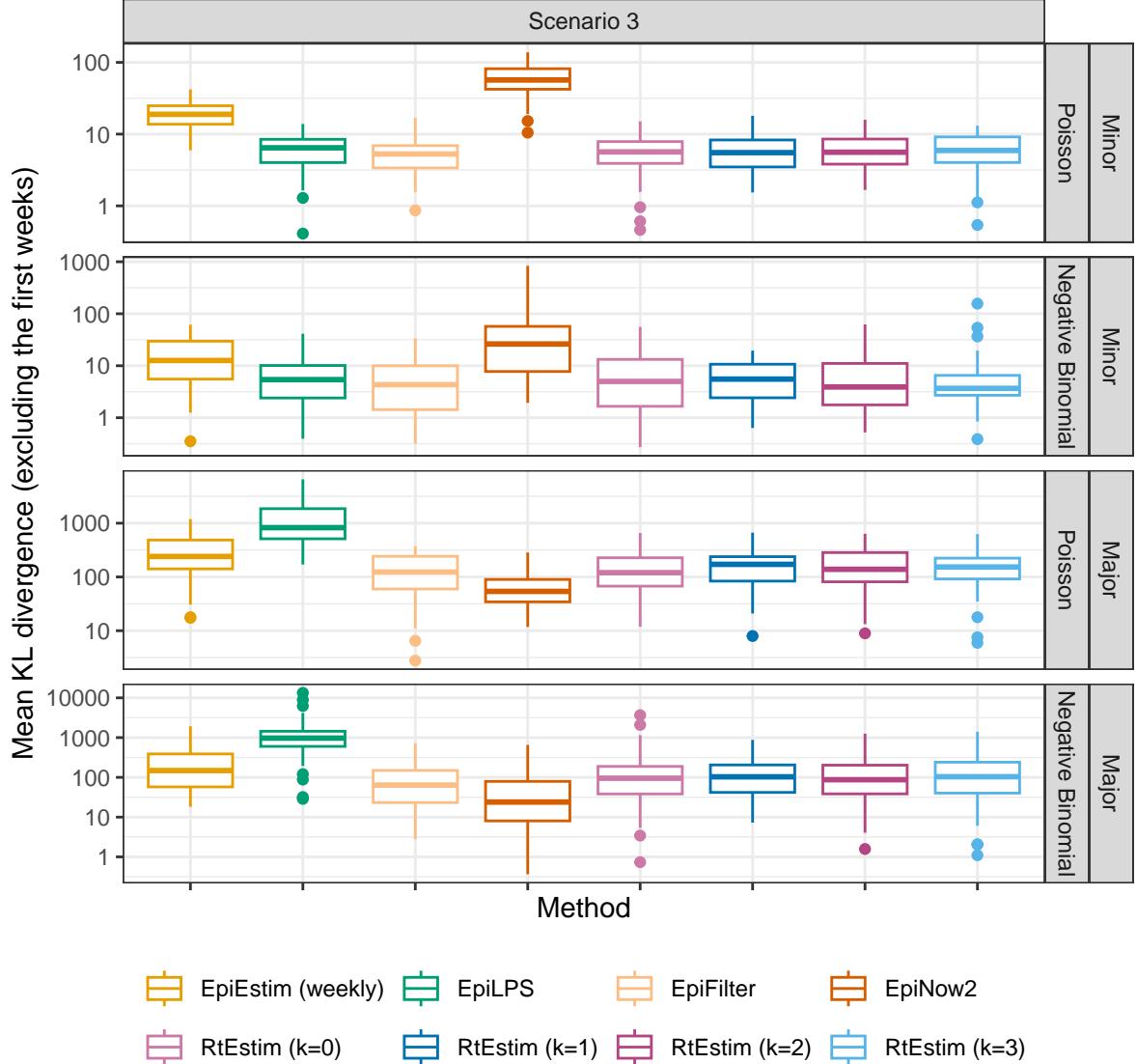


Figure A.4.3: Mean KL divergence excluding the first weeks for flu epidemics with SI misspecification, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Y-axis is on a logarithmic scale.

## A.6.2 Experimental settings on coverage level comparison of confidence intervals

We focus on a specific  $\mathcal{R}_t$  scenario, the piecewise linear case, and only long epidemics to compare the coverage of 95% confidence intervals across all 8 methods. We use the true serial interval distributions in modelling. Table 2 summarizes the experimental settings.

## A.6.3 Experimental results on interval coverage comparison

Figures A.6.5 and A.6.6 displays the percentages of coverage of 95% CI per coordinate over 50 random samples for measles and SARS epidemics respectively.

Figures A.6.7 and A.6.8 displays the percentages of coverage of 95% CI across all timepoints averaged over 50 random measles and SARS epidemics respectively.

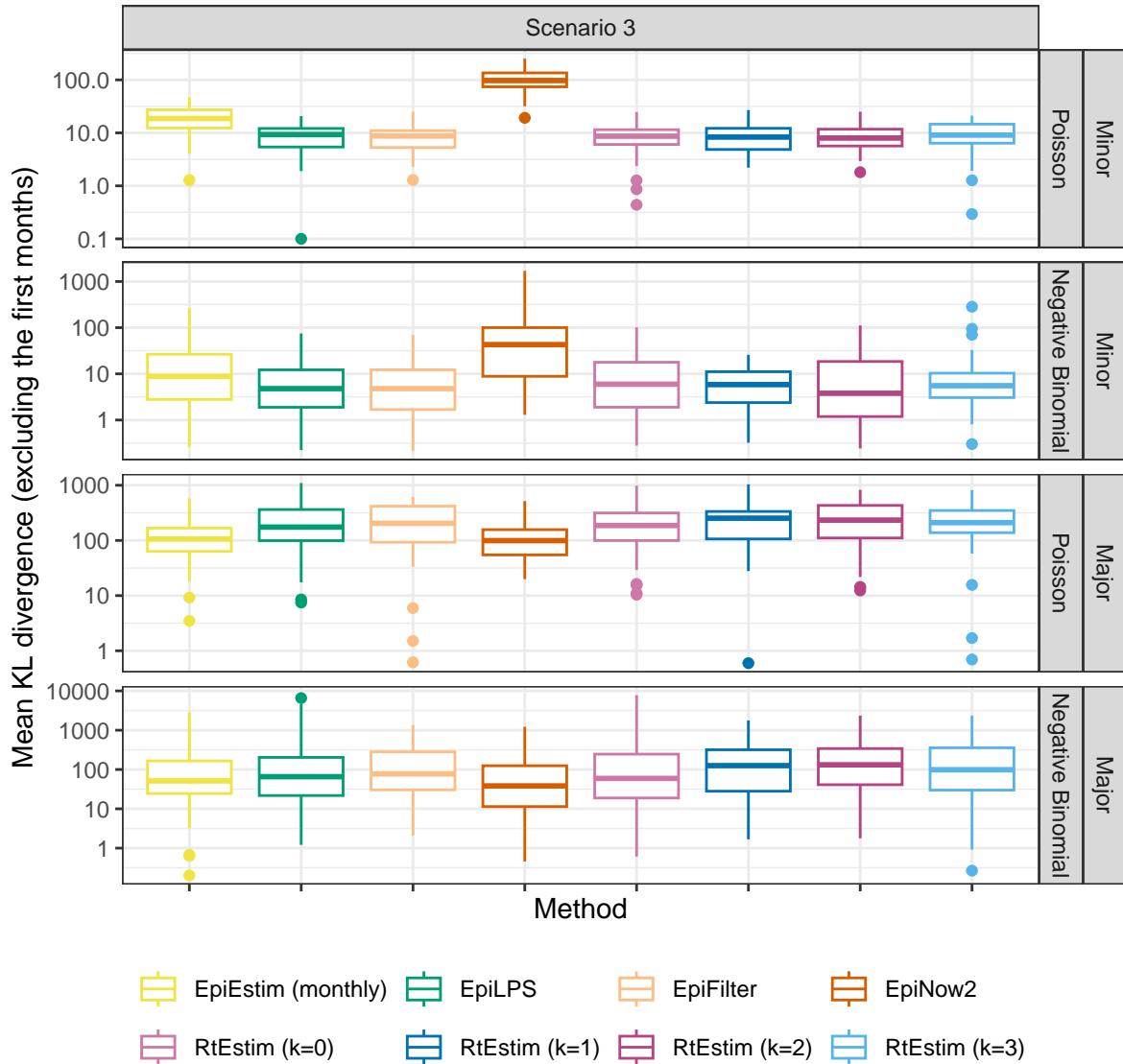


Figure A.4.4: Mean KL divergence excluding the first months for flu epidemics with SI misspecification, since EpiEstim with the monthly sliding window does not provide estimates for the first month. Y-axis is on a logarithmic scale.

We output a vector of the CI coverage for each timepoint per experiment, the percentage of coverage of all timepoints, and the interval score

$$score_\alpha(\mathcal{R}, u, l) = \frac{1}{n} \sum_{t=1}^n (u_t - l_t) + \frac{2}{\alpha} (l_t - \mathcal{R}_t) \mathbf{1}_{(\mathcal{R}_t < l_t)} + \frac{2}{\alpha} (\mathcal{R}_t - u_t) \mathbf{1}_{(\mathcal{R}_t > u_t)}$$

, where  $\alpha = 0.05$  is the significance level,  $l, u$  are the lower and upper bounds,  $\mathcal{R}_t$  is the true effective reproduction number, and  $\mathbf{1}_X$  is the indicator function of the condition  $X$ . (Bracher et al. 2021). Figures A.6.9 and A.6.10 displays the interval scores of 95% CI averaged over 50 random measles and SARS epidemics respectively.

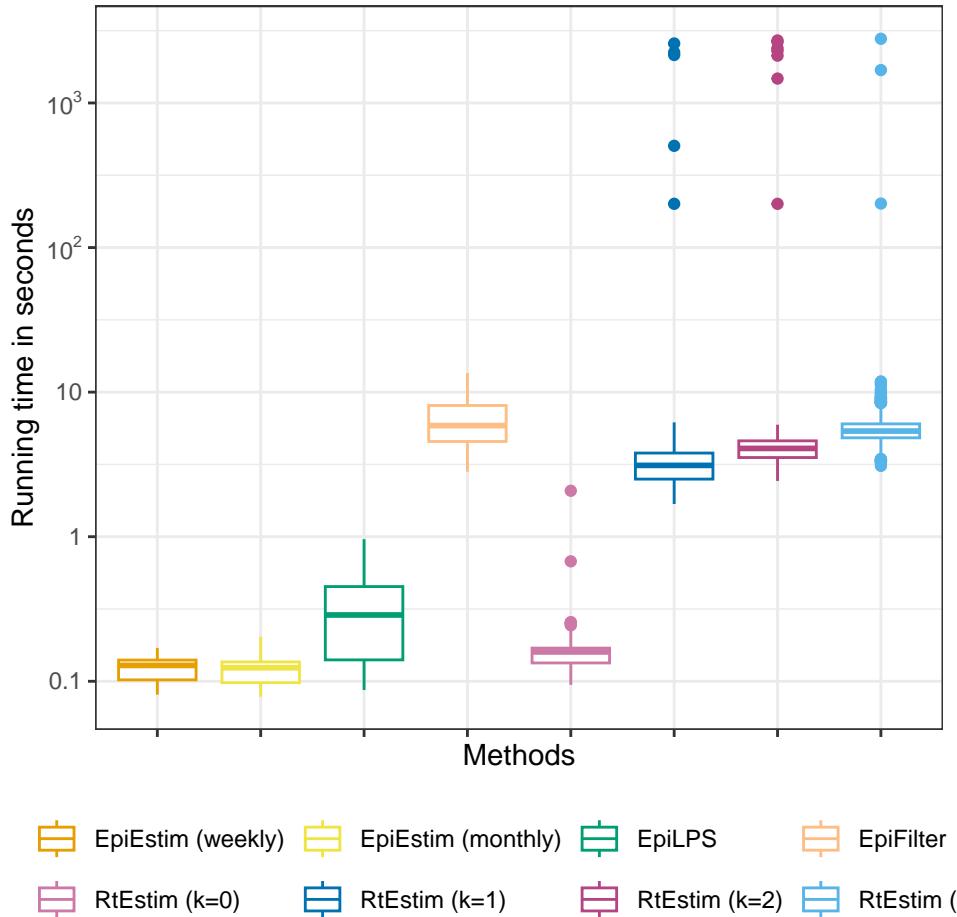


Figure A.5.1: Time comparisons of all methods for long (measles and SARS) epidemics across all cases. Y-axis is on a logarithmic scale.

## A.7 Data examples and alternative visualizations of Figs 5 and 6

### A.7.1 More visualization of example epidemics

We generate measles and SARS epidemics using Poisson and negative Binomial incidence distributions for each experimental settings. The condensed display of estimates for measles with Poisson incidence and SARS with negative Binomial incidence are provided in Fig 5 and Fig 6. A full visualization of each case is provided in Section A.6.1. The condensed visualization of other cases is provided below in Figures A.7.1 and A.7.2.

Table 2: Summary of experimental setting on coverage of confidence intervals

Length	SI	Rt scenario	Incidence	SI for modelling	Method
300	measles		3	Poisson, NB	measles
300	SARS		3	Poisson, NB	SARS

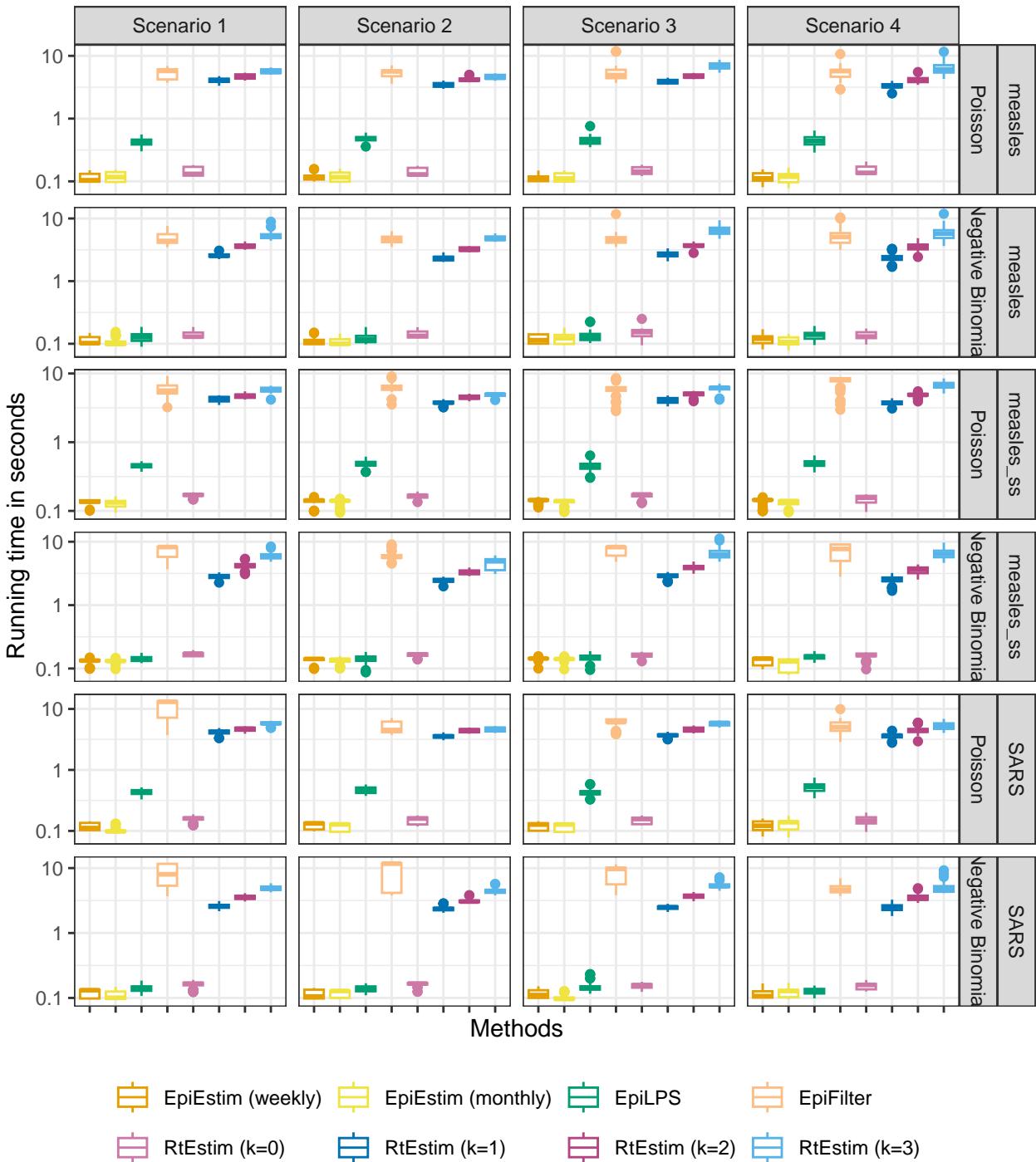


Figure A.5.2: Time comparisons of all methods for measles epidemics with each choice of SI parameter for modelling per incidence distribution per Rt scenario (excluding outliers for better illustration). Y-axis is on a logarithmic scale.

### A.7.2 Alternative view of difference between fitted and true Rt estimates

Here, we also provide an alternative view of Fig 5 & Fig 6 by plotting  $\mathcal{R}_t - \hat{\mathcal{R}}_t$  per coordinate  $t$  in A.7.3 and A.7.4 respectively. Figures A.7.5 and A.7.6 provide the alternative view of A.7.1 and A.7.2 respectively.

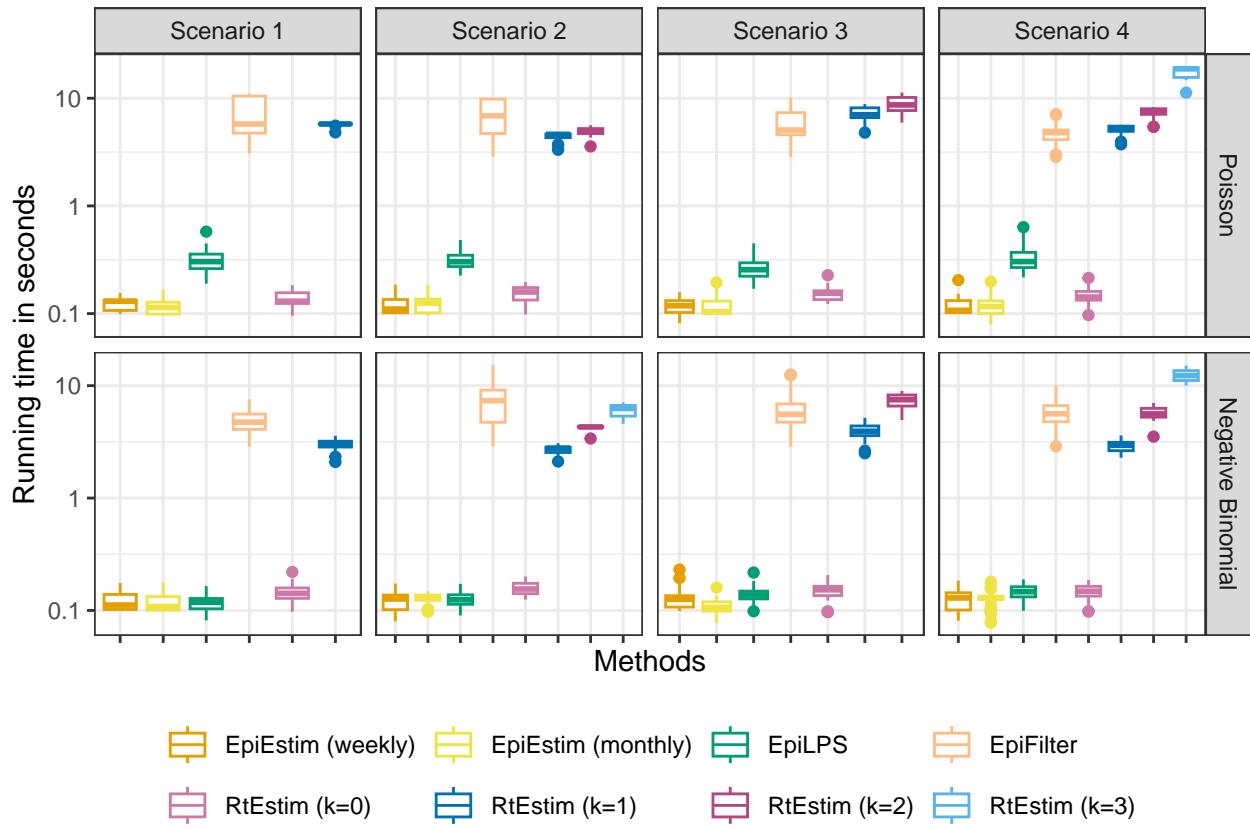


Figure A.5.3: Time comparisons of all methods for SARS epidemics with each choice of SI parameter for modelling per incidence distribution per Rt scenario (excluding outliers for better illustration). Y-axis is on a logarithmic scale.

## A.8 Application of RtEstim and all competitors on real epidemics

We apply all methods on Covid19 incidence in BC, and the estimated are displayed in ???. An alternative display which plot all estimated curves in one panel for an easier comparison is provided in ???.

We also apply all methods on Flu in 1918 as well. The results are visualized in Figures A.8.3 and A.8.4.

Boëlle, Pierre-Yves, Severine Ansart, Anne Cori, and Alain-Jacques Valleron. 2011. “Transmission Parameters of the a/H1N1 (2009) Influenza Virus Pandemic: A Review.” *Influenza and Other Respiratory Viruses* 5 (5): 306–16.

Bracher, Johannes, Evan L. Ray, Tilmann Gneiting, and Nicholas G. Reich. 2021. “Evaluating Epidemic Forecasts in an Interval Format.” Edited by Virginia E. Pitzer. *PLoS Computational Biology* 17 (2): e1008618. <https://doi.org/10.1371/journal.pcbi.1008618>.

Cori, Anne, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. 2013. “A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics.” *American Journal of Epidemiology* 178 (9): 1505–12.

Ferguson, Neil M, Derek AT Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S Burke. 2005. “Strategies for Containing an Emerging Influenza Pandemic in Southeast Asia.” *Nature* 437 (7056): 209–14.

Groendyke, Chris, David Welch, and David R Hunter. 2011. “Bayesian Inference for Contact Networks Given

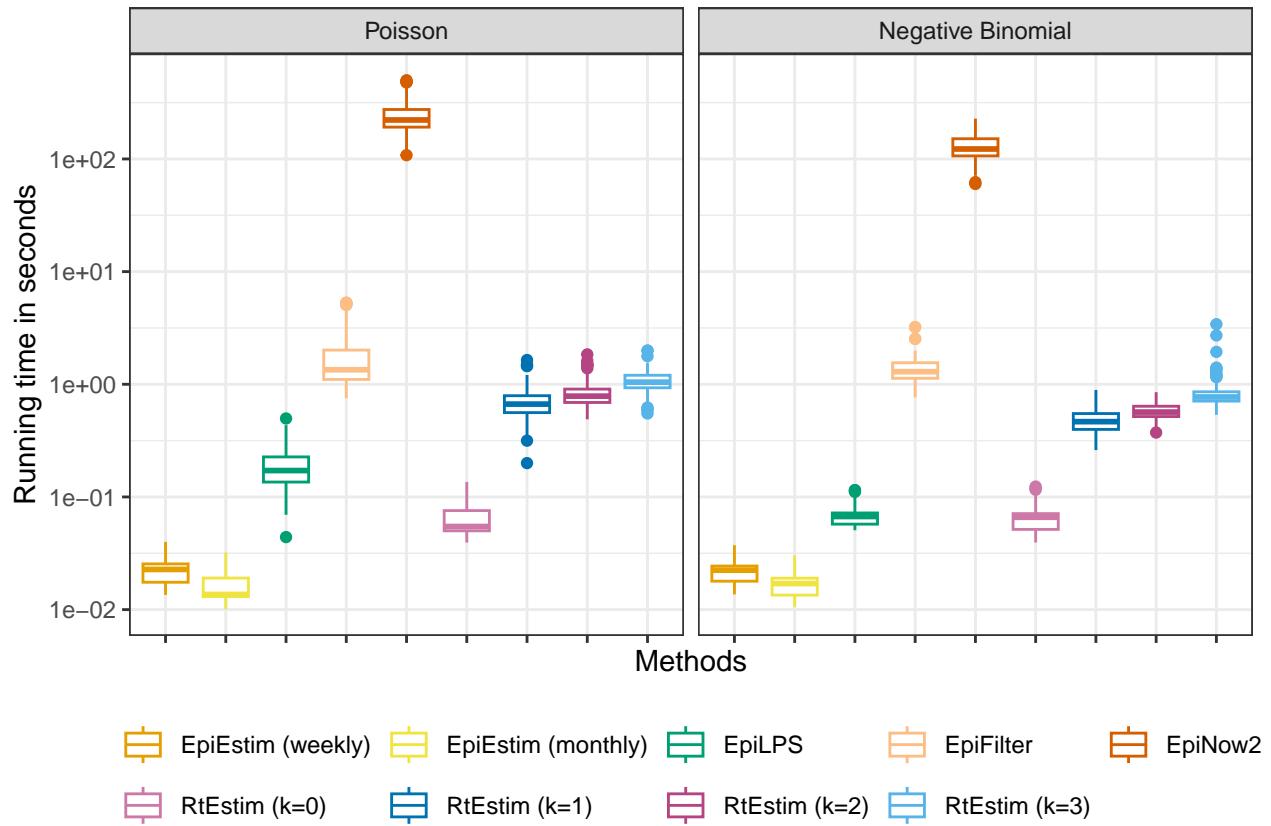


Figure A.5.4: Time comparisons of methods for short (flu) epidemics across all cases. Y-axis is on a logarithmic scale.

Epidemic Data." *Scandinavian Journal of Statistics* 38 (3): 600–616.

Lipsitch, Marc, Ted Cohen, Ben Cooper, James M Robins, Stefan Ma, Lyn James, Gowri Gopalakrishna, et al. 2003. "Transmission Dynamics and Control of Severe Acute Respiratory Syndrome." *Science* 300 (5627): 1966–70.

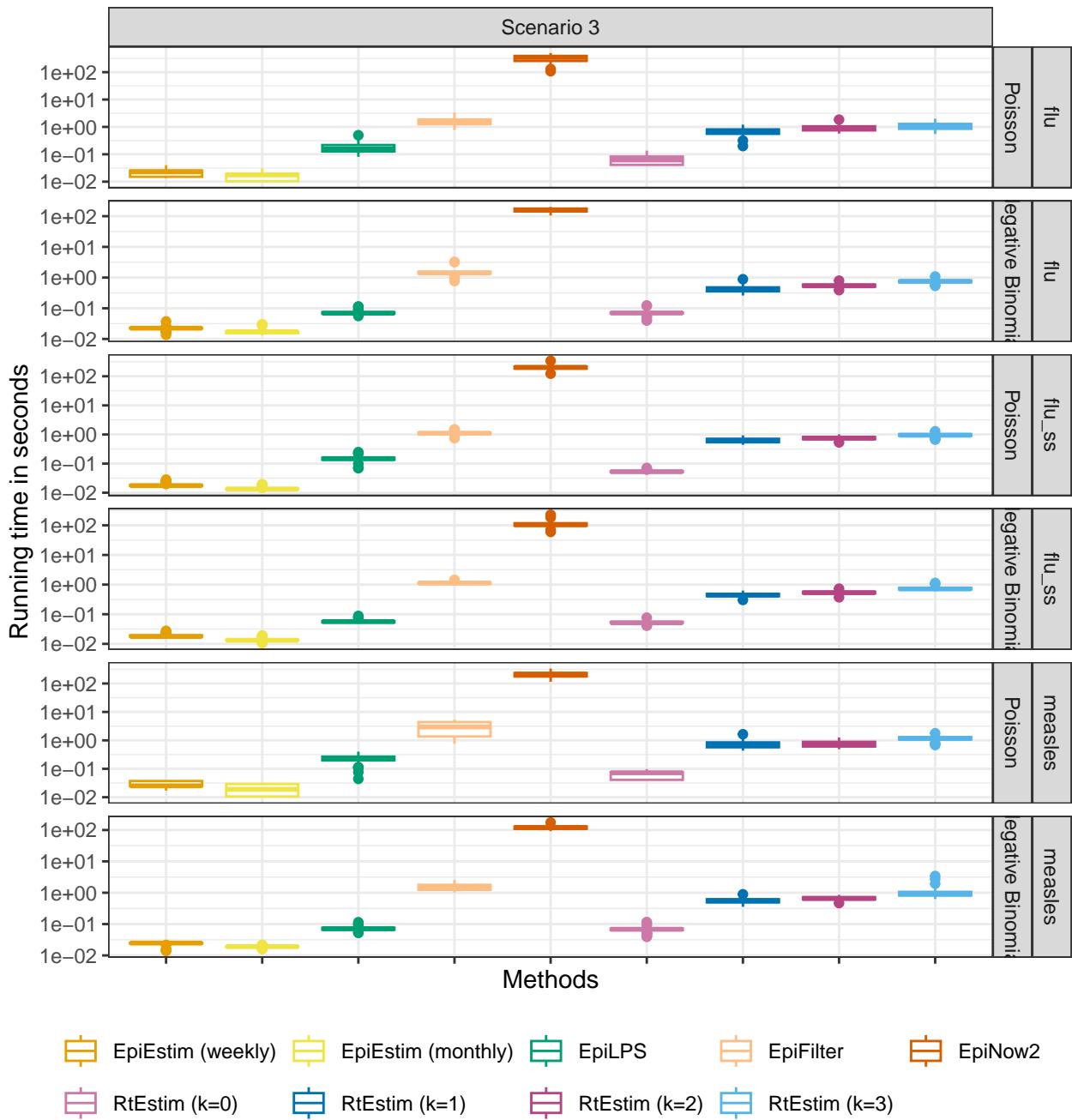


Figure A.5.5: Time comparisons of methods for short (flu) epidemics with different cases in different panels. Y-axis is on a logarithmic scale.

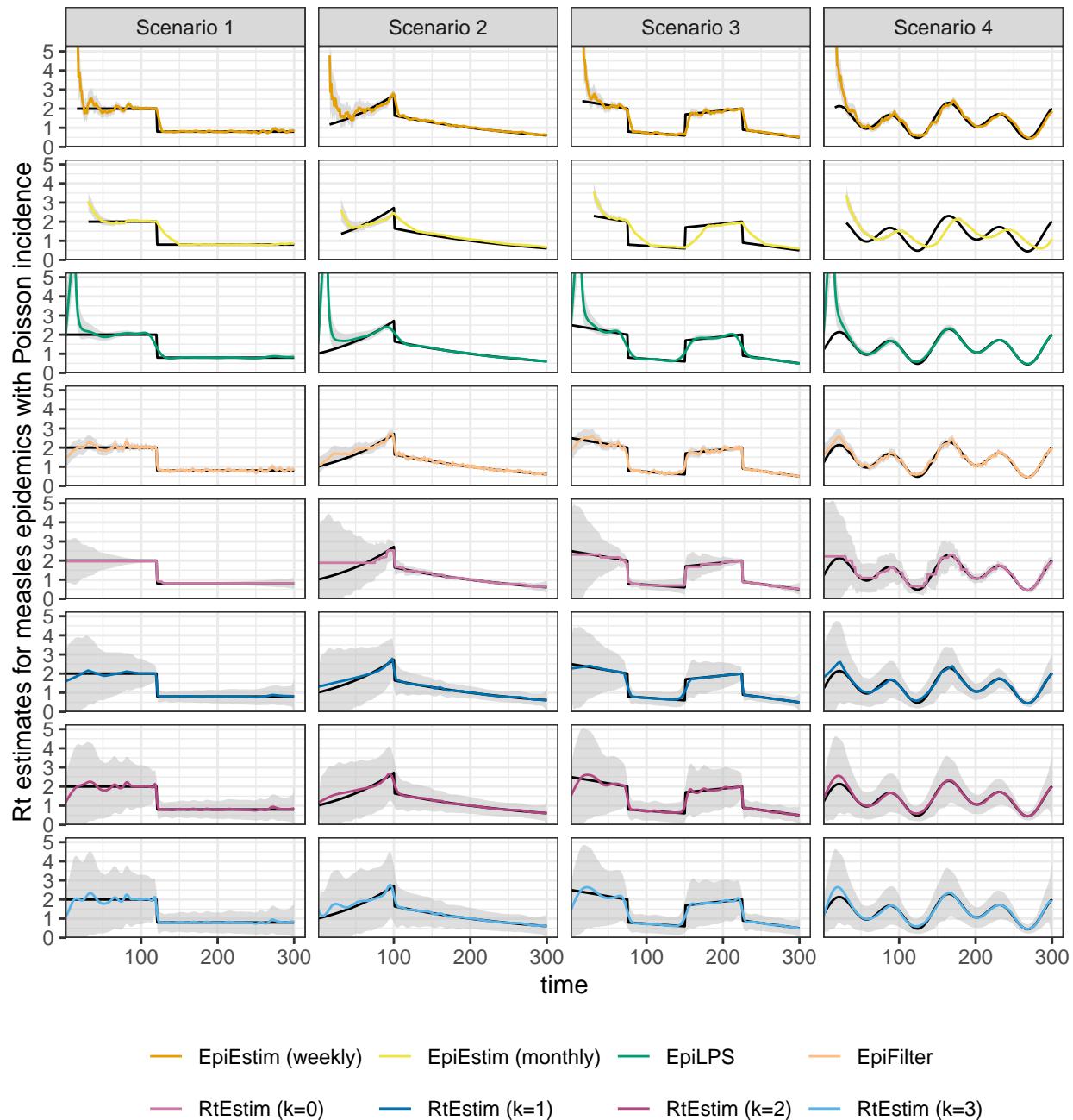


Figure A.6.1: Example measles epidemics with Poisson incidence.

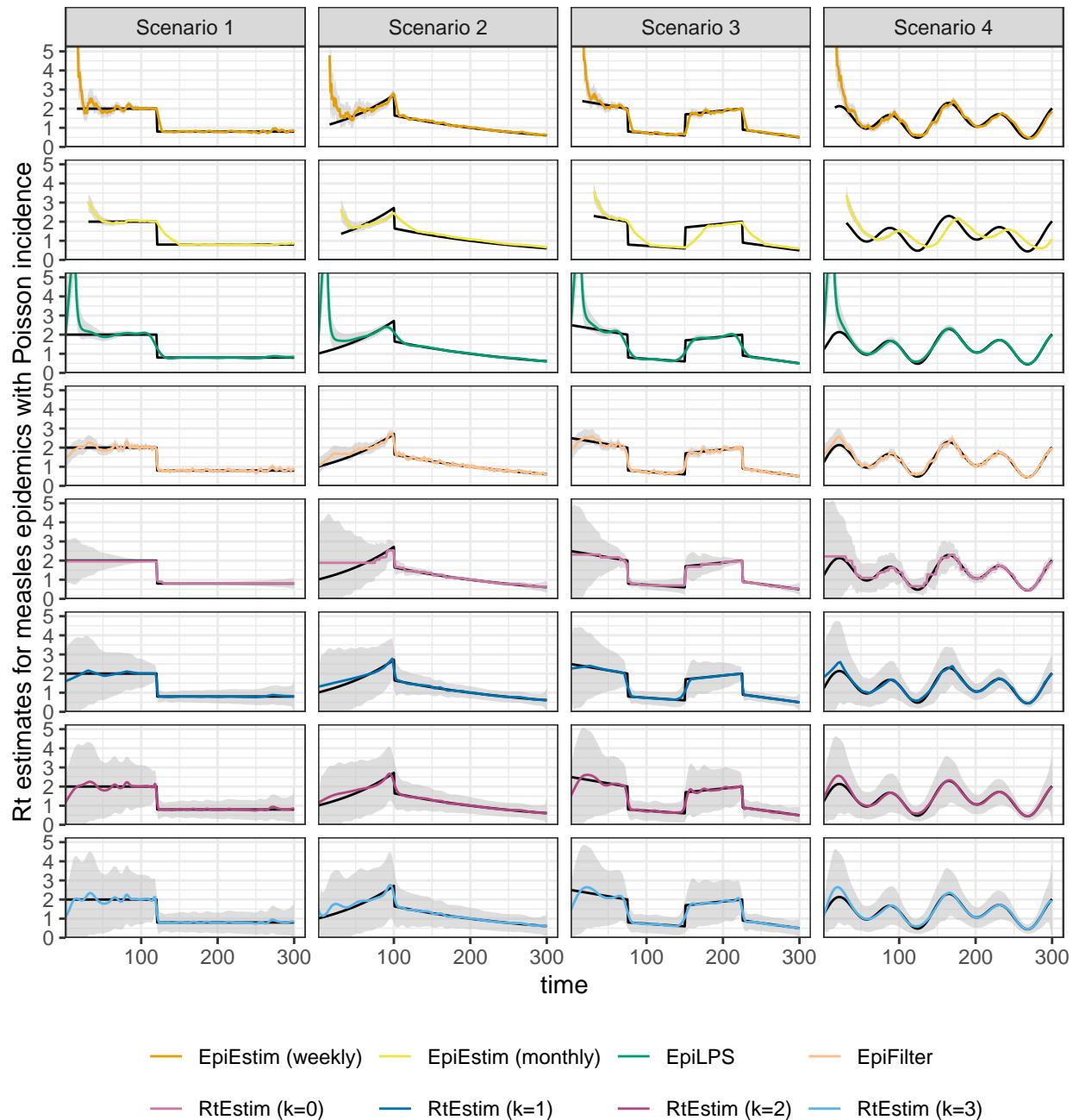


Figure A.6.2: Example measles epidemics with negative Binomial incidence.

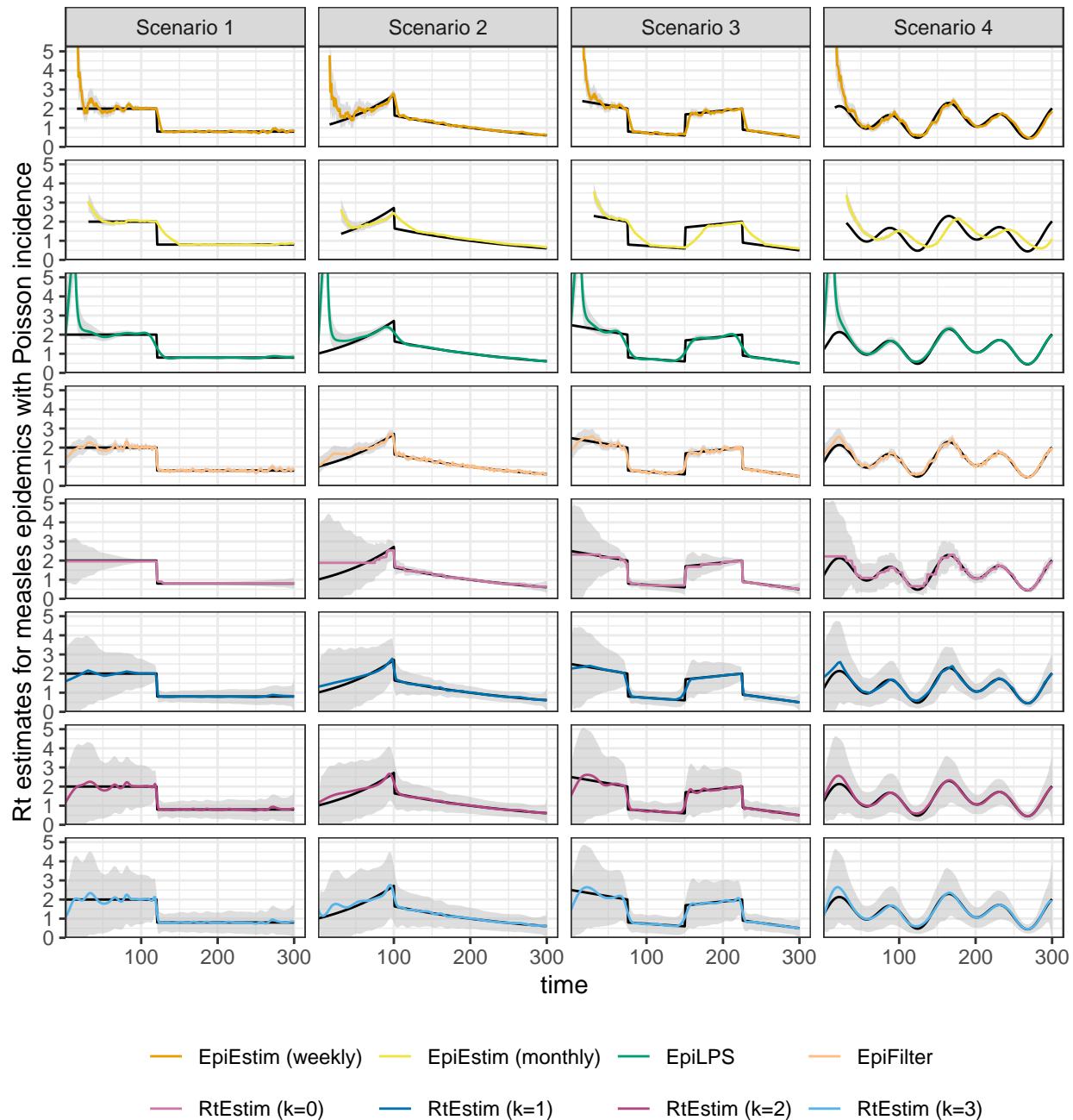


Figure A.6.3: Example SARS epidemics with Poisson incidence.

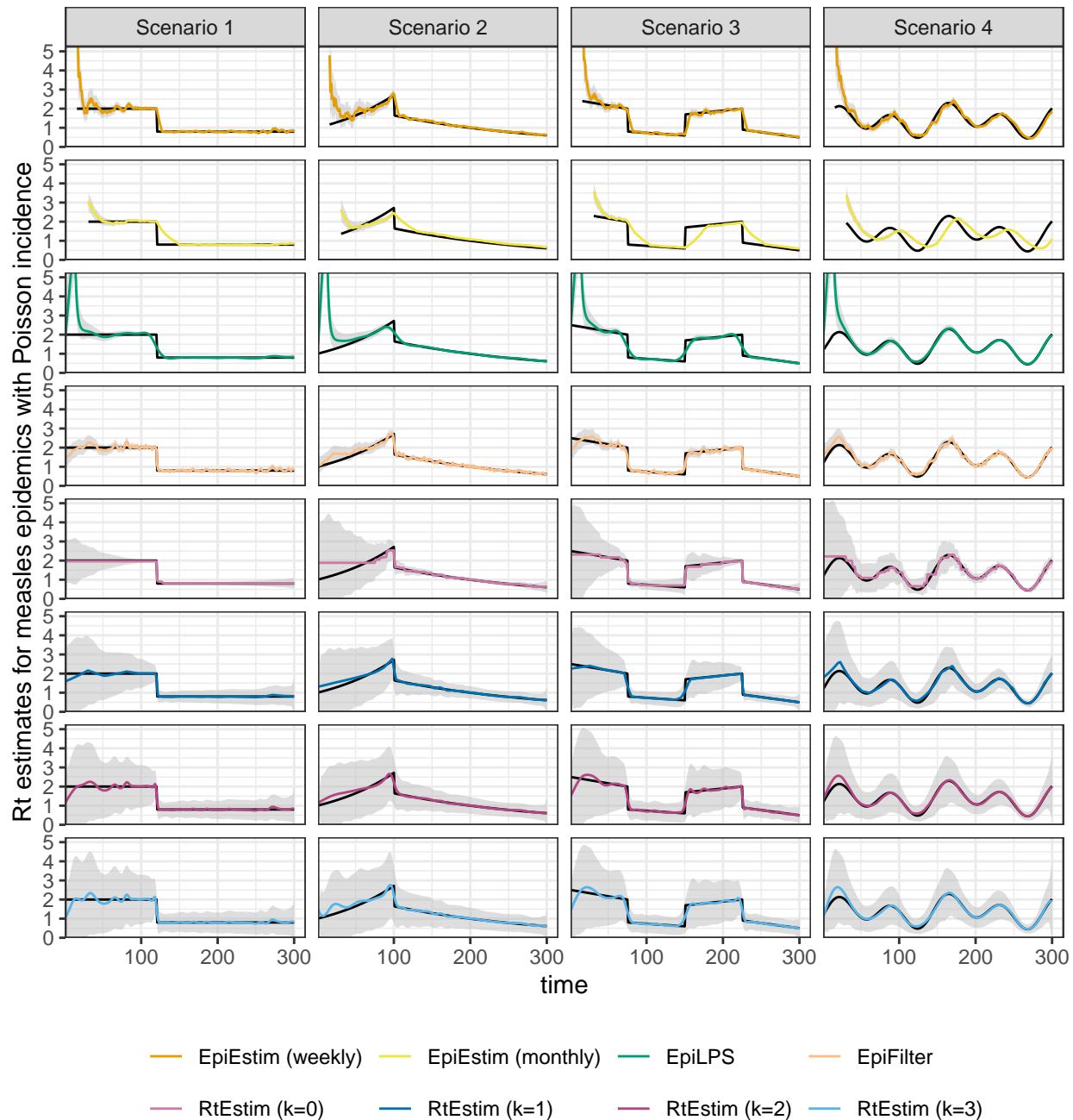


Figure A.6.4: Example SARS epidemics with negative Binomial incidence.

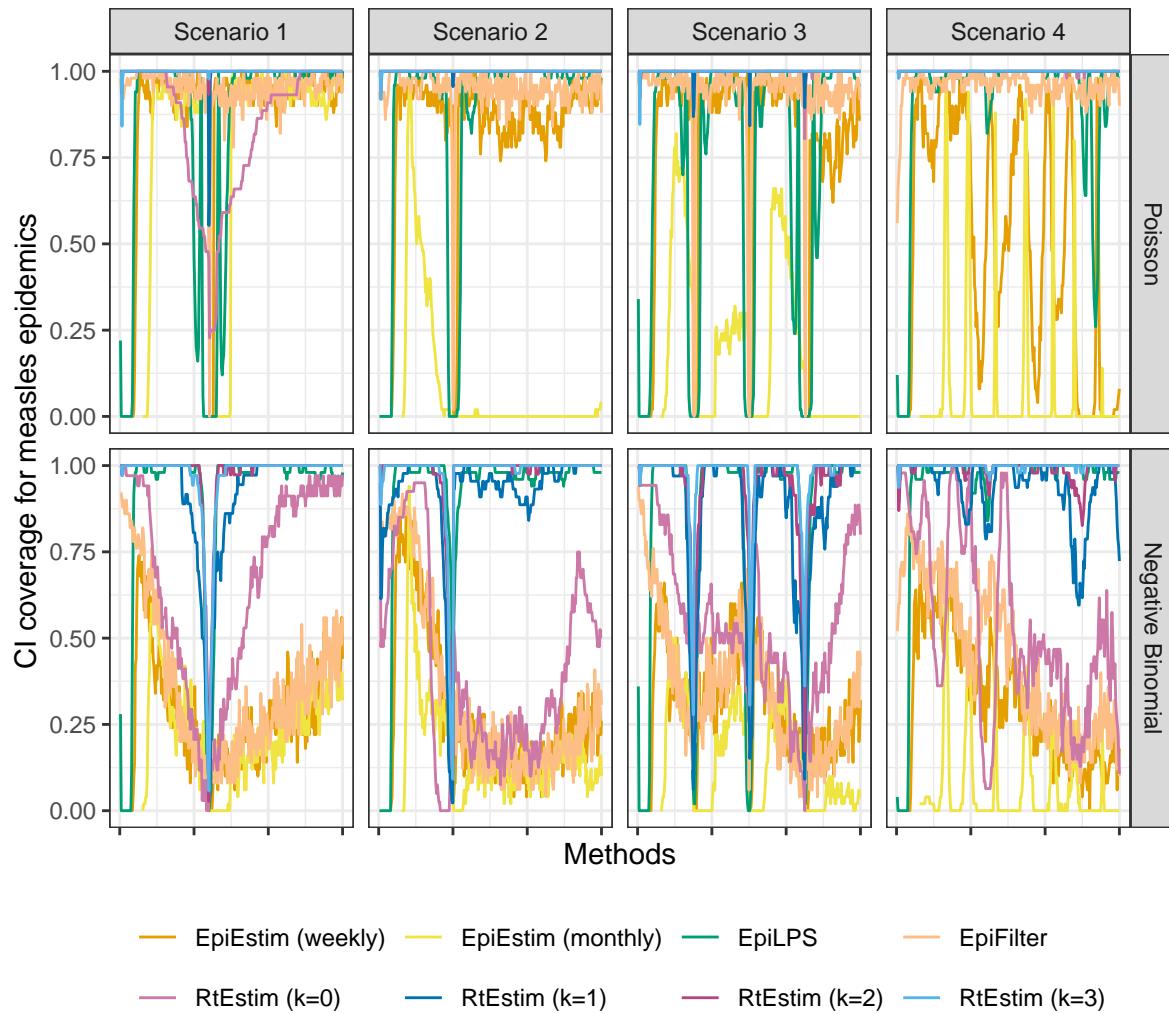


Figure A.6.5: Averaged coverage of CI per coordinate with measles epidemics.

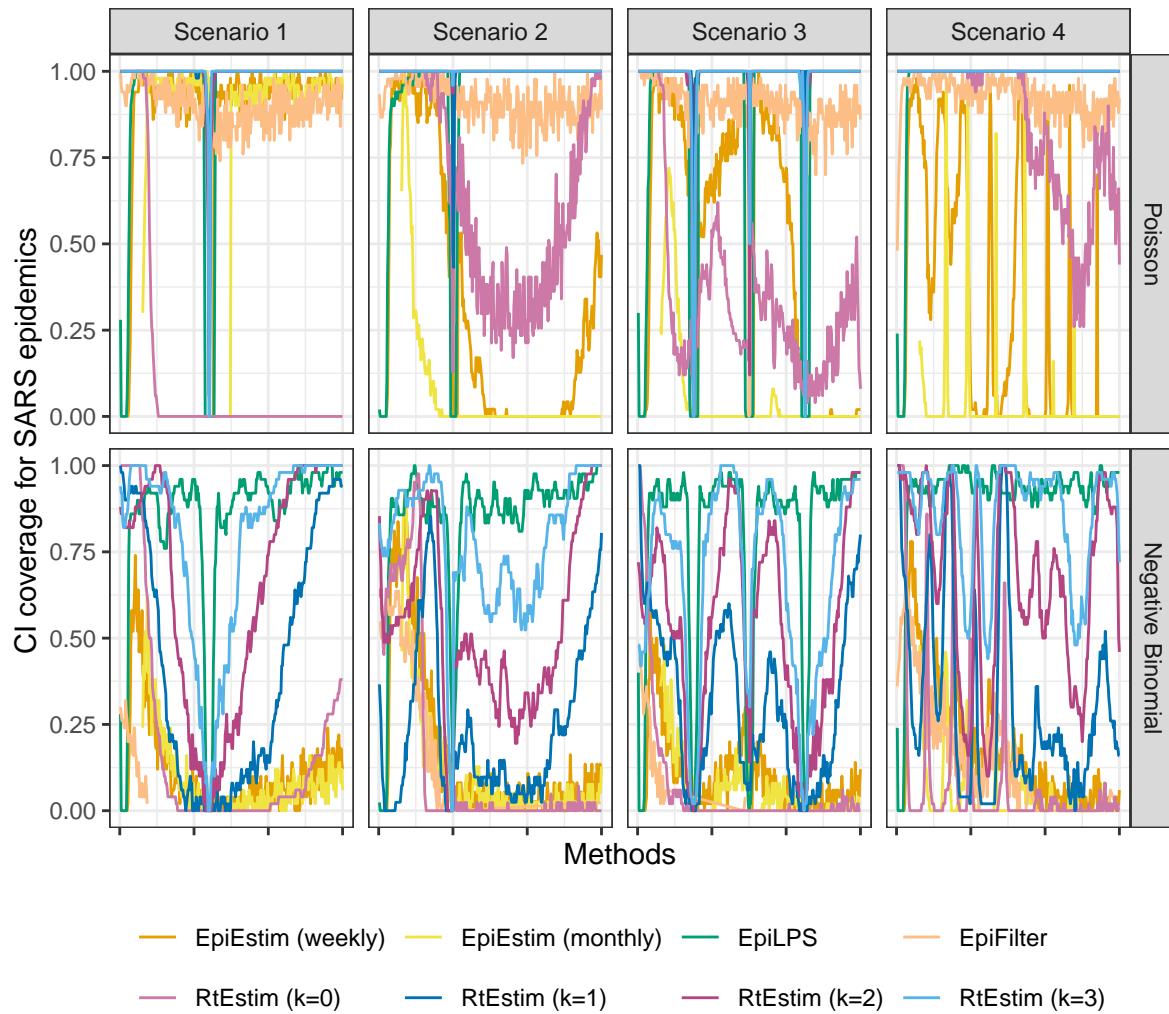


Figure A.6.6: Averaged coverage of CI per coordinate with SARS epidemics.

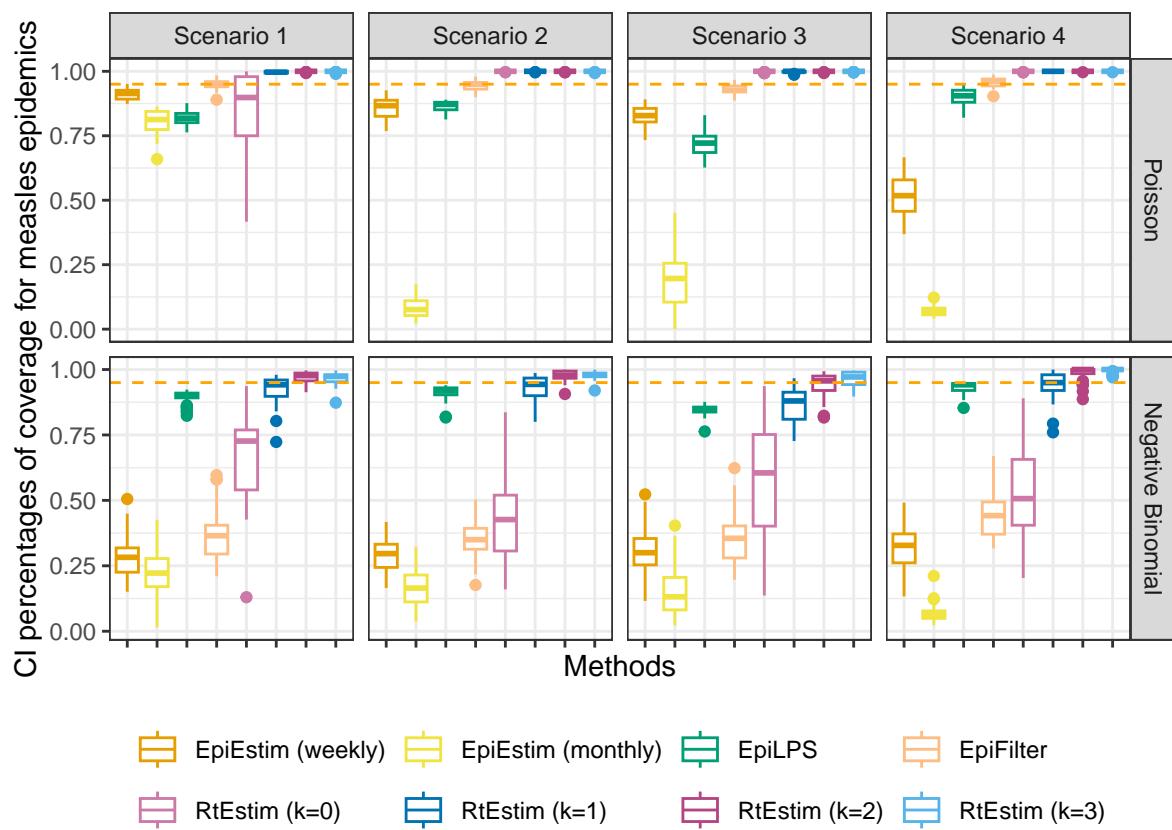


Figure A.6.7: Averaged percentages of CI coverage with measles epidemics.

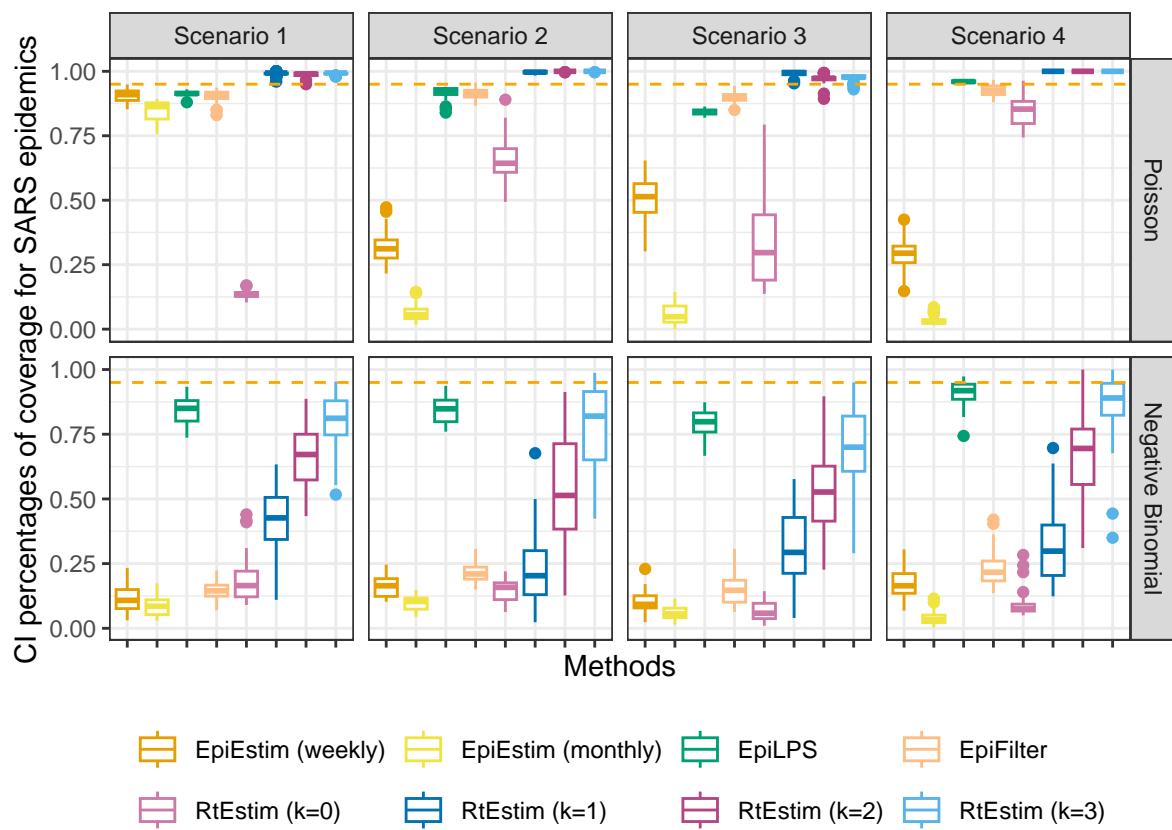


Figure A.6.8: Averaged percentages of CI coverage with SARS epidemics.

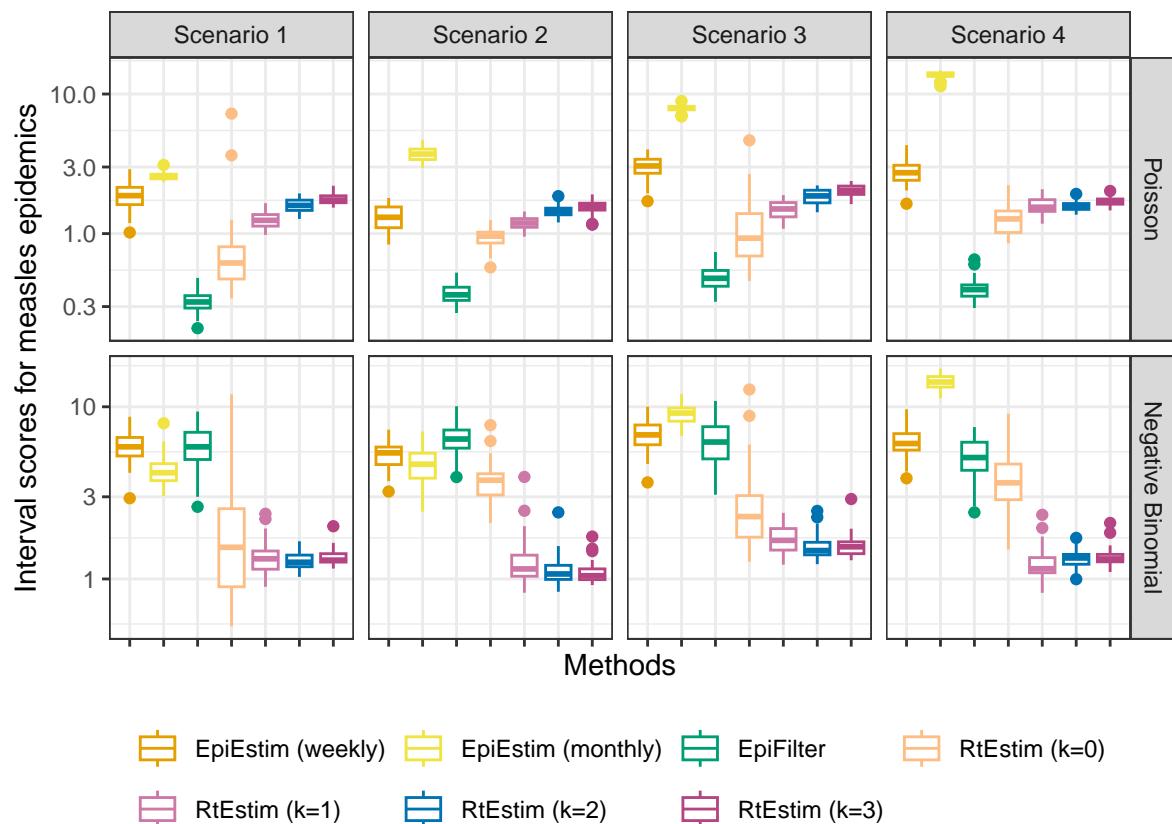


Figure A.6.9: Averaged interval scores with measles epidemics. EpiLPS is excluded, since it's scores are always larger than 100.

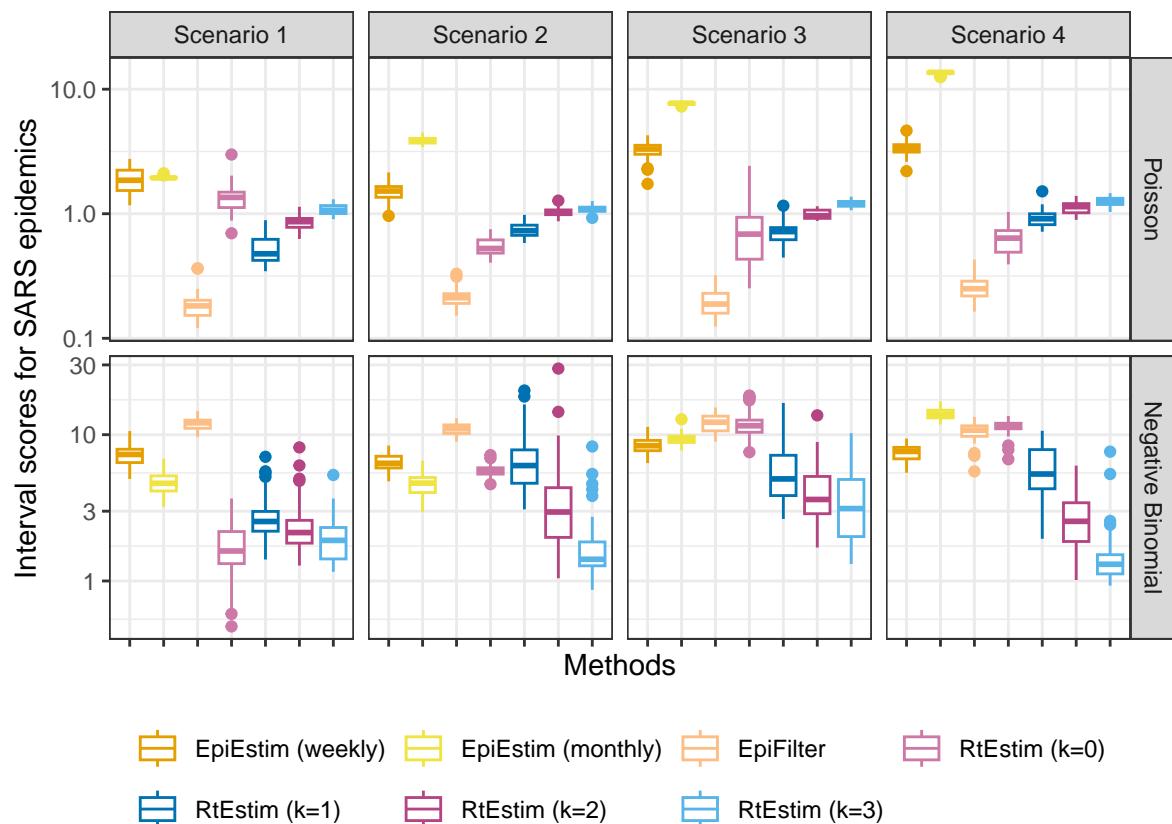


Figure A.6.10: Averaged interval scores with SARS epidemics. EpiLPS is excluded, since it's scores are always larger than 100.

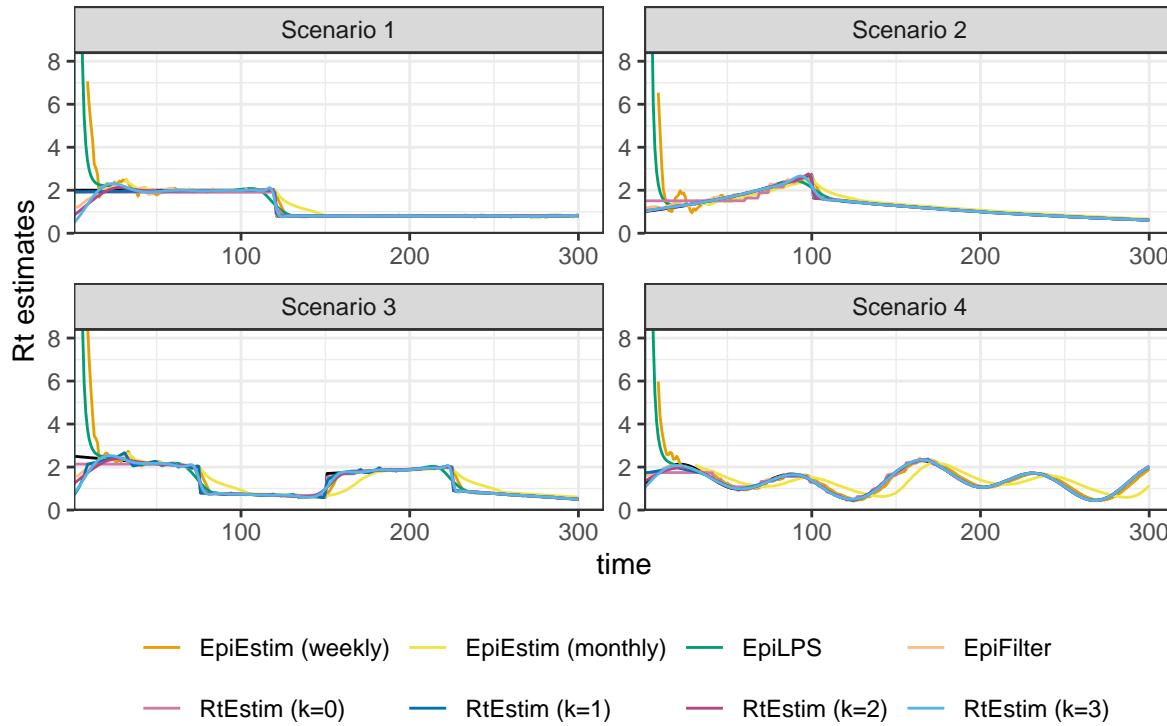


Figure A.7.1: Example of effective reproduction number estimation for SARS epidemics with Poisson observations.

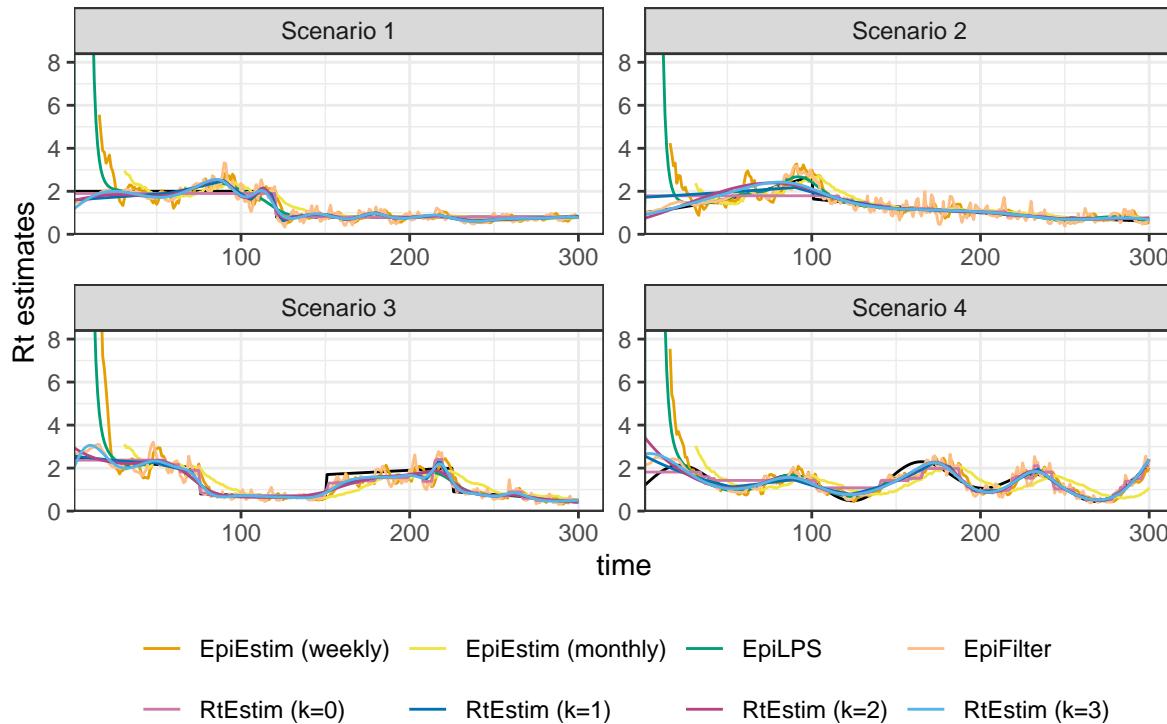


Figure A.7.2: Example of effective reproduction number estimation for measles epidemics with negative Binomial observations.

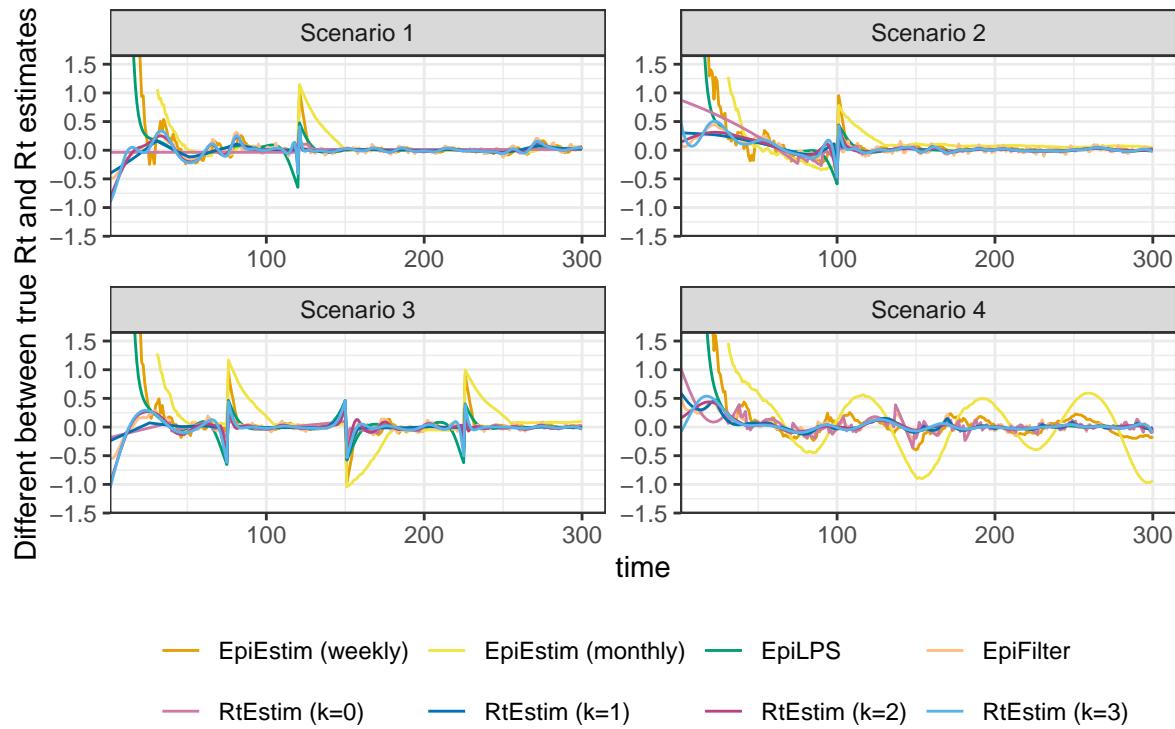


Figure A.7.3: Difference between of the true effective reproduction number and its estimation for measles epidemics with Poisson observations.

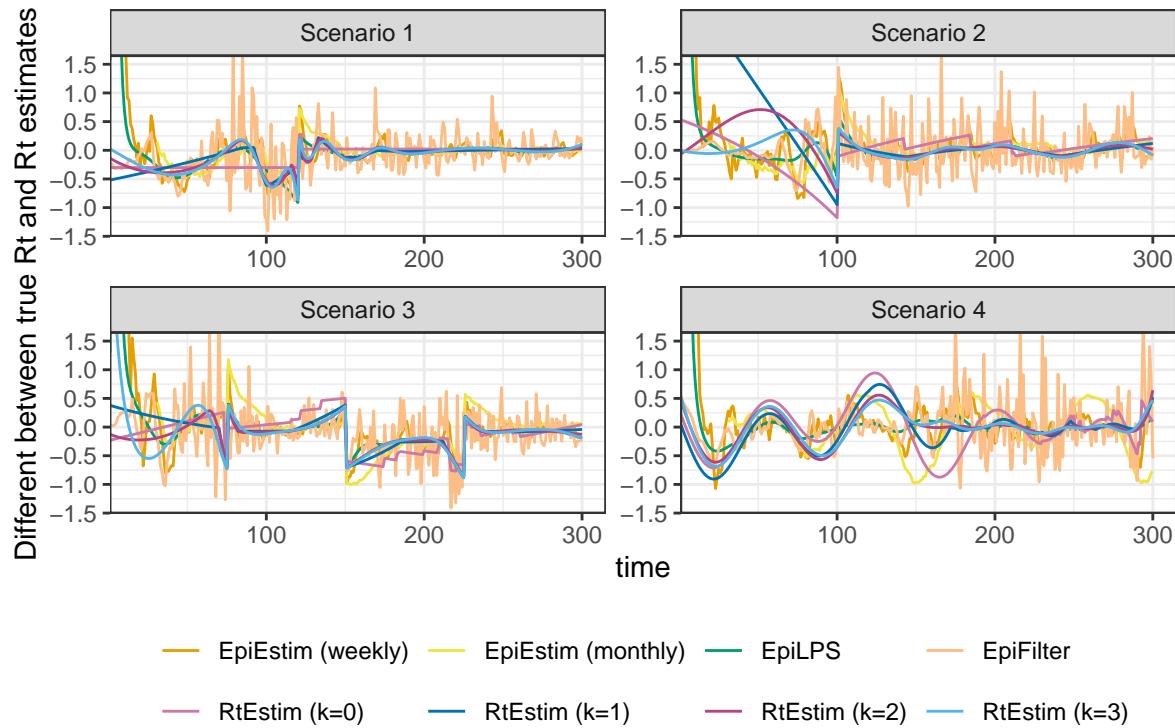


Figure A.7.4: Difference between of the true effective reproduction number and its estimation for SARS epidemics with negative Binomial observations.

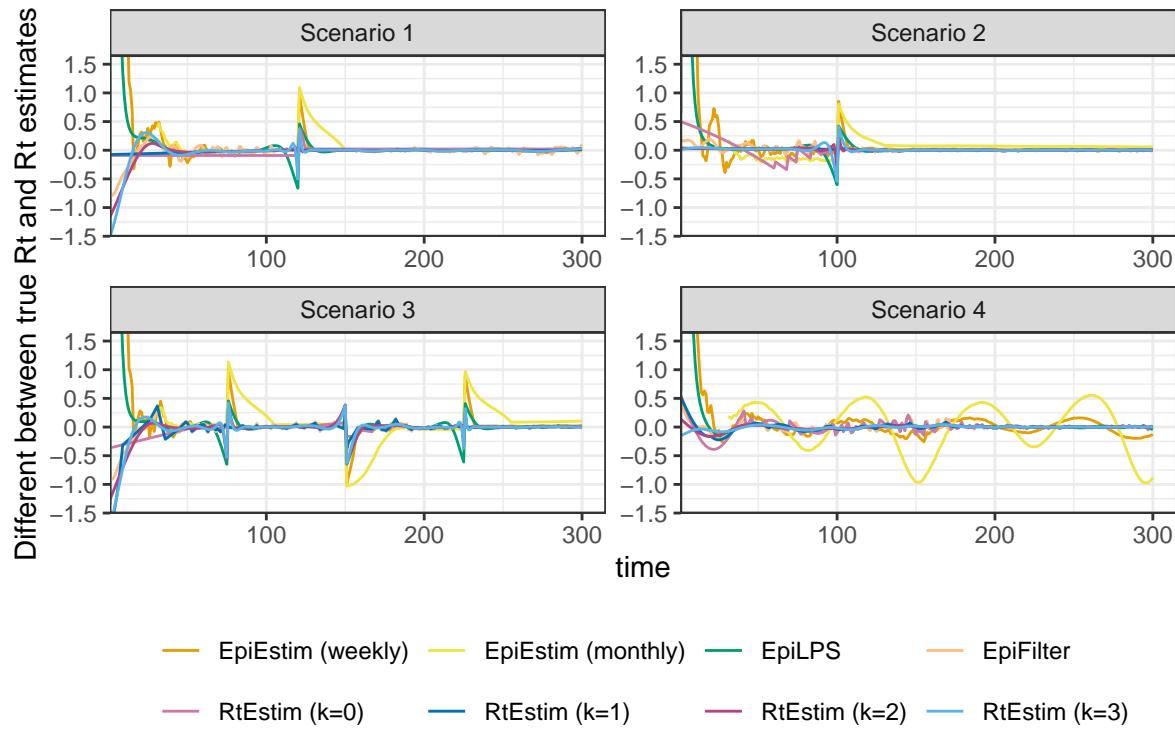


Figure A.7.5: Difference between of the true effective reproduction number and its estimation for SARS epidemics with Poisson observations.

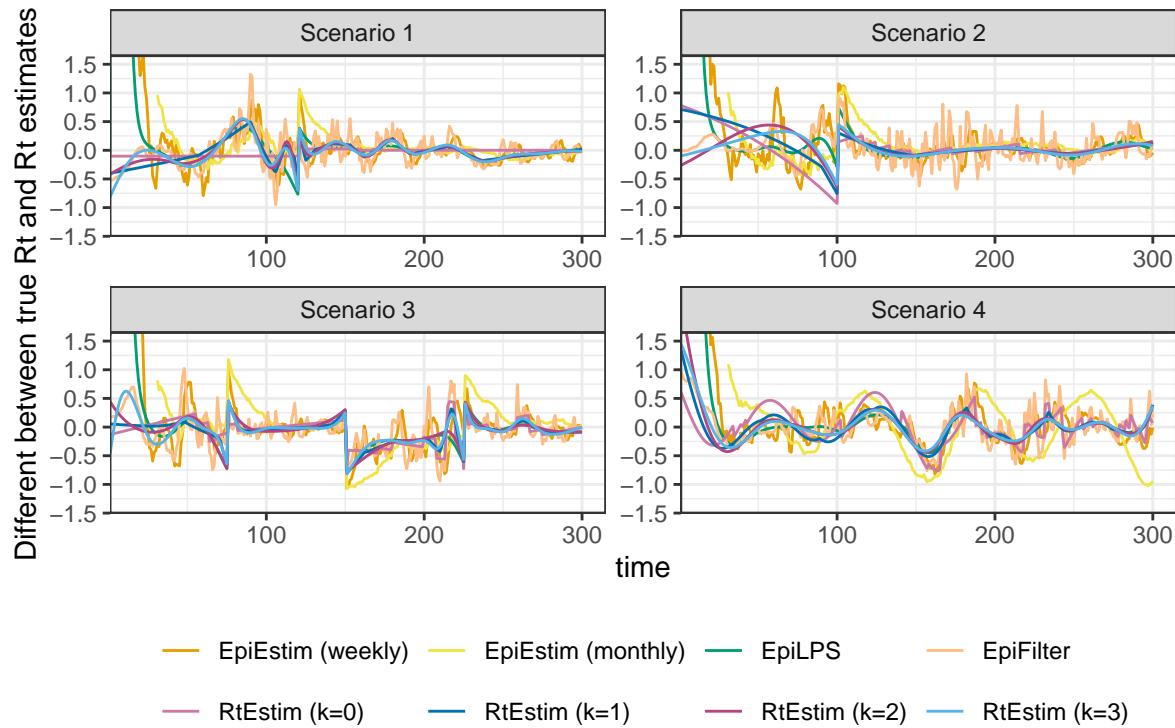


Figure A.7.6: Difference between of the true effective reproduction number and its estimation for measles epidemics with negative Binomial observations.

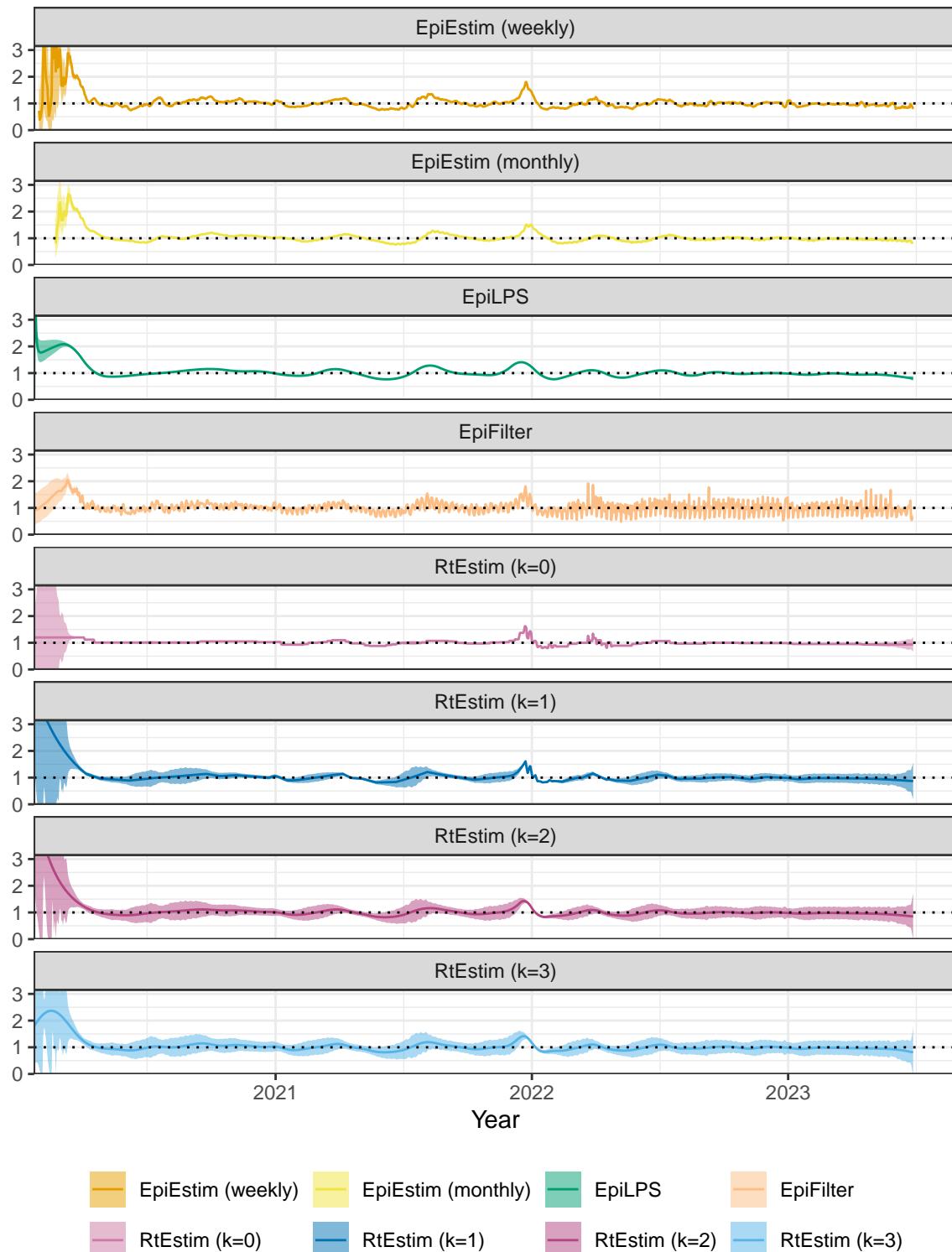


Figure A.8.1: Rt estimates with CIs for Covid19.

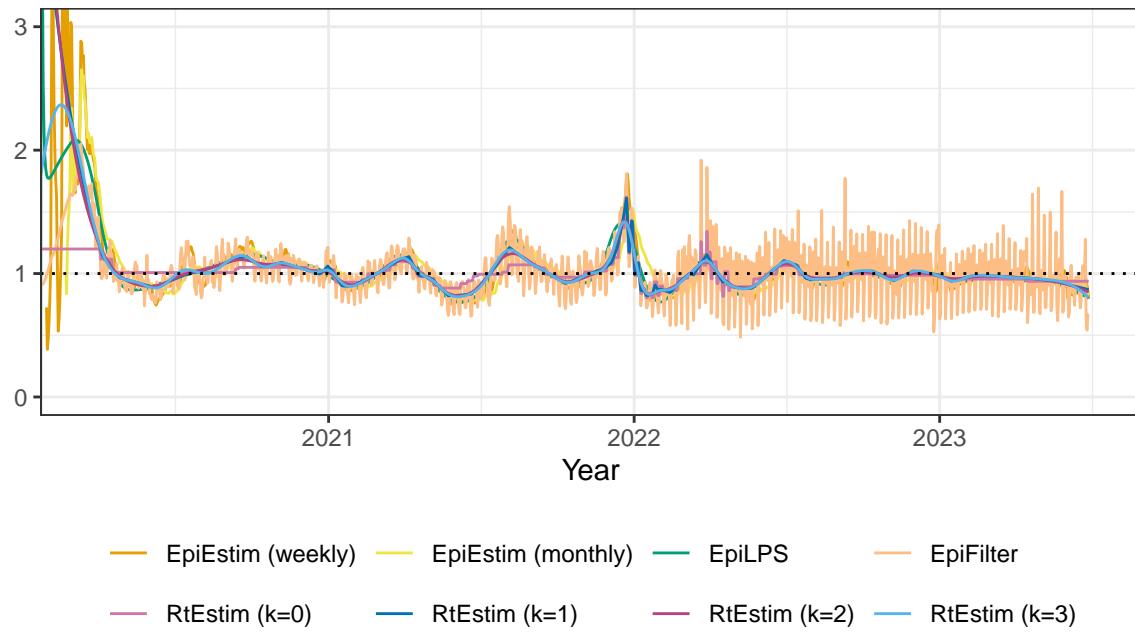
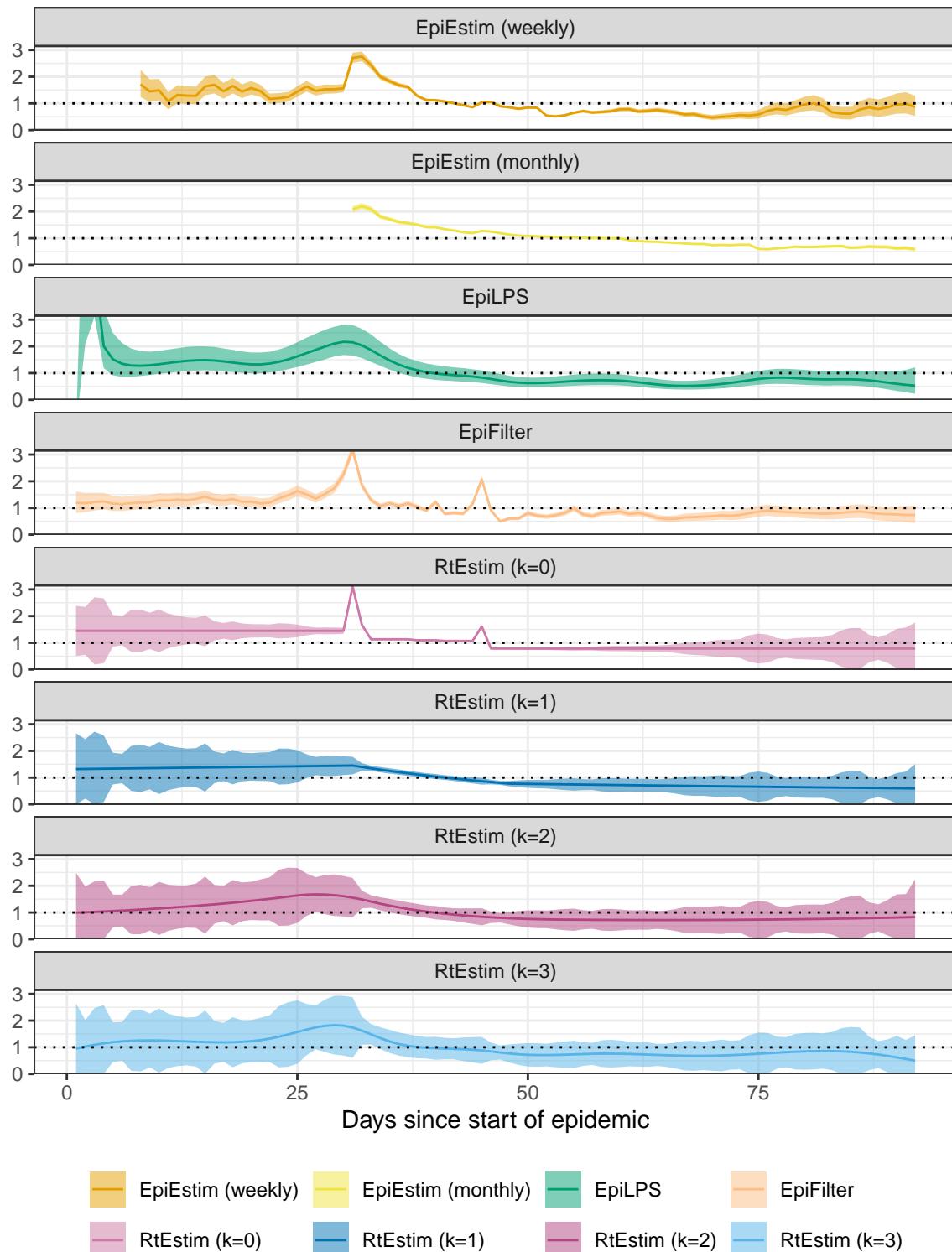


Figure A.8.2: Rt estimates for Covid19.


Figure A.8.3:  $R_t$  estimates with CIs for Flu 1918.

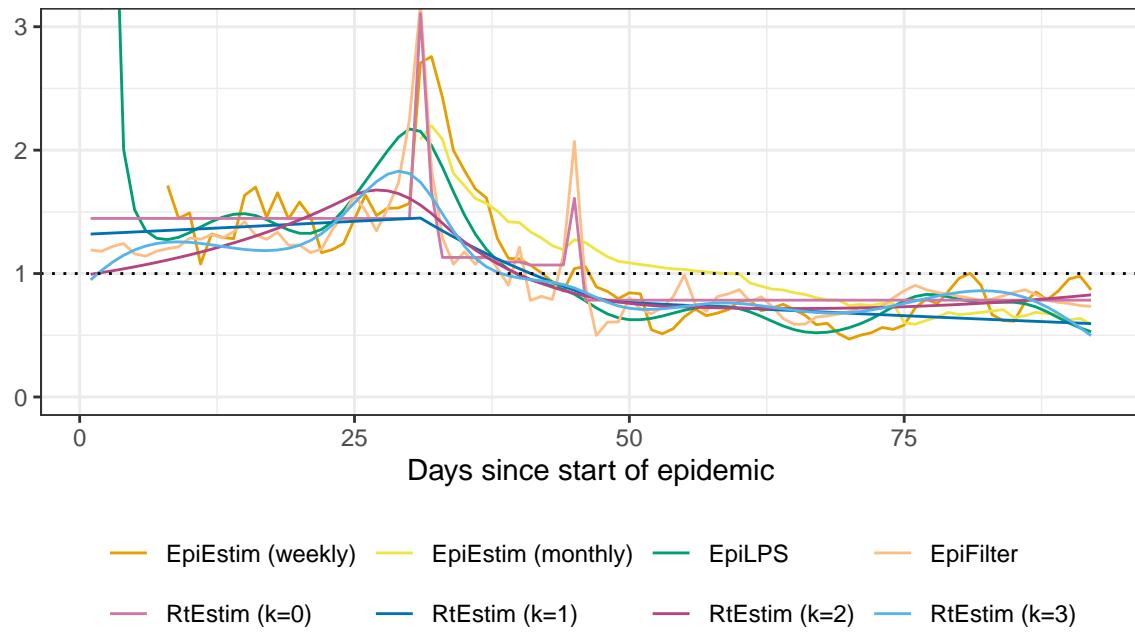


Figure A.8.4: Rt estimates for Flu 1918.