Supplementary details of time-varying reproduction number estimation with trend filtering

Jiaping 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) in the manuscript that is used to compare the accuracy of the estimated time-varying instantaneous reproduction number with the true ones. Given the total infectiousness η, we compare the distance between the Poisson distributions $y \sim Pois(\eta\hat{\mathcal{R}})$ and $y \sim Pois(\eta\mathcal{R})$, where $y, \mathcal{R} \in \mathbb{N}_0^n$ are natural numbers including 0, $\eta \in \mathbb{R}^n$, and $f_0(y;\eta,\mathcal{R}) = \prod_{t=1}^n \frac{(\eta_t\mathcal{R}_t)^{y_t}e^{-\eta_t\mathcal{R}_t}}{y_t!}$, $f_1(y;\eta,\hat{\mathcal{R}}) = \prod_{t=1}^n \frac{(\eta_t\hat{\mathcal{R}}_t)^{y_t}e^{-\eta_t\hat{\mathcal{R}}_t}}{y_t!}$ are the corresponding density mass functions for independent $y_t, t = 1, \cdots, n$. Then, the KL divergence between them is defined as

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

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

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 Supplmentary details on experimental settings

We compare the accuracy of the estimated instantaneous reproduction numbers using the mean Kullback Leibler (KL) divergence with Poisson distributional assumption on incidence (we say (mean) KL divergence for short in the following) in (11) 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` takes too long to converge (e.g., for a long `measles` epidemic, it takes almost 2 hours (specifically, 115 minutes computed on Cedar cluster provided by Compute Canada)), we only compare it with other methods for short `flu` epidemics.

We consider the serial interval (SI) distributions of `measles` and `SARS` to generate long synthetic epidemics, and `flu` for short epidemics, inspired by Cori et al. (2013) which used SI from real epidemics to illustrate the performance of their method. The means and standard deviations of SI distributions are estimated by existing studies; 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 and 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,

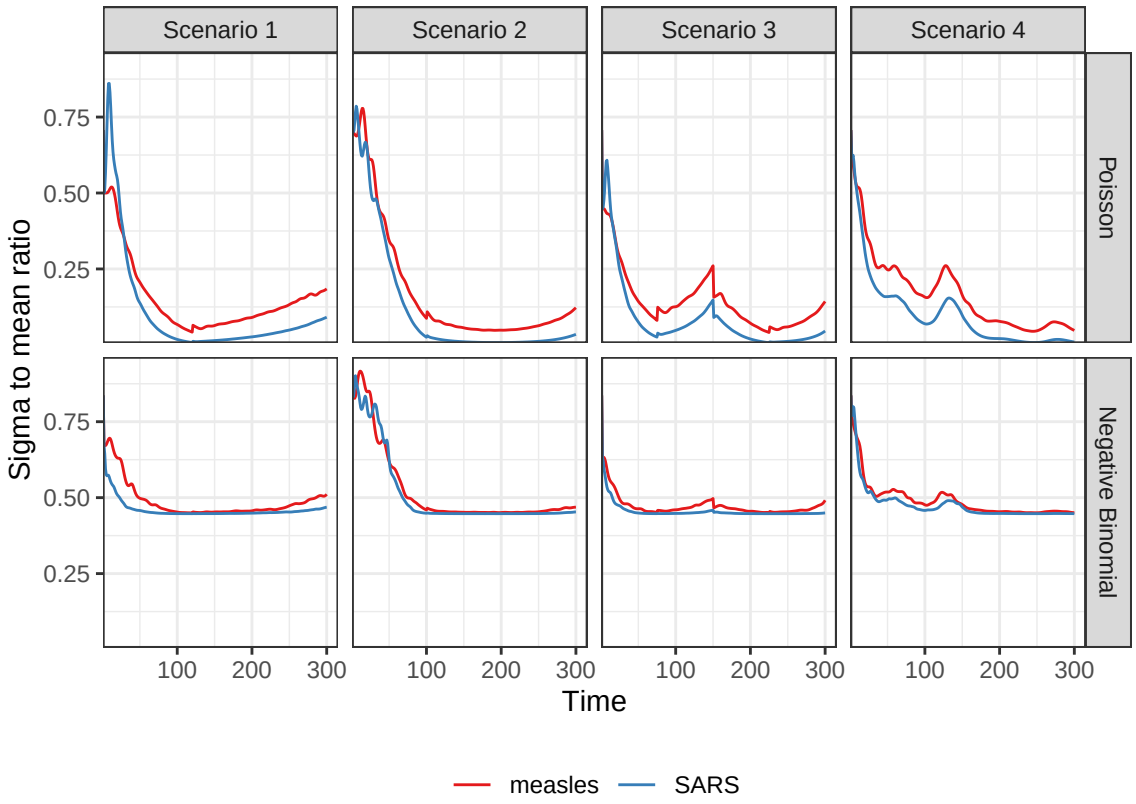

Figure A.2.1: Dispersion level of incidence of sample epidemics.

Table 1: Summary of experimental settings 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

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 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 four $\mathcal{R}_t$ scenarios are used for all long epidemics. We only consider one $\mathcal{R}_t$ scenario (Scenario 3: piecewise linear $\mathcal{R}_t$) 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 in the manuscript. Other main experimental results are displayed in Section A.3.

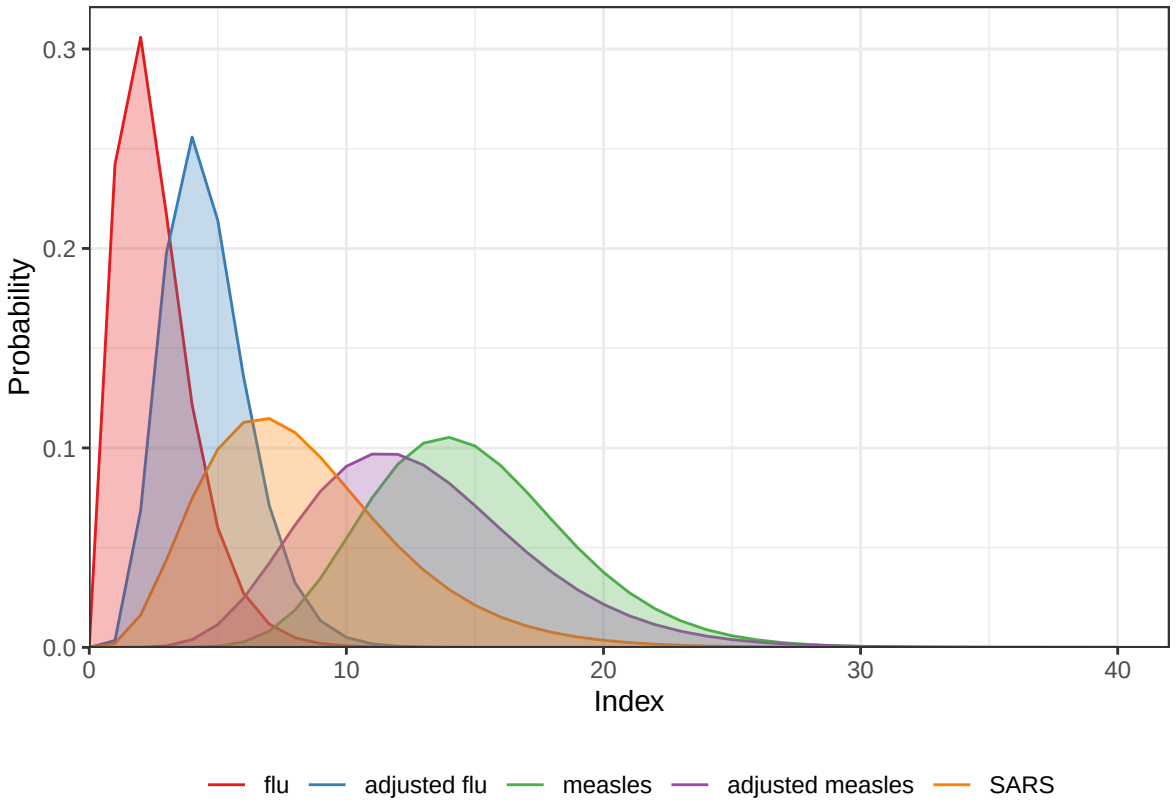

Figure A.2.2: Density curves of serial interval distributions used in the experiments.

A.3 Supplementary experimental results on accuracy comparison

A.3.1 Long epidemics

We have displayed the accuracy of all methods (where EpiEstim uses weekly sliding window) for measles and SARS sample epidemics using KL divergence excluding the first weeks since EpiEstim does not provide estimates in the first weeks in Fig 3 and Fig 4 in the manuscript, where we exclude the outliers. A full visualization including the outliers is in Figure 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.

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. It can be explained that `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.

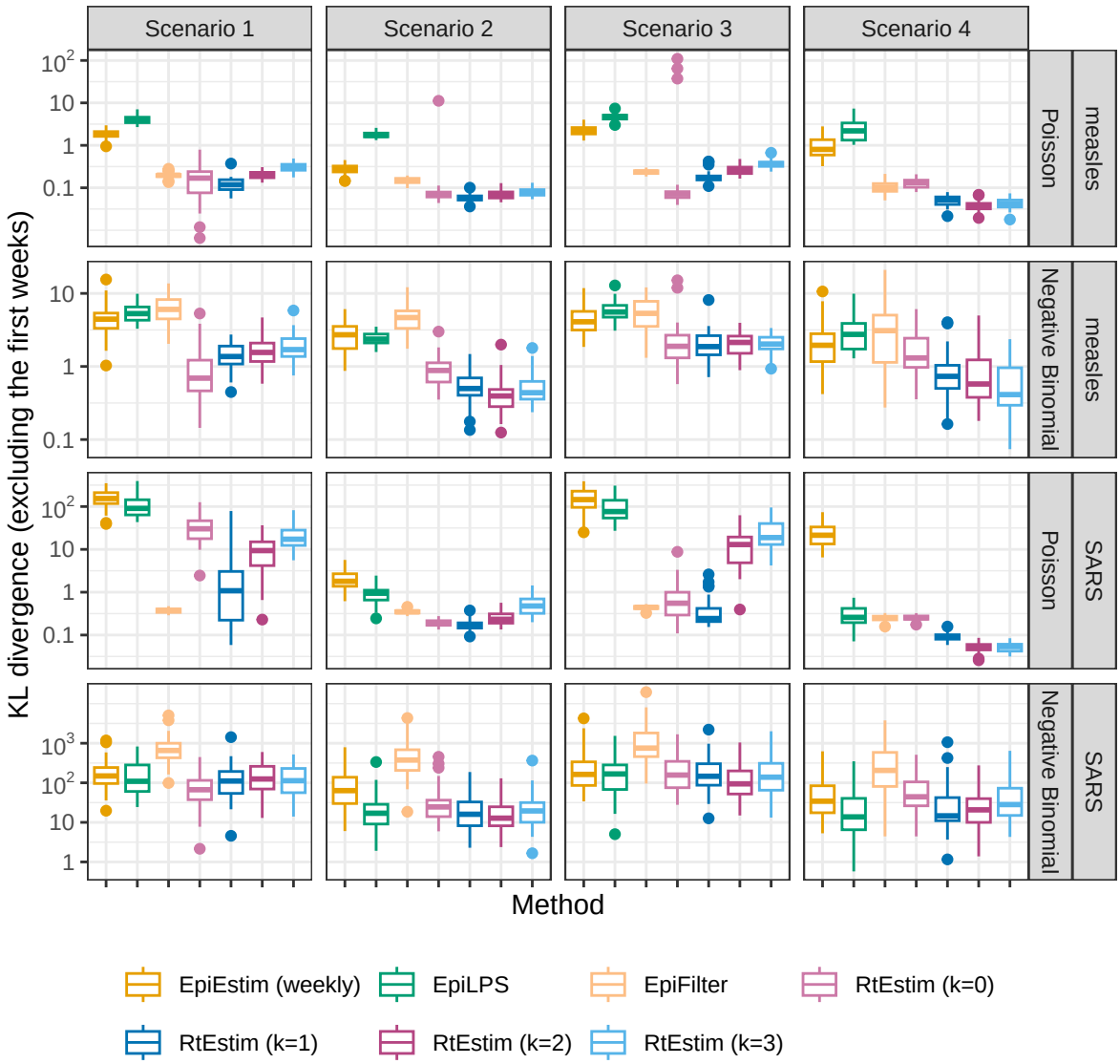

Figure A.3.1: The mean KL divergence excluding the first weeks for measles and SARS epidemics, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Y-axis is on a logarithmic scale.

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 to compare EpiEstim with weekly and monthly sliding windows with other methods including `EpiNow2`. The difference in accuracy is more obvious given Poisson distributional assumption in incidence. To estimate "true" piecewise linear $\mathcal{R}_t$, piecewise constant and linear `RtEstim` (with $k = 0, 1$) are most accurate given Poisson incidence, `RtEstim` ($k = 2, 3$), `EpiLPS` and `EpiFilter` are accurate as well with median KL estimates around 1. Given negative binomial incidence, the advantage of `RtEstim` is less obvious, but `RtEstim` with all degrees still have the lowest median with a short IQR.

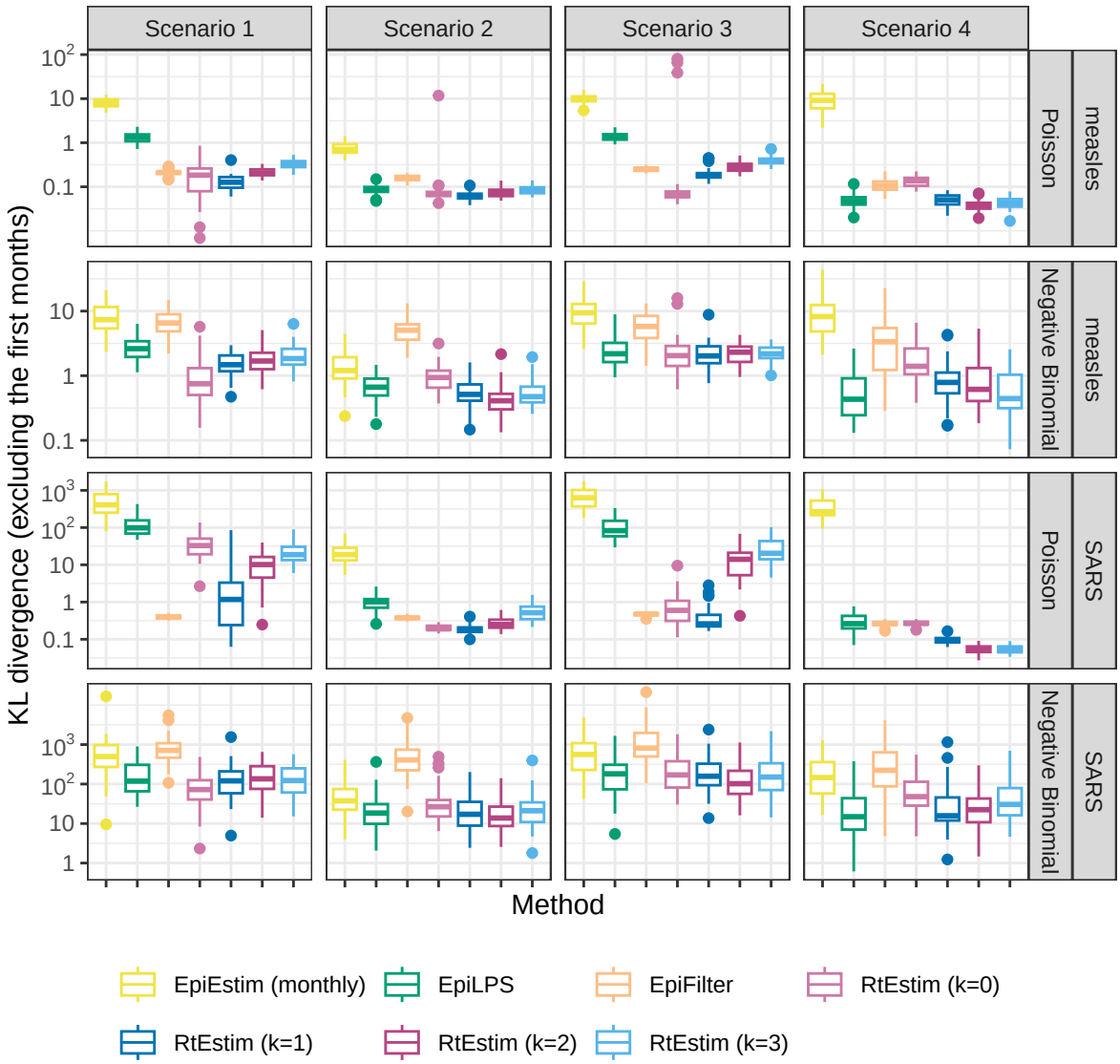

Figure A.3.2: The mean KL divergence excluding the first months for measles and SARS epidemics, since EpiEstim with the monthly sliding window does not provide estimates for the first month. Y-axis is on a logarithmic scale.

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 the first weeks and the first months respectively) for all 8 methods with "mild" misspecification (using shaped and scaled `measles` SI parameters) and "major" misspecification (using `SARS` SI parameters) for long `measles` epidemics across all settings. `RtEstim` performs robust to misspecification of SI parameters, since the median for each problem design is almost always the lowest with the lowest IQR. `EpiLPS` is a strong competitor given negative binomial incidence, since it assumes incidence to follow negative binomial distributions. `EpiFilter` is also quite robust to SI misspecification

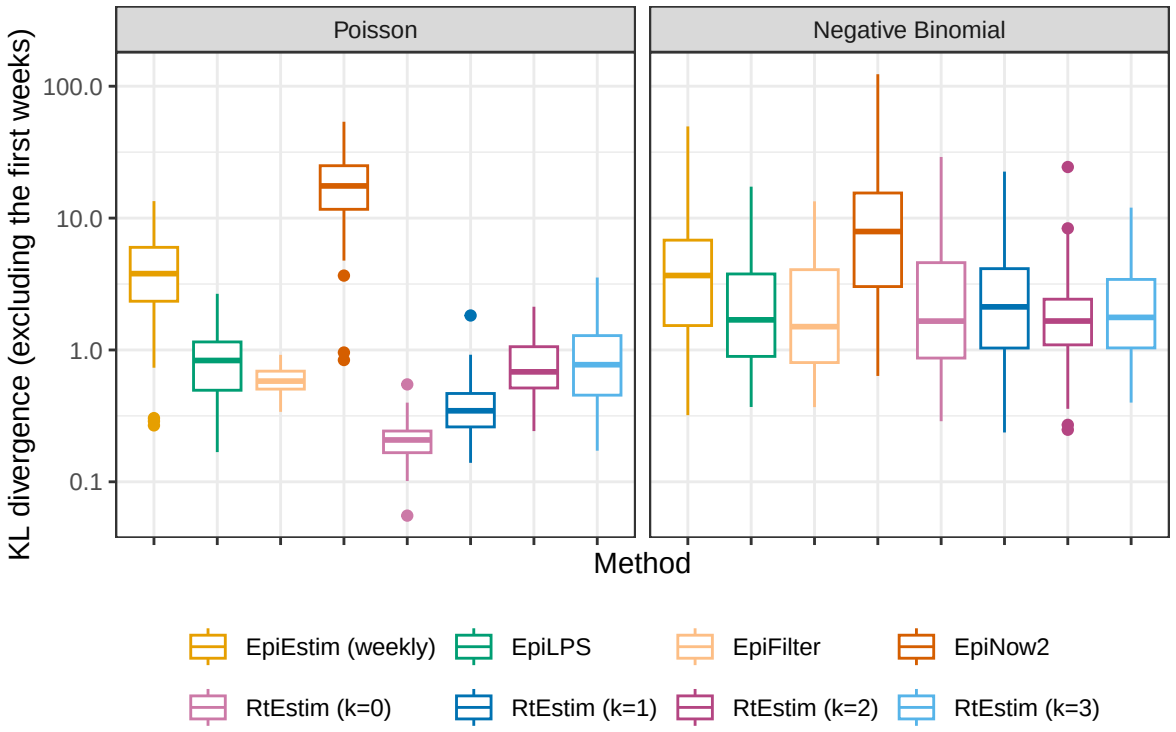

Figure A.3.3: The mean KL divergence excluding the first weeks for flu epidemics, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Y-axis is on a logarithmic scale.

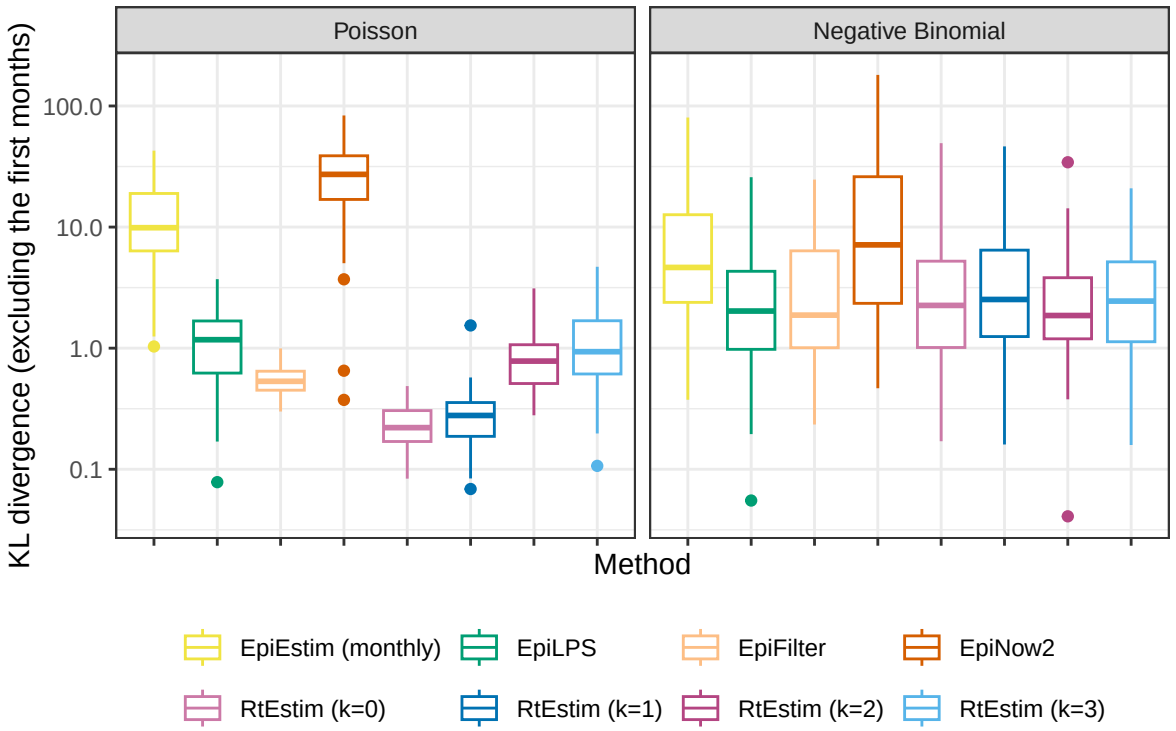

Figure A.3.4: The mean KL divergence excluding the first months for flu epidemics, since EpiEstim with the monthly sliding window does not provide estimates for the first month. Y-axis is on a logarithmic scale.

given Poisson incidence.

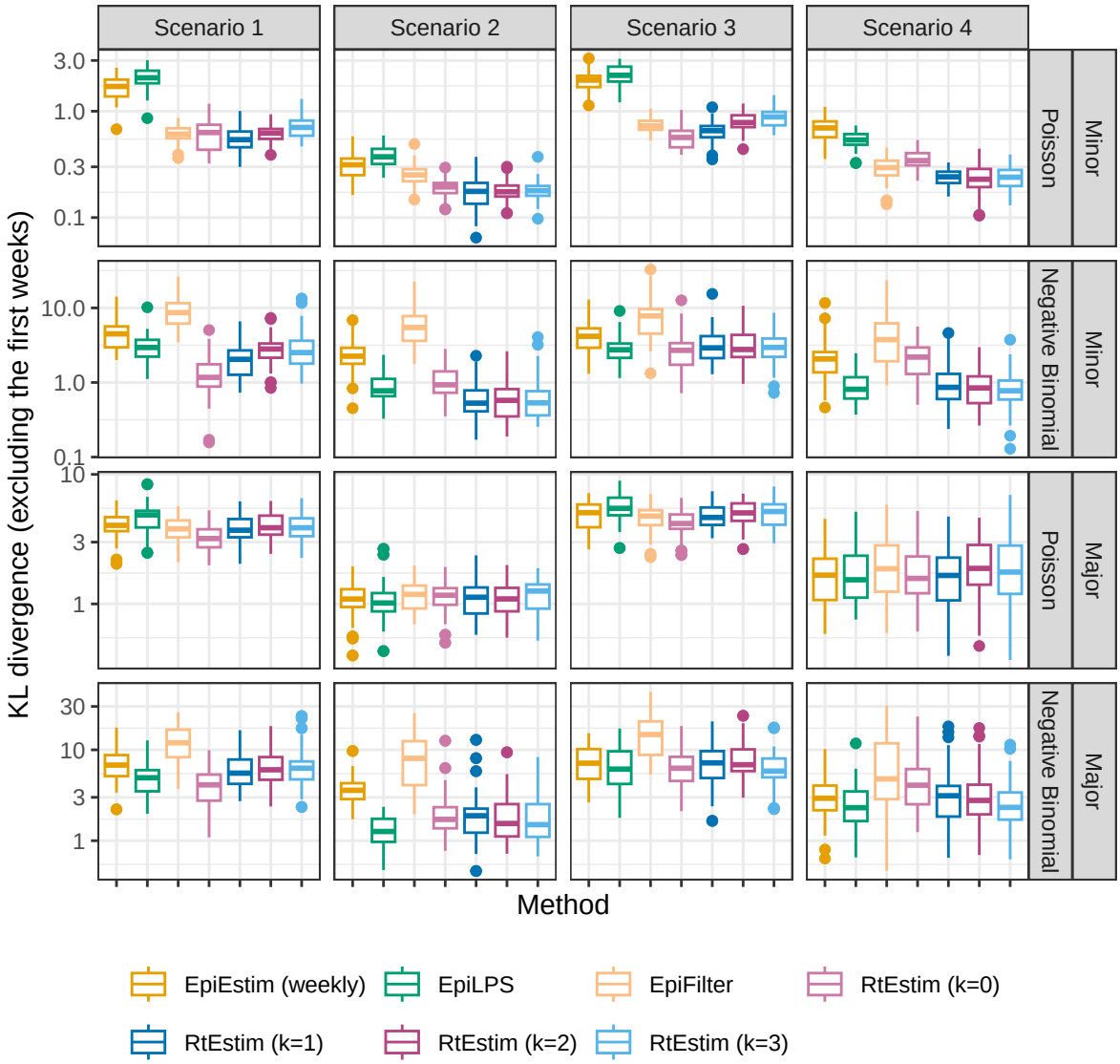

Figure A.4.1: The mean KL divergence excluding the first weeks for measles epidemics with SI misspecification, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Y-axis is on a logarithmic scale.

A.4.2 SI misspecification for short epidemics

Figures A.4.3 and A.4.4 display KL divergence (excluding the first weeks and the first months respectively) for all 9 methods with "minor" misspecification (using shaped and scaled `flu` SI parameters) and "major" misspecification (using `measles` parameters) for short `flu` epidemics across all settings. We yield similar conclusion in short epidemics. We also note that `EpiNow2` is quite robust to major misspecification in SI parameters, while `EpiLPS` is less satisfactory in major misspecification excluding the first weeks in KL computation. It might be due to the large estimates at the beginning of the epidemics beyond the first weeks, but eliminated within the first months.

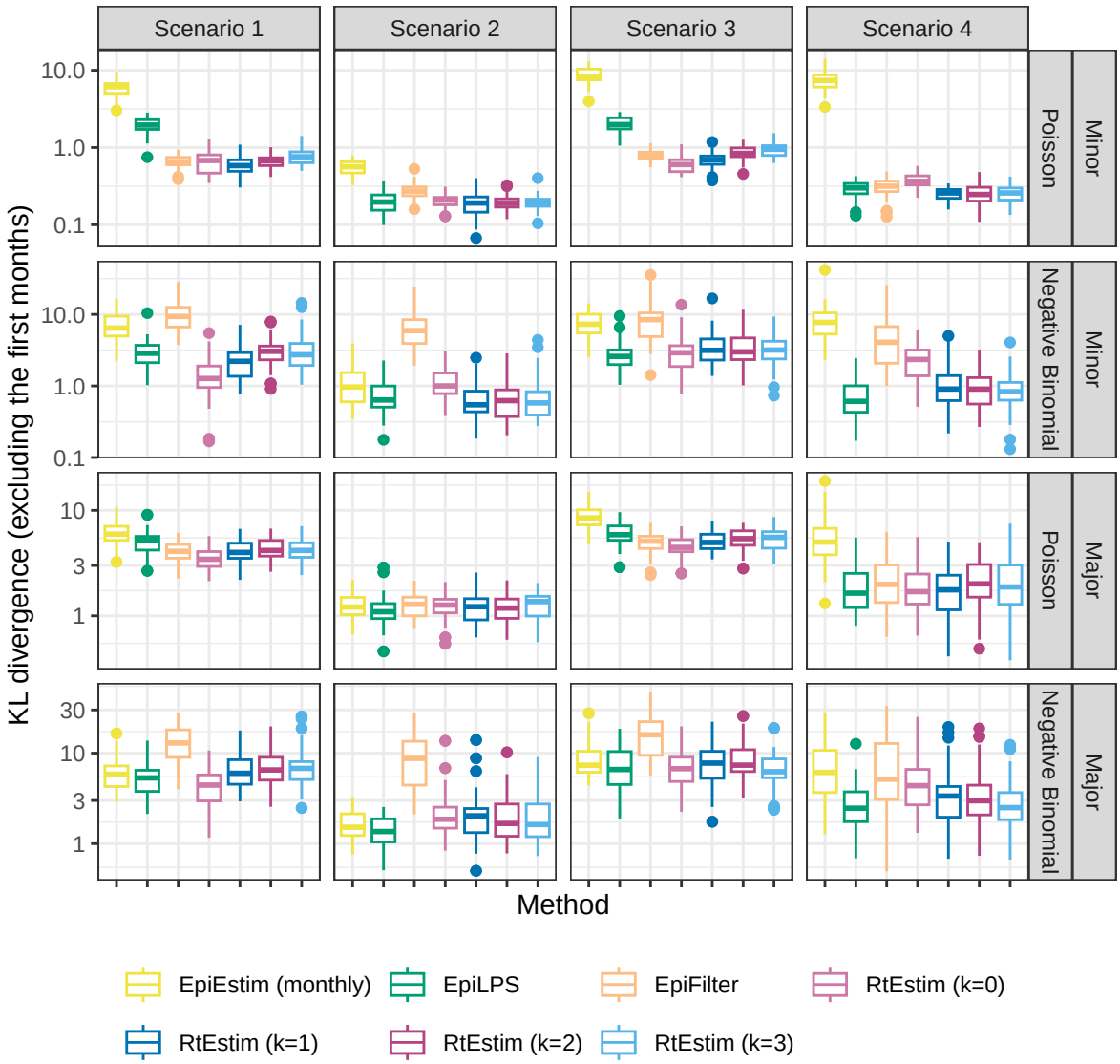

Figure A.4.2: The mean KL divergence excluding the first months for measles epidemics with SI misspecification, since EpiEstim with the monthly sliding window does not provide estimates for the first month. Y-axis is on a logarithmic scale.

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 sliding windows are very fast and converge in less than 0.1 seconds. Piecewise 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. `EpiFilter` is in a similar scale of our method with higher than 0 degrees. Piecewise linear and cubic `RtEstim` (with $k = 1$ and $k = 3$ respectively) are slower, but mostly within 10 seconds. We also provide an alternative view with the running time of each case in a separate panel in Figures A.5.2 and A.5.3 for `measles` and `SARS` epidemics respectively. We find similar results as in Figure A.5.1 in all panels.

It is remarkable that our `RtEstim` computes 50 lambda values with 10-fold CV for each experiment, which

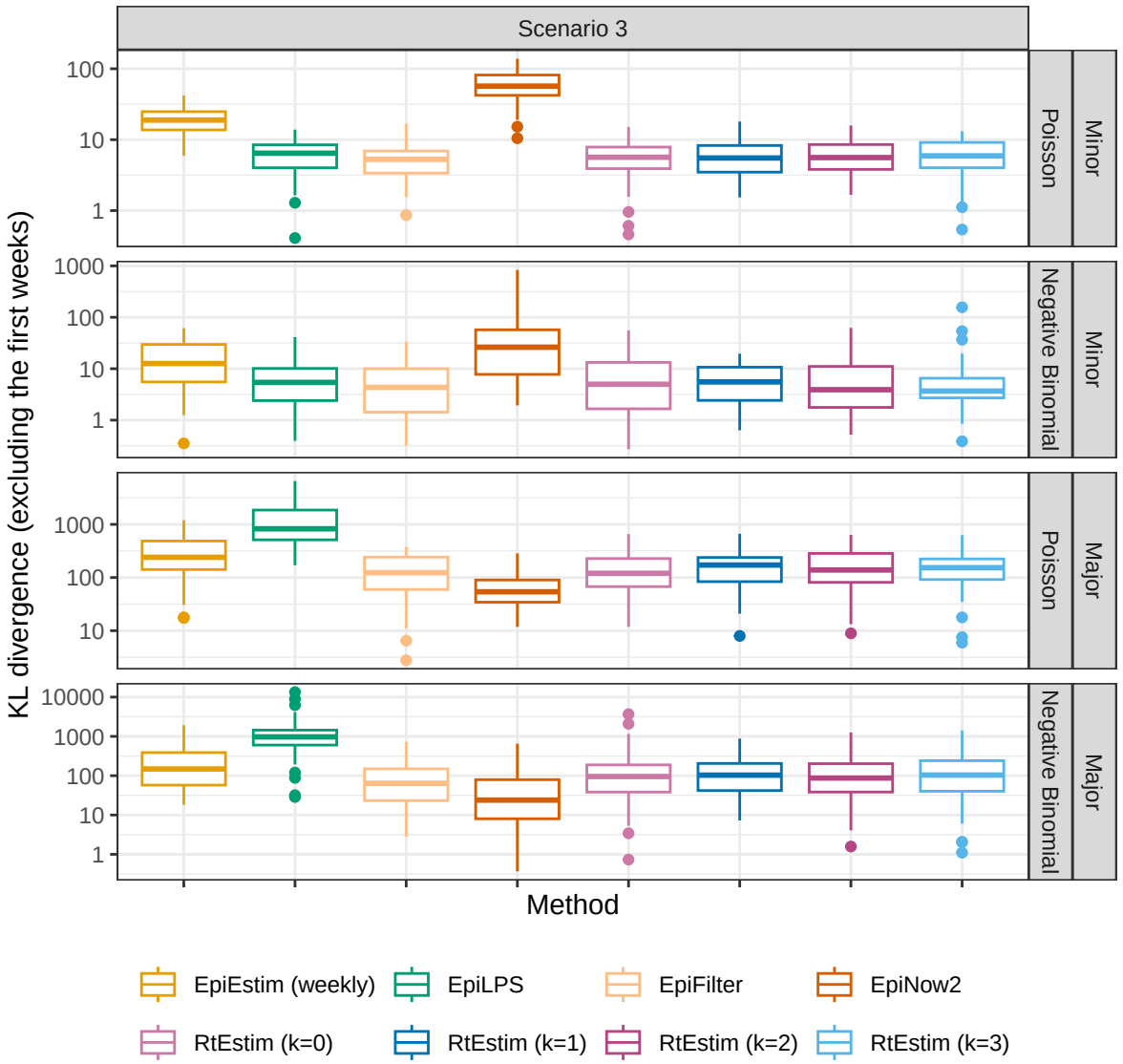

Figure A.4.3: The mean KL divergence excluding the first weeks for flu epidemics with SI misspecification, since EpiEstim with the weekly sliding window does not provide estimates for the first week. Y-axis is on a logarithmic scale.

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.

Figure A.5.4 displays the running time of all methods for short (`flu`) epidemics. All methods except `EpiNow2` can converge with in around 1 second (within 10 seconds). Figure A.5.5 displays the running times for each setting separately, and finds similar results as in the overall running time comparison.

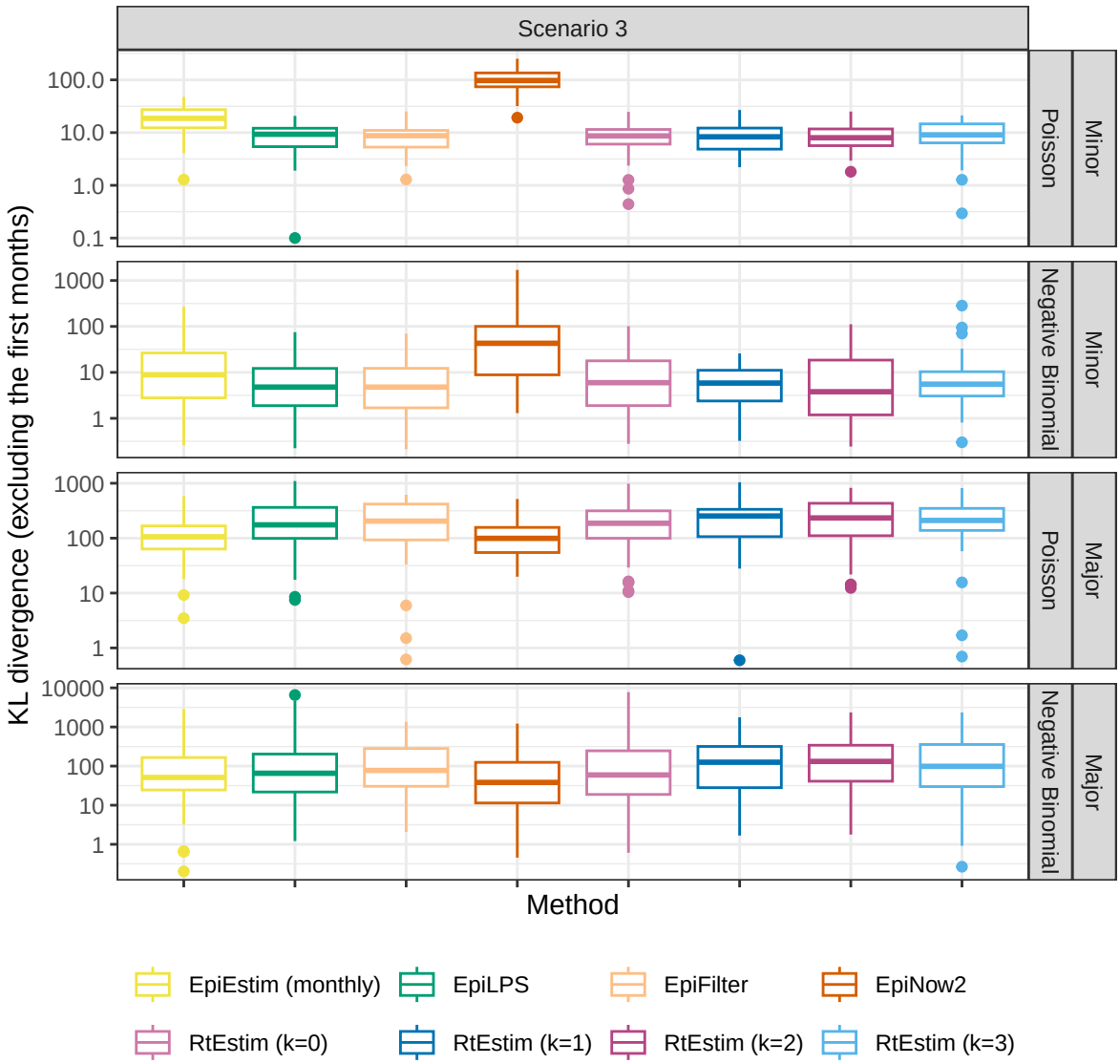

Figure A.4.4: The mean KL divergence excluding the first months for flu epidemics with SI misspecification, since EpiEstim with the monthly sliding window does not provide estimates for the first month. Y-axis is on a logarithmic scale.

A.6 Confidence interval coverage

A.6.1 Display estimates and confidence intervals for sample epidemics

Fig 5 and Fig 6 in the manuscript provided $\mathcal{R}_t$ estimates by all methods on sample `measles` epidemics with Poisson incidence and `SARS` epidemics with negative binomial incidence respectively. Figures A.6.1 and A.6.4 provide a clearer view of each method with its 95% confidence interval in a separate panel. The full display of sample epidemics for other settings are visualized in Figures A.6.2 and A.6.3.

All methods (except EpiEstim with the monthly sliding window) fit the epidemics with Poisson incidence well with estimate $\hat{\mathcal{R}}_t$ close to the true $\mathcal{R}_t$ and 95% CI covering the true value at most timepoints. While, given negative binomial incidence, `RtEstim` with k=0 misses to recover the curvature in $\mathcal{R}_t$ curves, especially in the

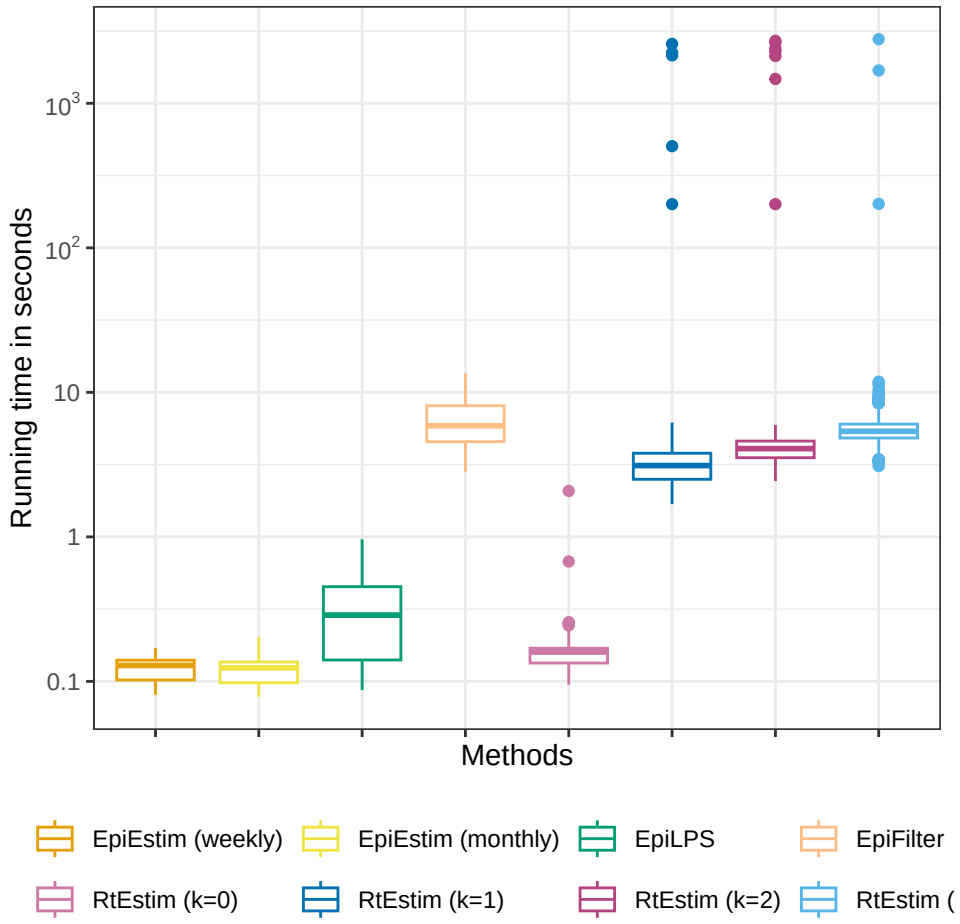

Figure A.5.1: Running time comparison of all methods for long (measles and SARS) epidemics across all cases. Y-axis is on a logarithmic scale.

exponential and periodic scenarios. `EpiEstim` with weekly sliding windows and `EpiFilter` are more wiggly, and `EpiLPS` has wider confidence intervals given negative binomial incidence compared to Poisson incidence. For large incidence with negative binomial distribution, `EpiFilter` is extremely wiggly and it is difficult for `RtEstim` (k=0) to recover many changepoints and the curvature especially in exponential and periodic scenarios. `EpiLPS` performs well overall, but returns large estimates at the beginning of the epidemics, which remains after the first week. Overall, our method with different degrees can recover the changepoints and graphical curvature of $\mathcal{R}_t$ in all scenarios, except in the case of the periodic $\mathcal{R}_t$ curve with large incidence from negative binomial distribution, where `EpiLPS` has a clear win ignoring the large estimates at the early stage. The performance of sample curves across different settings by different methods generally coincides with the findings in the KL divergence estimates.

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, which are used to generate the synthetic epidemics, in modelling. Table 2 summarizes the experimental settings.

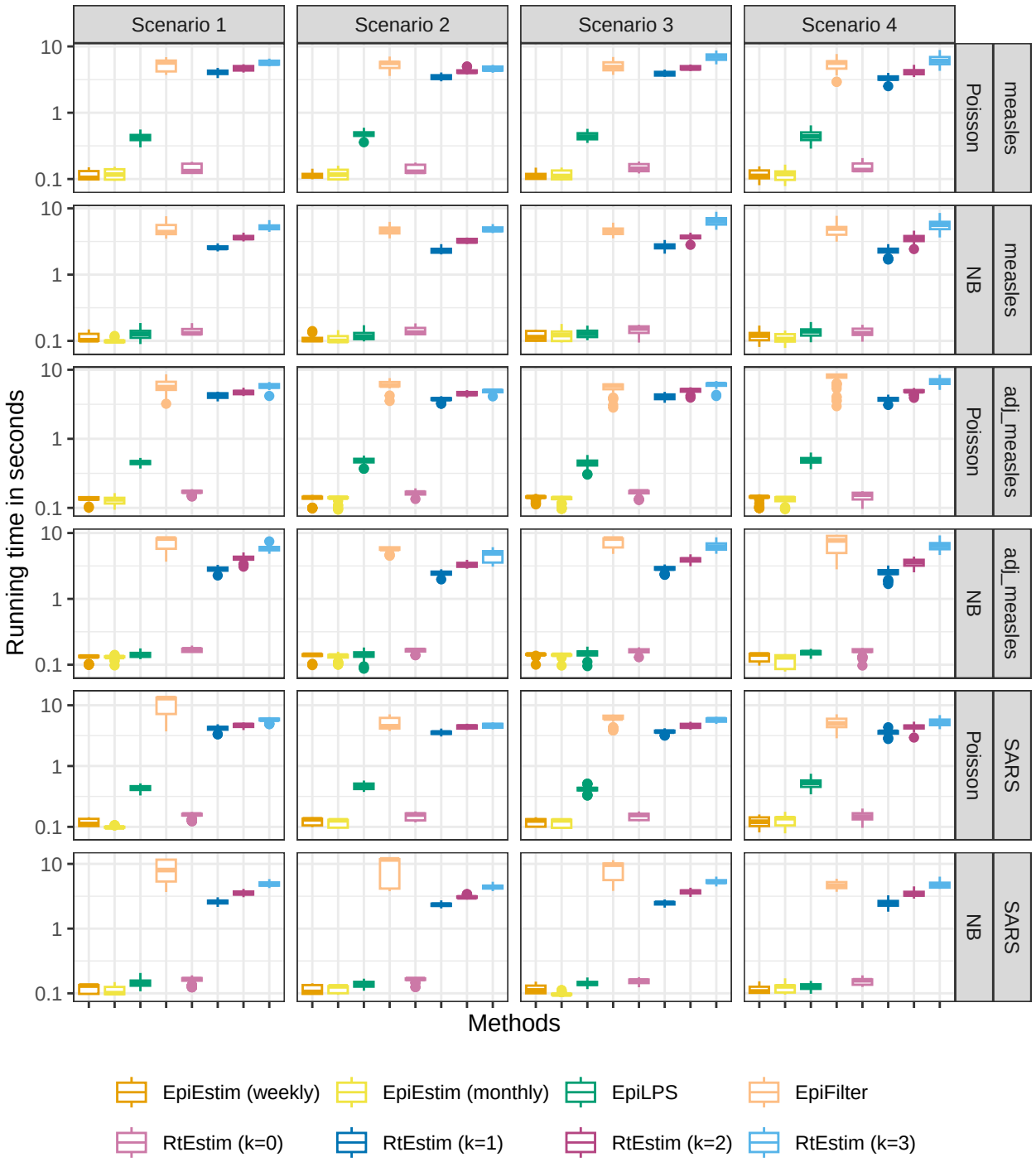

Figure A.5.2: Running time comparison of all methods for measles epidemics with each pair of SI parameters (measles, adjusted measles, and SARS) for modelling per incidence distribution per Rt scenario (excluding outliers for better illustration). Y-axes are on a logarithmic scale.

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. Low Poisson incidence is the easiest for all methods,

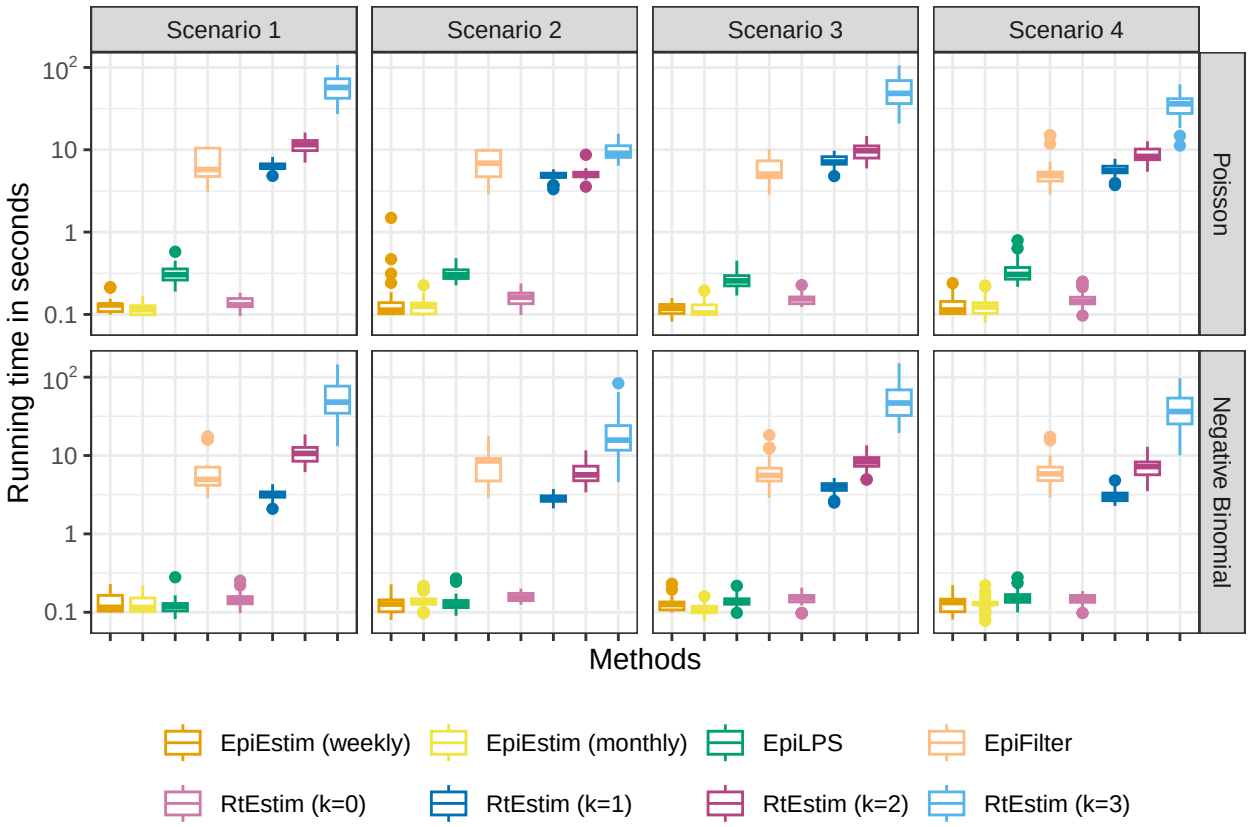

Figure A.5.3: Running time comparison of all methods for SARS epidemics with each choice of SI parameter for modelling per incidence distribution per Rt scenario. Y-axes are on a logarithmic scale.

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	8 methods
300	SARS	3	Poisson, NB	SARS	8 methods

with coverage near 100% at most timepoints and 0 elsewhere. Large negative Binomial incidence is the "hardest", while `EpiLPS` does the best here with avearged coverage at all timepoints close to 1. It is consistent to the findings in the accuracy comparison (using KL values) and the illustration of sample epidemics in Figures A.6.1–A.6.4, where `EpiLPS` is the most accurate in this case. `RtEstim` with degrees k=1,2,3 have 100% coverage at most timepoints except the changepoints except in the "hardest" case, where larger degrees tend to have higher percentages of coverge at most timepoints. `RtEstim` with $k = 0$ tends to produce overly narrow intervals, leading to lower coverage. `EpiEstim` with weekly sliding windows has a higher chance to fail to cover the true $\mathcal{R}_t$ give negative binomial incidence compared to Poisson and given larger incidence, its point estimates are quite accurate, but since its 95% confidence band is too narrow and the estimate curves are quite wiggly, it fails to cover the true values. `EpiEstim` with monthly sliding windows has low percentages of interval coverage at more timepoints than other methods, especially given negative binomial incidence, which is also consistent to the findings in Section A.6.1, where it misses to recover the $\mathcal{R}_t$ values at many timepoints and also has narrow 95% confidence intervals. `EpiFilter` has lower percentages of coverage given negative binomial incidence than given Poisson incidence, which is consistent to its performance in

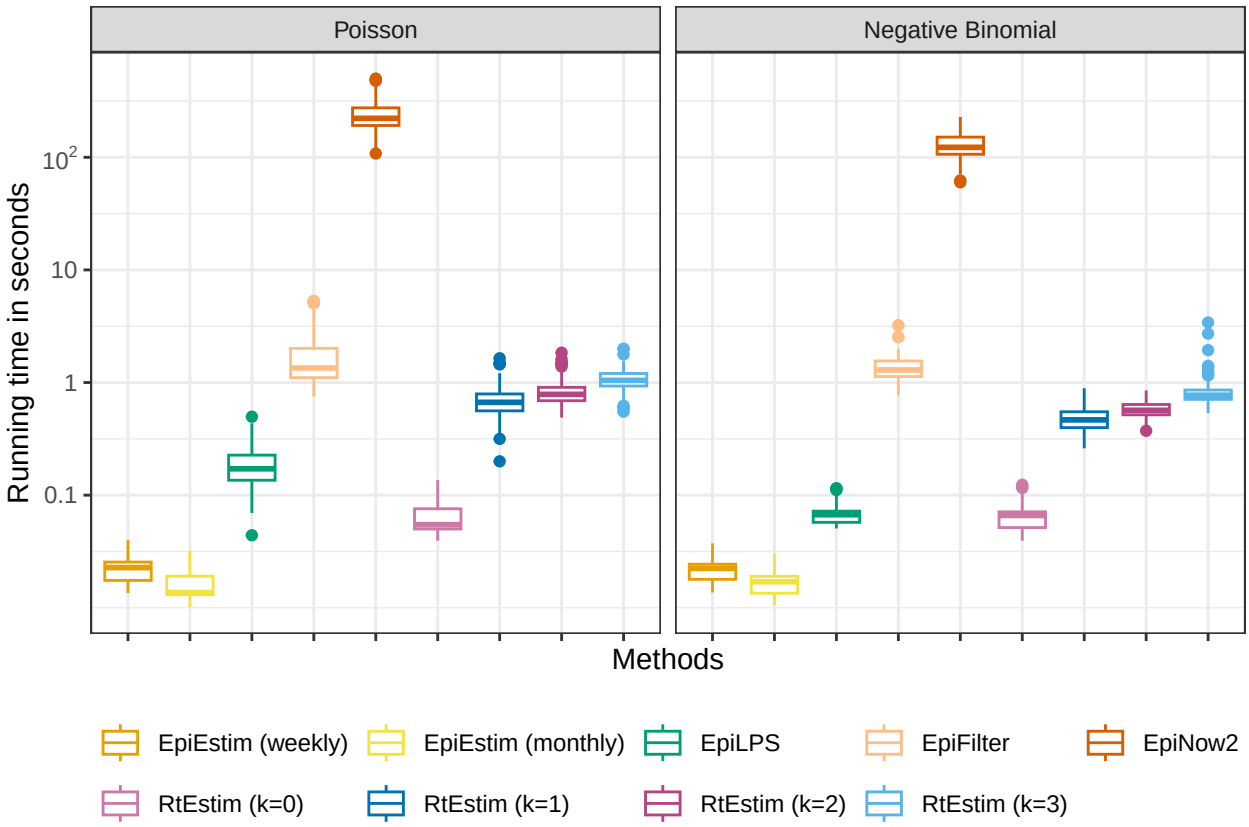

Figure A.5.4: Time comparisons of methods for short (flu) epidemics across all cases. Y-axis is on a logarithmic scale.

accuracy of point estimation.

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. CIs of `RtEstim` with $k = 1, 2, 3$ have nearly 100% coverage across all timepoints for all random samples except in the "hardest" problem, where the incidence is large and overdispersed. The coverage of `RtEstim` $k = 0$ is lower than for other degrees, similar to the above. `EpiFilter` has larger averaged percentages across all timepoints given Poisson incidence compared to negative binomial incidence. `EpiEstim` with weekly sliding windows has higher coverage compared to monthly windows, while their percentages of coverage across all timepoints are less than 95% in most cases. The percentage of coverage of `EpiLPS` is close to 95% in most cases, and even in the "hardest" problem, its coverage is roughly higher than 70%.

We also output the interval score (Bracher et al. 2021)

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

where $\alpha = 0.05$ is the significance level, l, u are the lower and upper bounds, $\mathcal{R}_t$ is the true instantaneous reproduction number, and $\mathbf{1}_X$ is the indicator function of the condition X. A confidence band that covers the true value more frequently with a shorter interval width will have a lower interval score. Figures A.6.9 and A.6.10 displays the interval scores of 95% CI averaged over 50 random measles and SARS epidemics respectively.

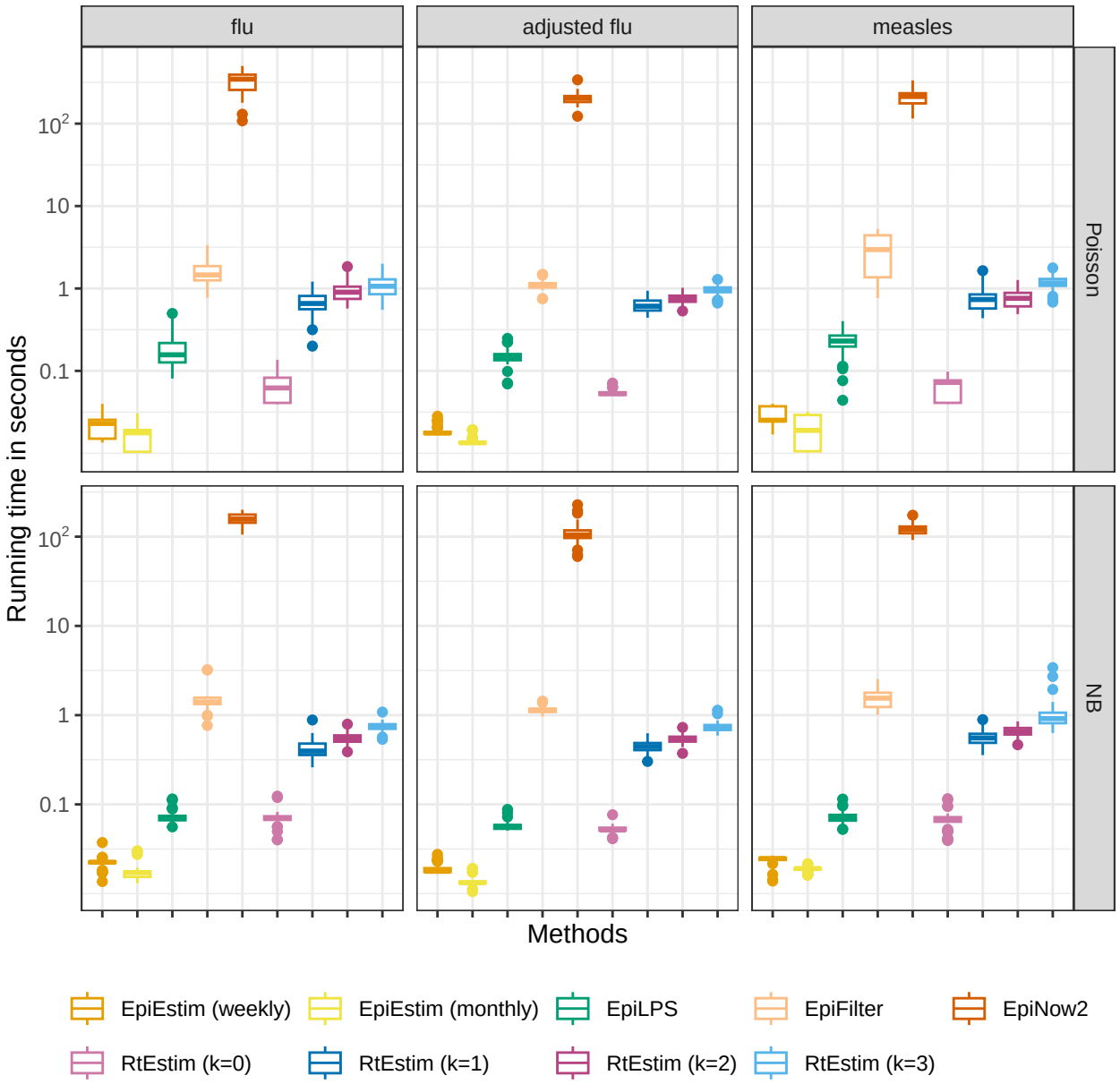

Figure A.5.5: Time comparisons of methods for short (flu) epidemics for pieciewise linear Rt (Scenario 3) for different pairs of SI parameters (flu, adjusted flu, and measles) and incidence distributions in different panels. Y-axes are on a logarithmic scale.

`RtEstim` always has the lowest or close to the lowest interval scores. In Poisson cases, `EpiFilter` has the lowest interval scores (less than 1), and the scores of `RtEstim` are always around 1.

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`

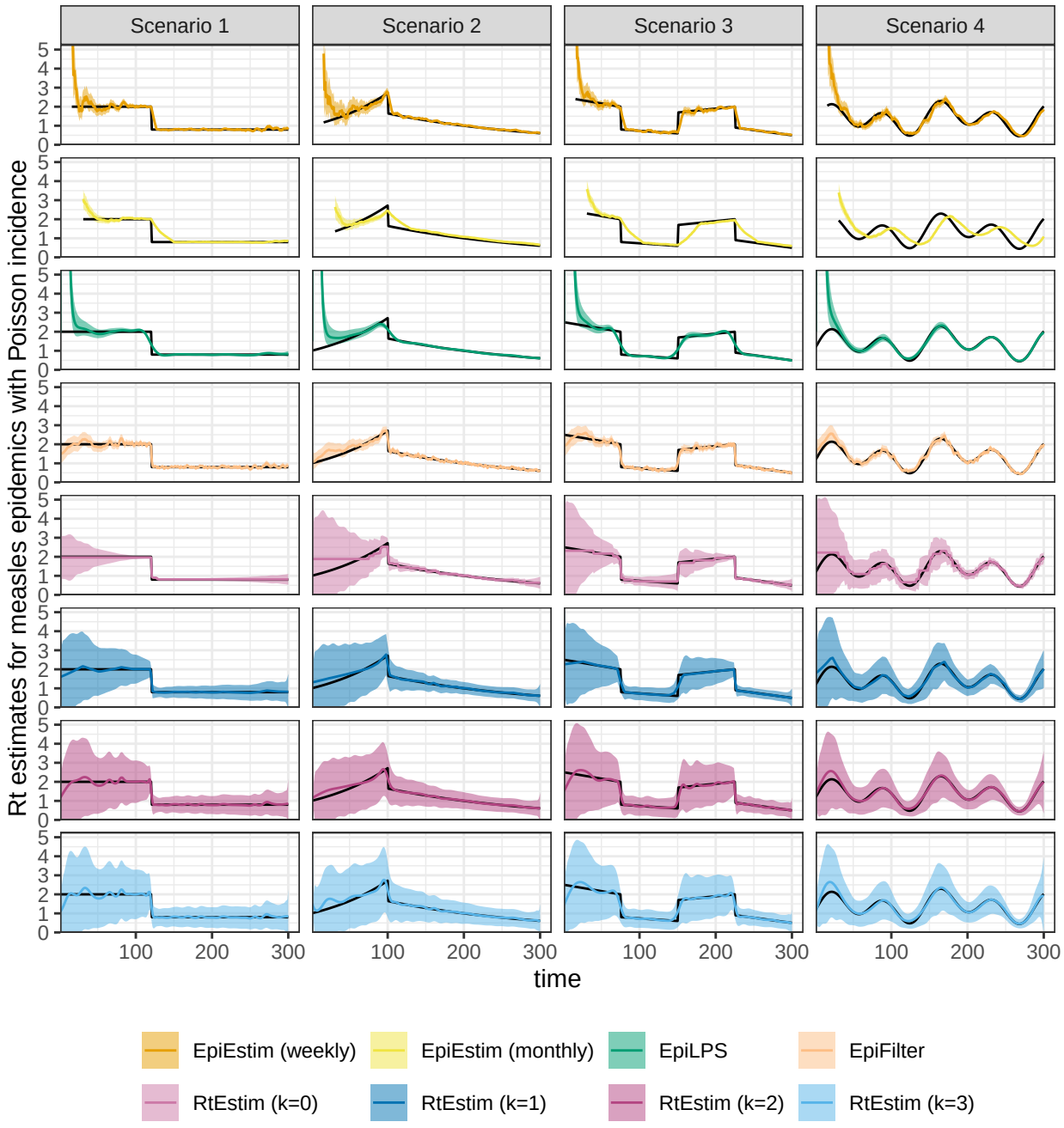

Figure A.6.1: Example measles epidemics with Poisson incidence.

with negative binomial incidence are provided in Fig 5 and Fig 6 in the manuscript. A full visualization of each case is provided in Section A.6.1. Here, we provide the condensed visualization of other cases in Figures A.7.1 and A.7.2. All methods provide accurate point estimates given large incidence from Poisson distribution, while `EpiEstim` (with weekly sliding window) and `EpiFilter` are more wiggly given negative binomial incidence.

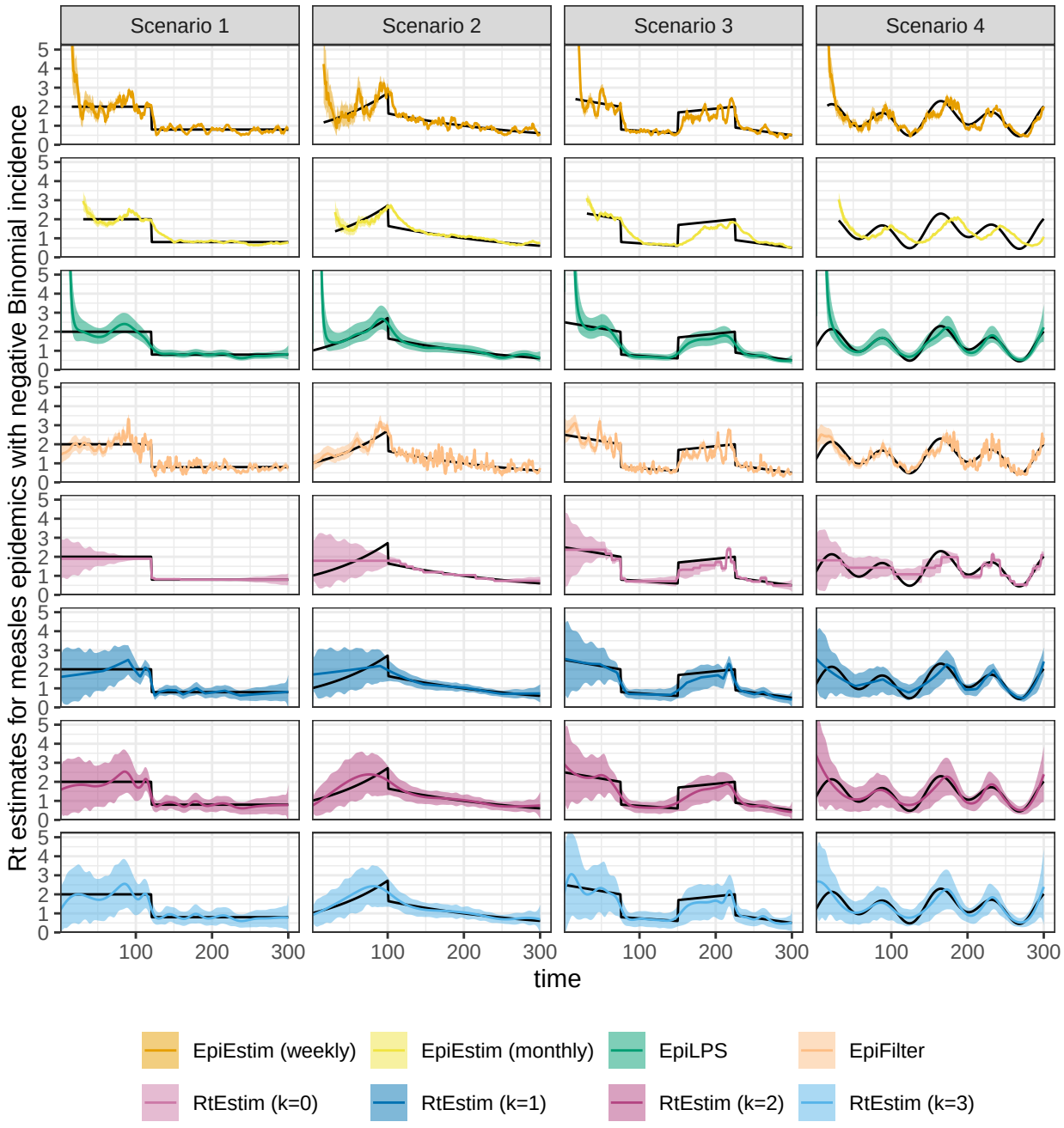

Figure A.6.2: Example measles epidemics with negative binomial incidence.

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

We also provide an alternative view of Fig 5 & Fig 6 in the manuscript by plotting $\mathcal{R}_t - \hat{\mathcal{R}}_t$ per coordinate t in Figures 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. We notice the different is larger at the changepoints for most methods. In the sinusoidal periodic scenario, the different of many method shows a periodic pattern, which implies that there is a periodic pattern that fails to be filtered by many methods. It makes sense since the "true" $calR_t$ curve is sinusoidal, while each method only filters the curve to a fixed polynomial degree.

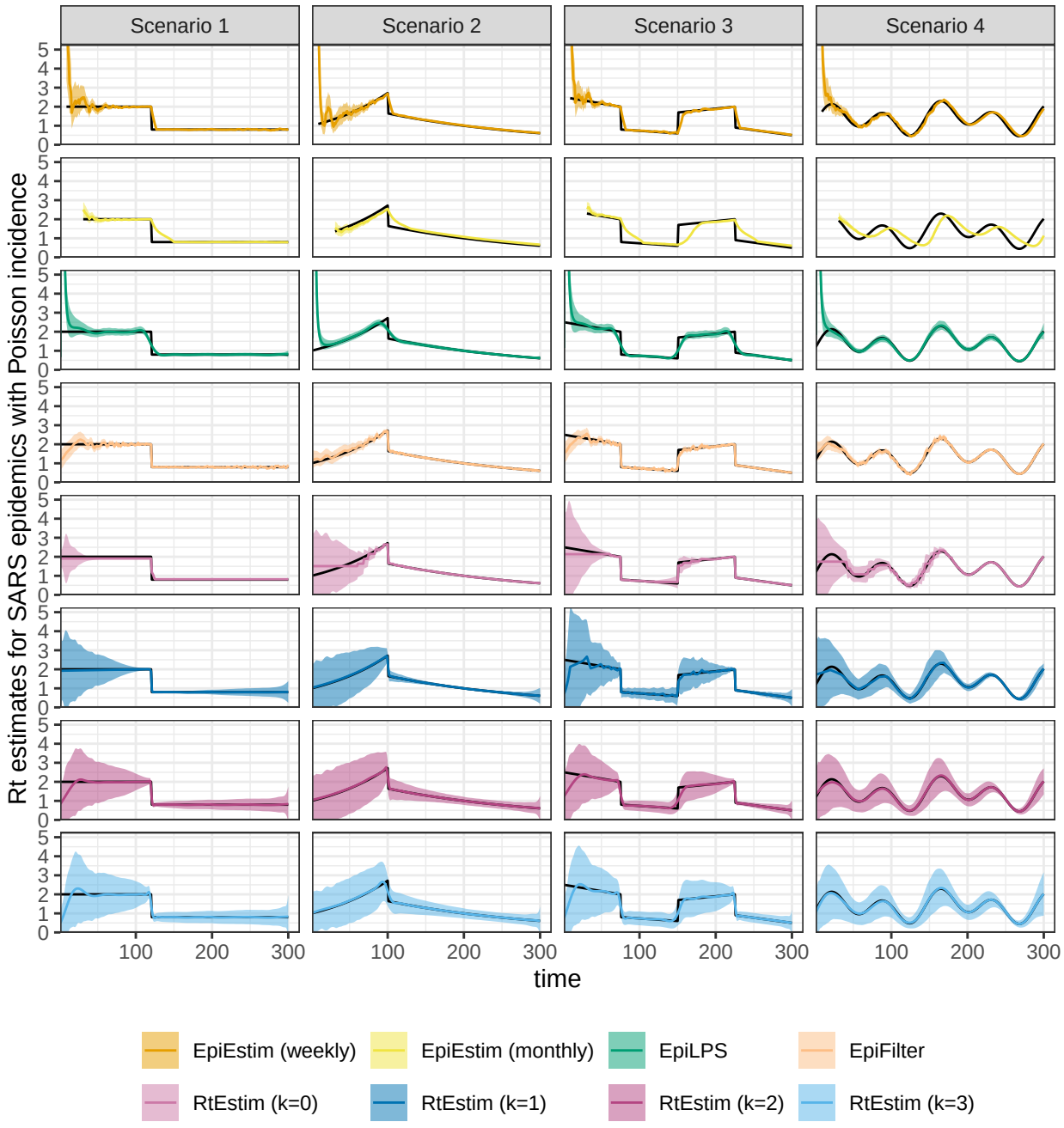

Figure A.6.3: Example SARS epidemics with Poisson incidence.

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

We apply all methods on Covid19 incidence in Canada, and the estimated are displayed in A.8.1. An alternative display which plots all estimated curves in one panel for an easier comparison is provided in A.8.2. All methods provide similar $\hat{\mathcal{R}}_t$ curves beyond the early stage. Many methods, including `RtEstim` (k=1,2), `EpiLPS`, and `EpiEstim` (weekly sliding window), all have large estimates (larger than 3) at the early stage of the epidemic. `EpiFilter` is much more wiggly than other estimated $\hat{\mathcal{R}}_t$ curves. All methods agree that the instantaneous reproduction number of Covid19 in Canada decreases to below 1 near June 2021 and reaches a

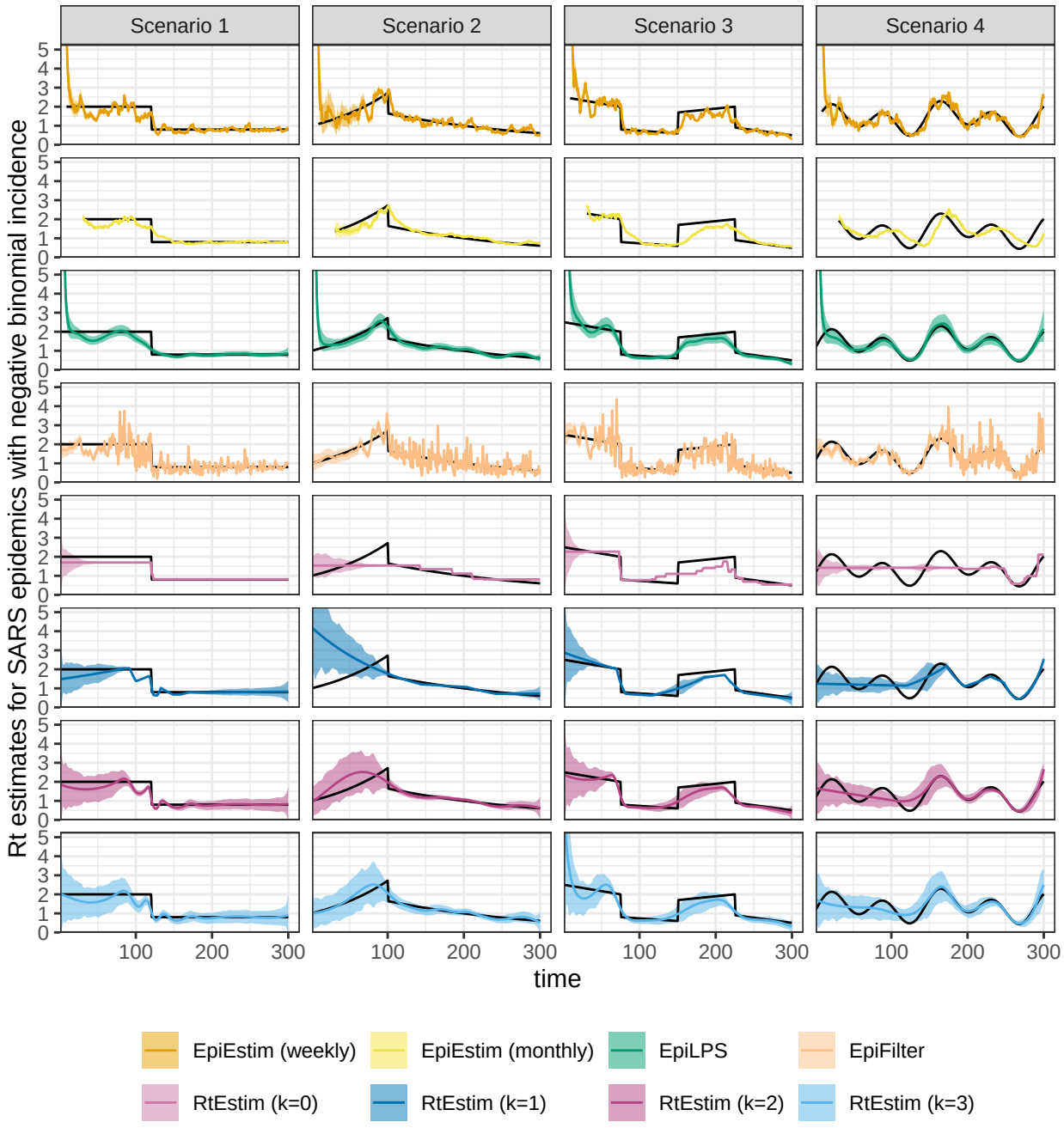

Figure A.6.4: Example SARS epidemics with negative binomial incidence.

small peak afterwards, and then decreases slowly until an outbreak at the end of 2021. The instantaneous reproduction number decreases slowly, and remains close to but below 1.

We also apply all methods on Flu in 1918 as well. The results are visualized in Figures A.8.3 and A.8.4. `EpiEstim` with weekly sliding windows, `EpiFilter` and `RtEstim` (k=0) captures the peak of $\mathcal{R}_t$ (close to 3) at around day 30 since the start of the epidemic. While `EpiEstim` with monthly sliding windows, `EpiLPS`, `RtEstim` (k=2,3) captures the increase around day 30, but have smaller estimates. Most methods agree that after around day 50, the instantaneous reproduction number decreases to and remains below 1.

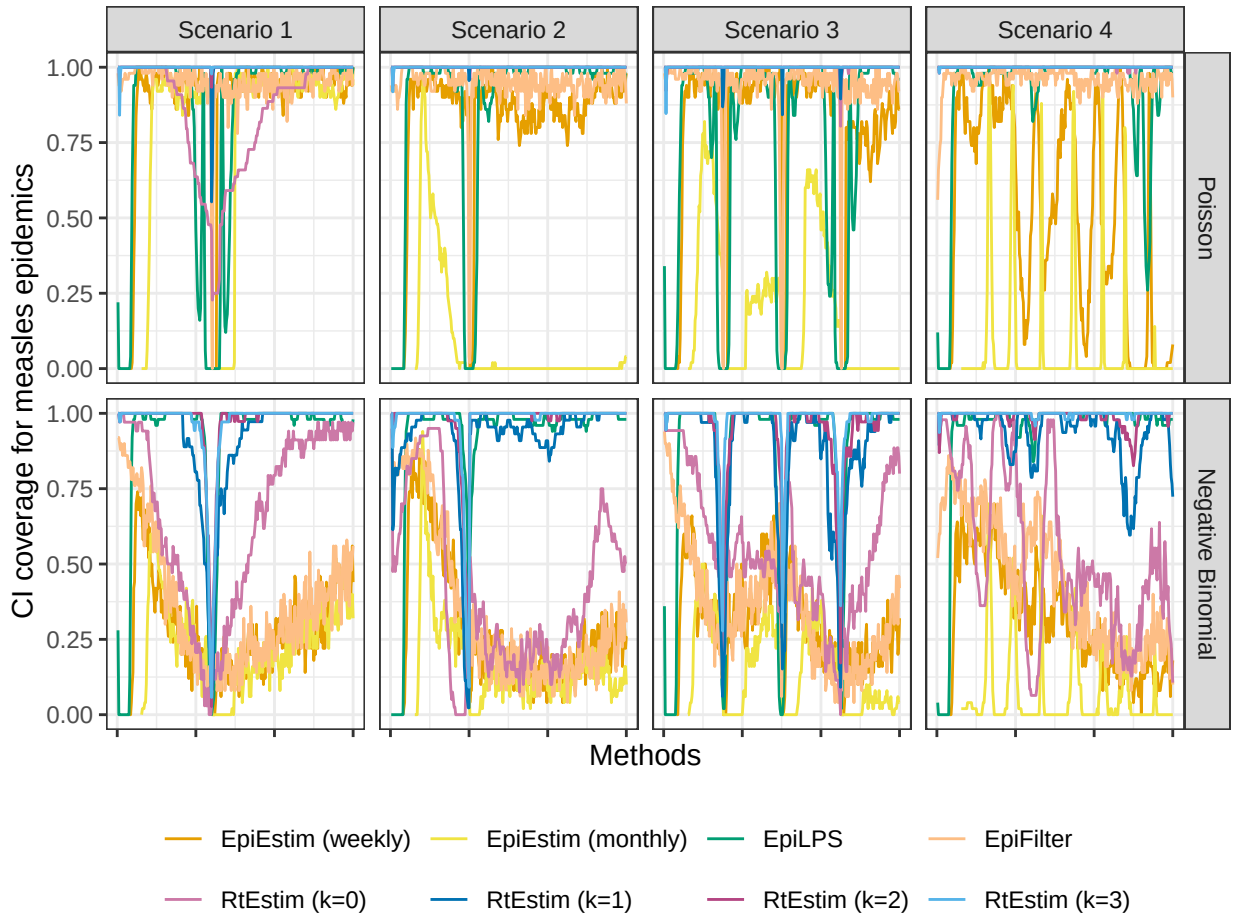

Figure A.6.5: Averaged coverage of CI per coordinate with measles epidemics.

#* References

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 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.

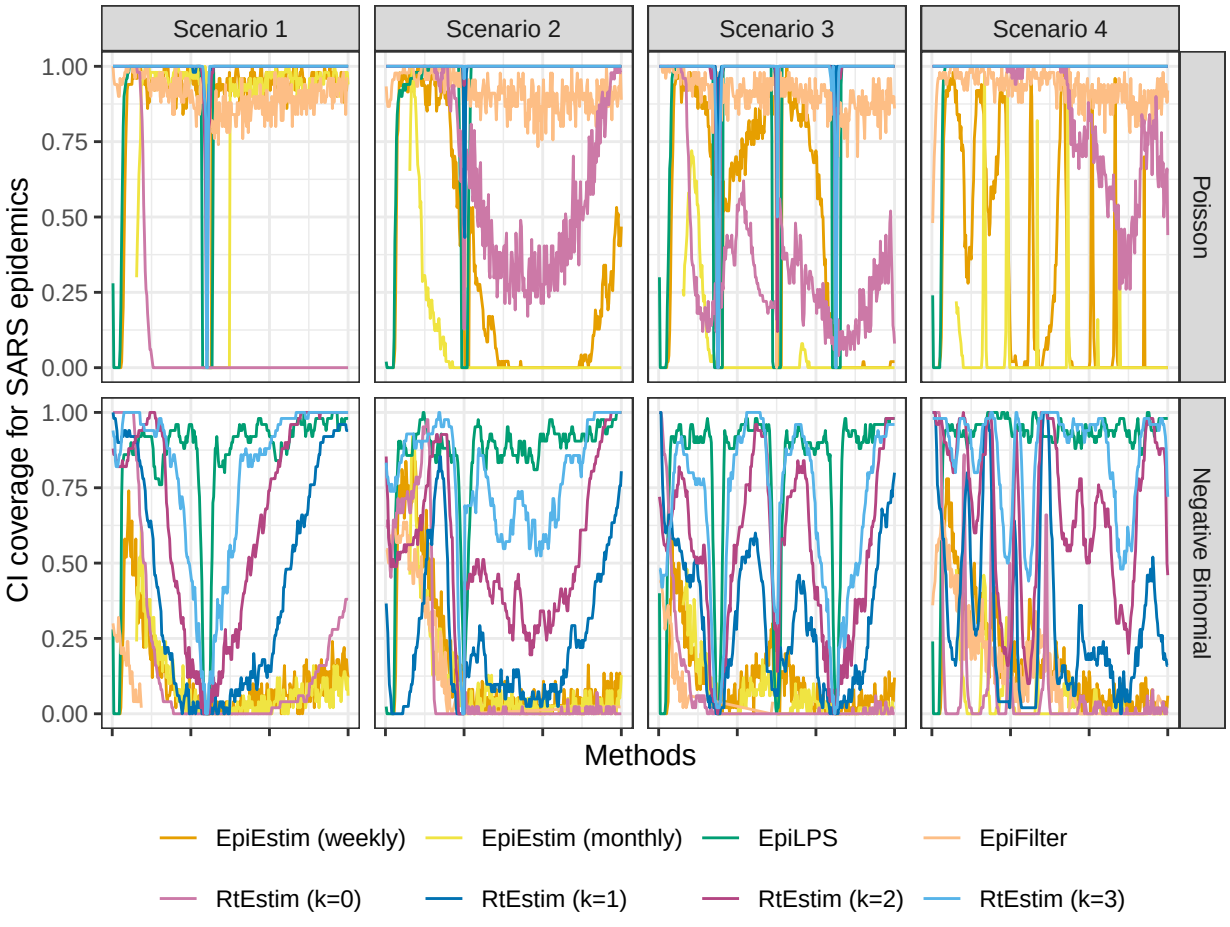

Figure A.6.6: Averaged coverage of CI per coordinate with SARS epidemics.

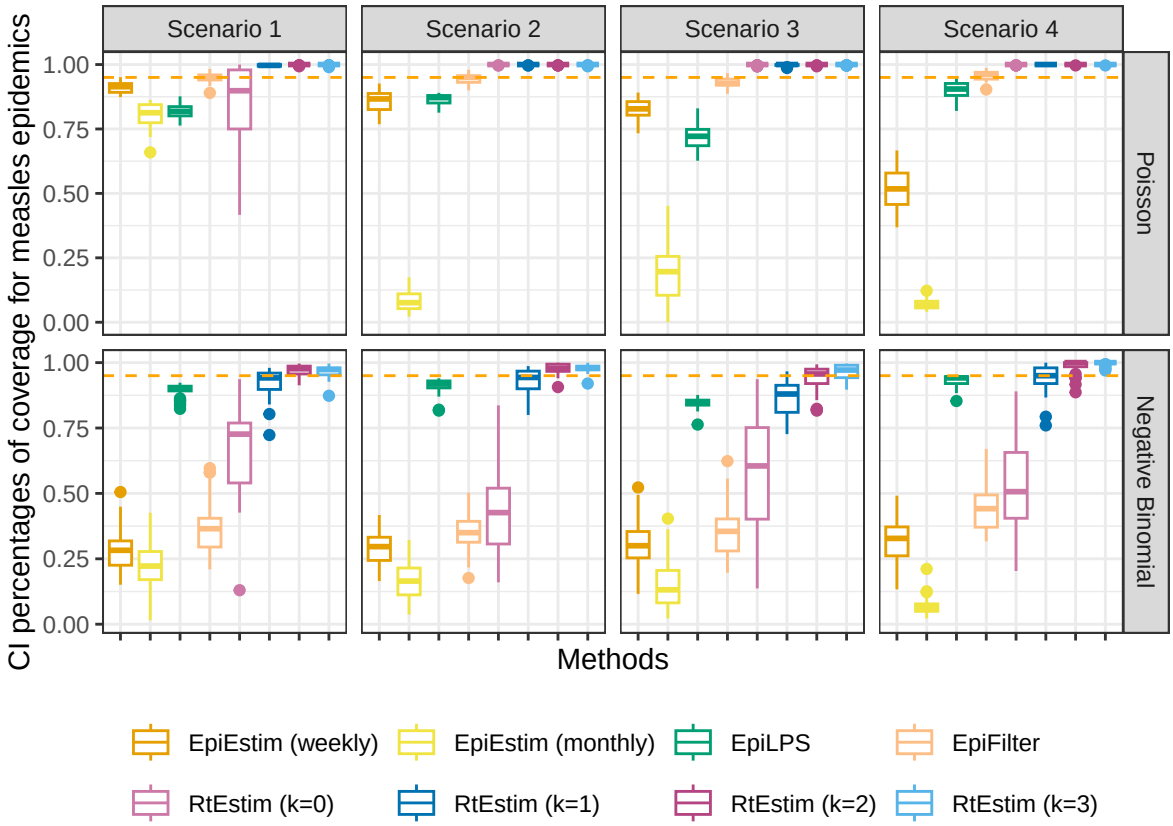

Figure A.6.7: Averaged percentages of CI coverage with measles epidemics. The orange dashed line represents 95% percentage of coverage across all timepoints.

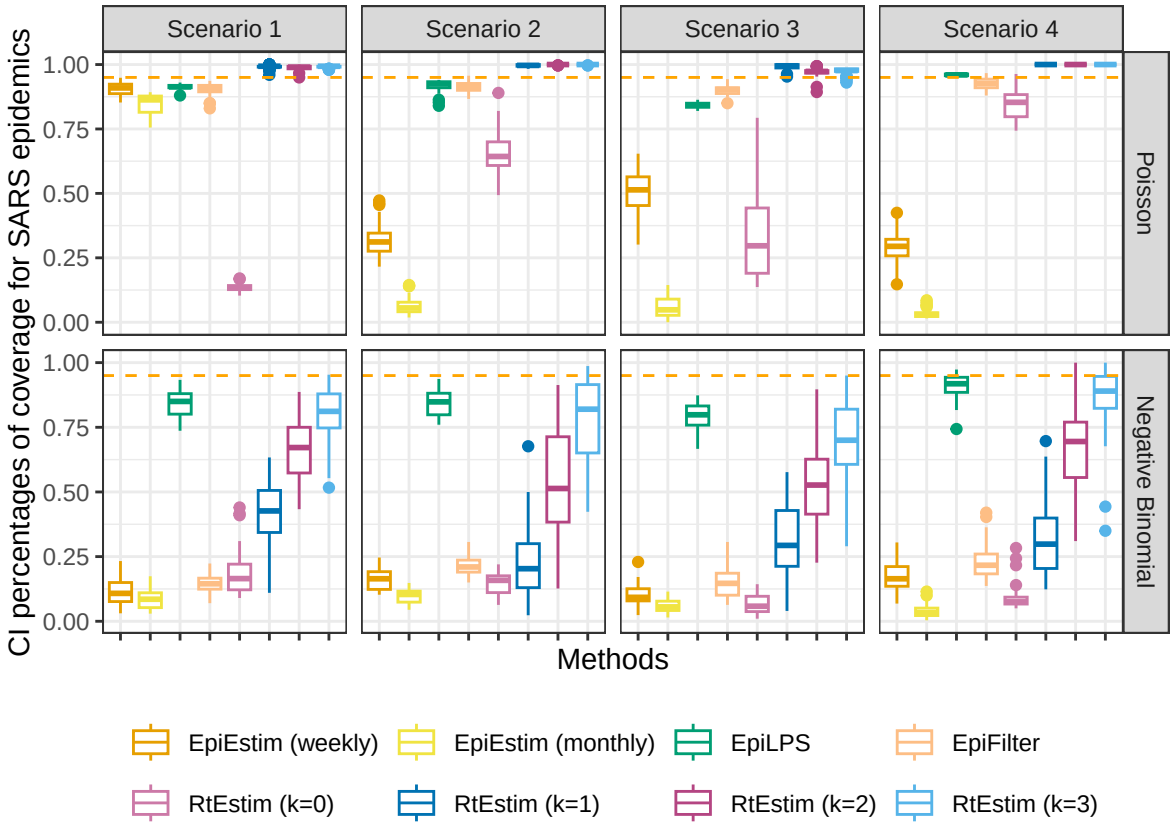

Figure A.6.8: Averaged percentages of CI coverage with SARS epidemics. The orange dashed line represents 95% percentage of coverage across all timepoints.

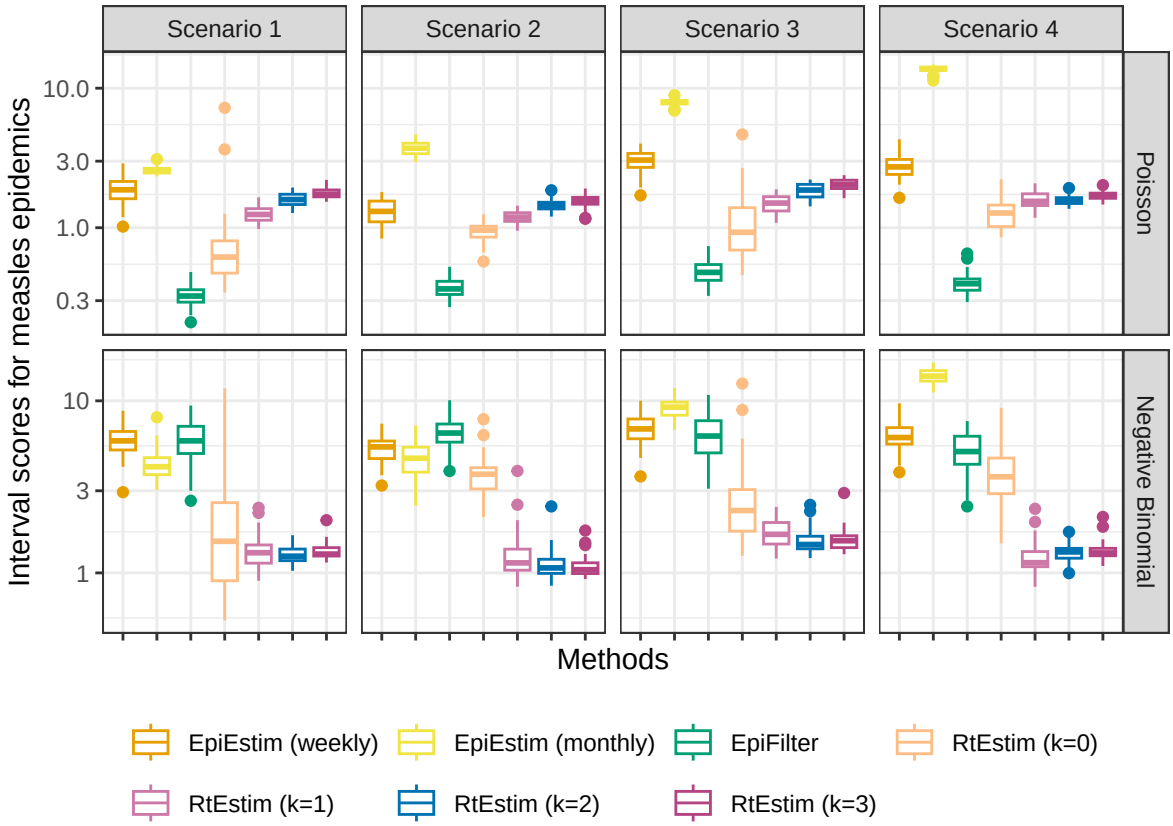

Figure A.6.9: Averaged interval scores with measles epidemics. EpiLPS is excluded, since it's scores are always larger than 100.

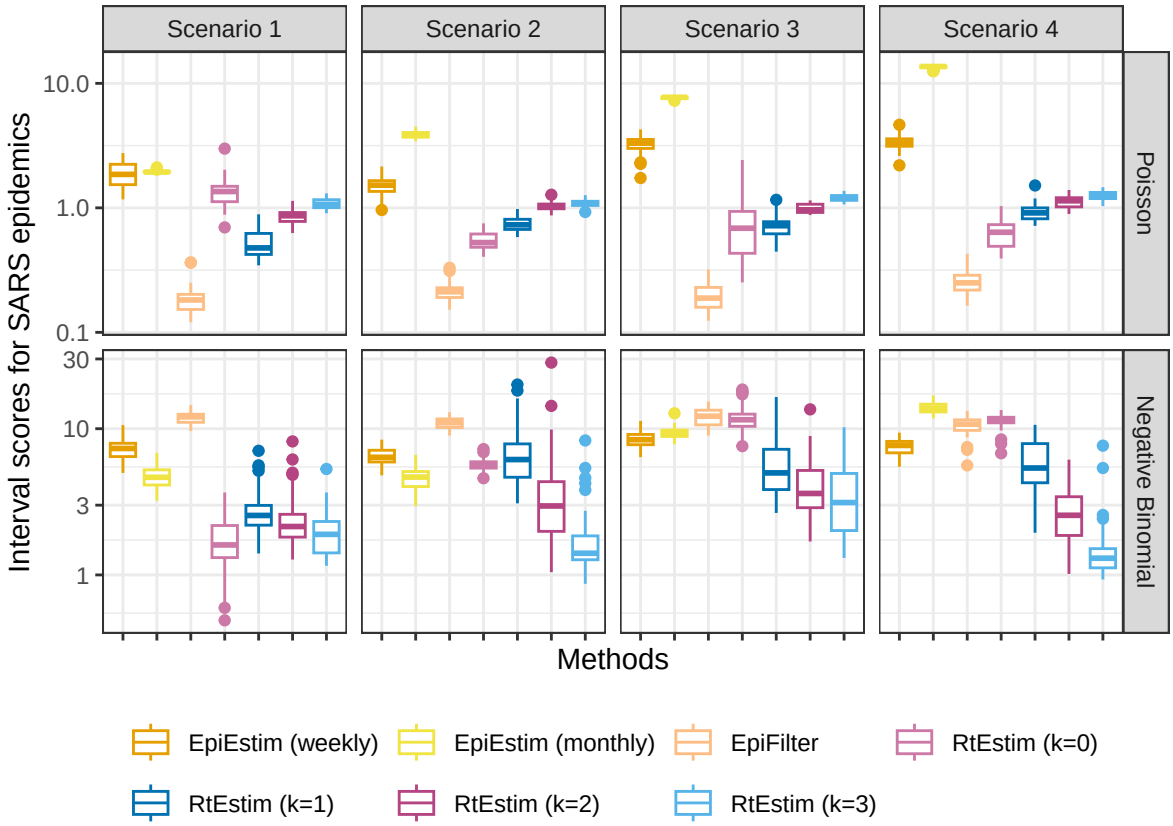

Figure A.6.10: Averaged interval scores with SARS epidemics. EpiLPS is excluded, since it's scores are always larger than 100.

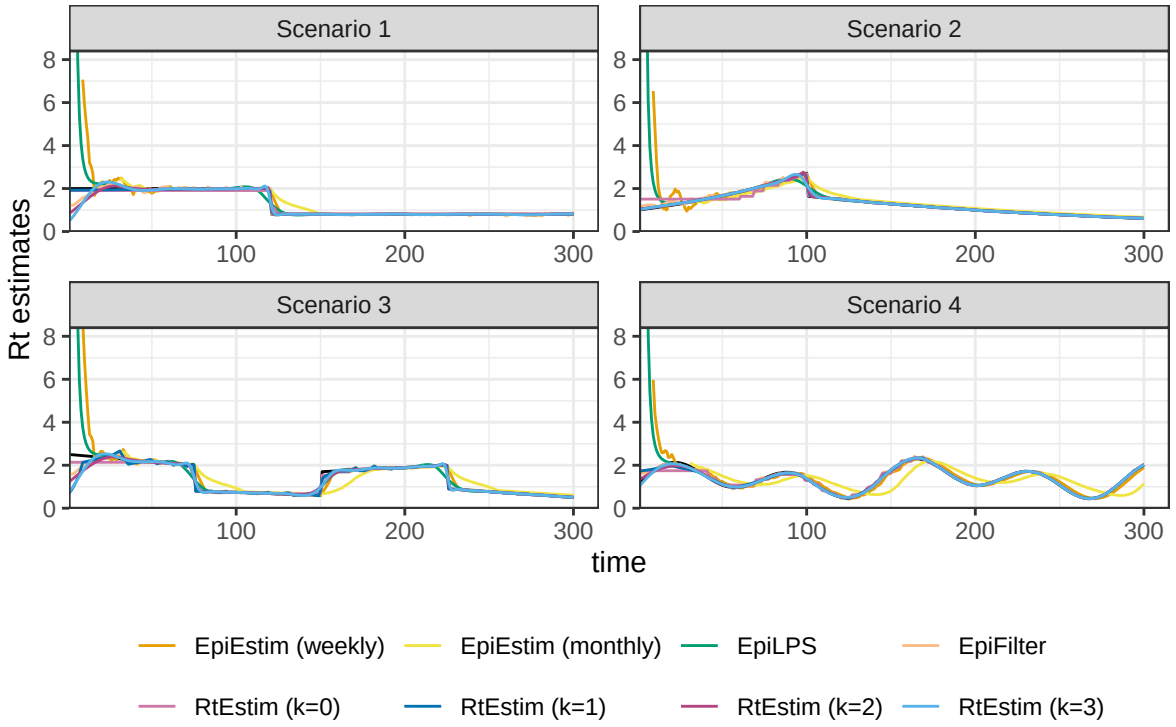

Figure A.7.1: Example of instantaneous reproduction number estimation for SARS epidemics with Poisson observations.

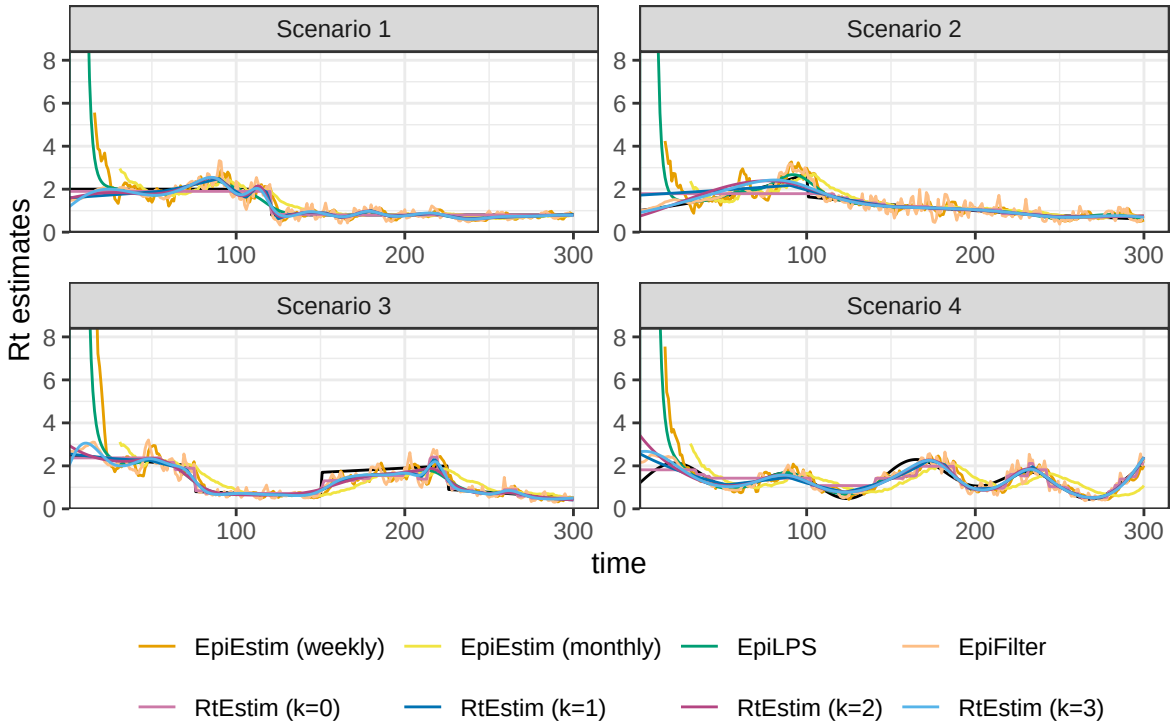

Figure A.7.2: Example of instantaneous reproduction number estimation for measles epidemics with negative binomial observations.

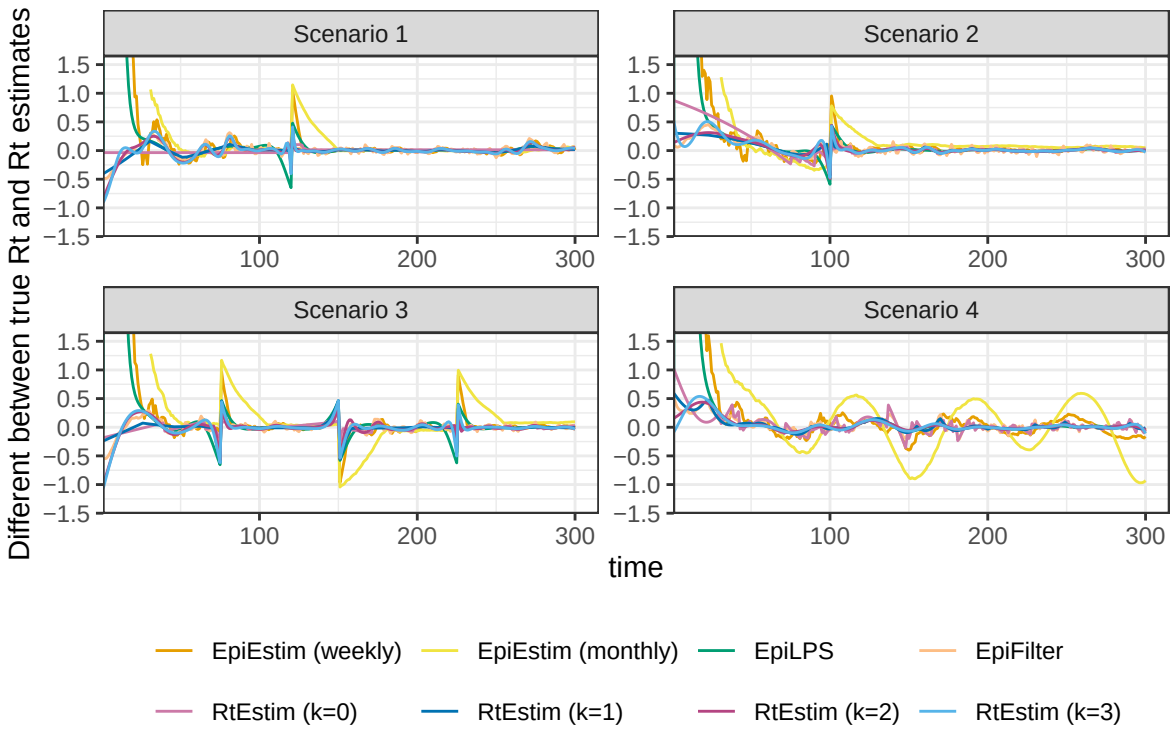

Figure A.7.3: Difference between of the true instantaneous reproduction number and its estimation for measles epidemics with Poisson observations.

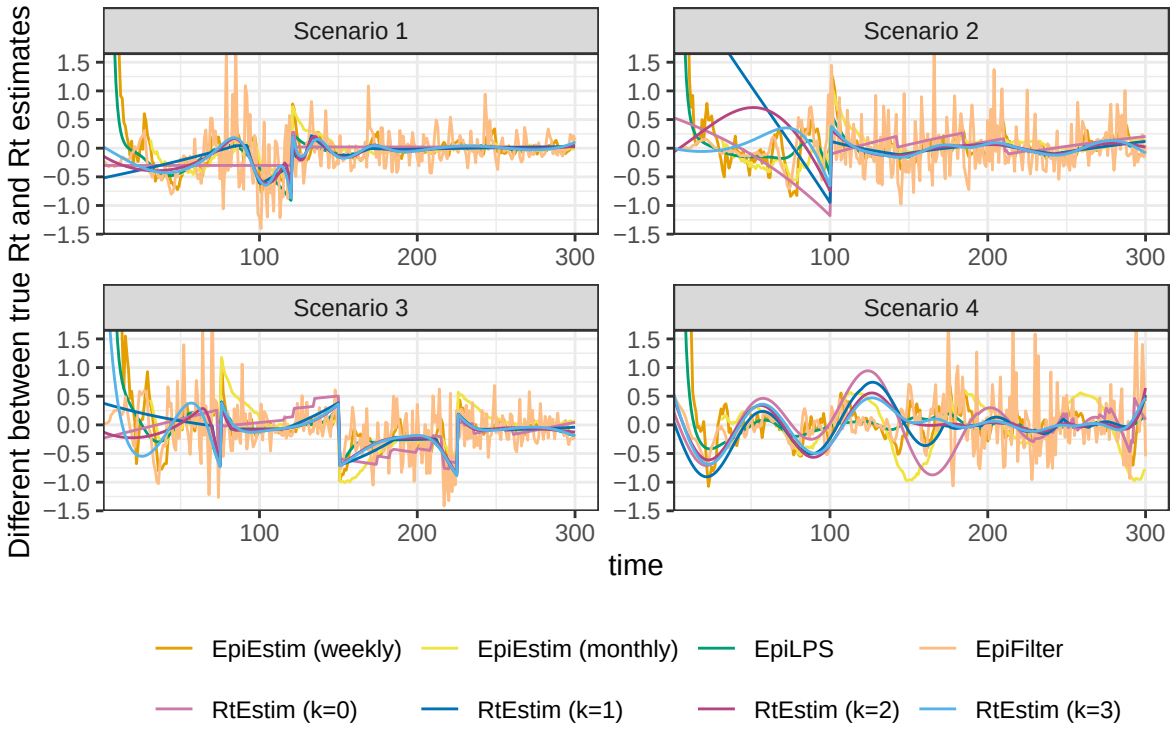

Figure A.7.4: Difference between of the true instantaneous reproduction number and its estimation for SARS epidemics with negative binomial observations.

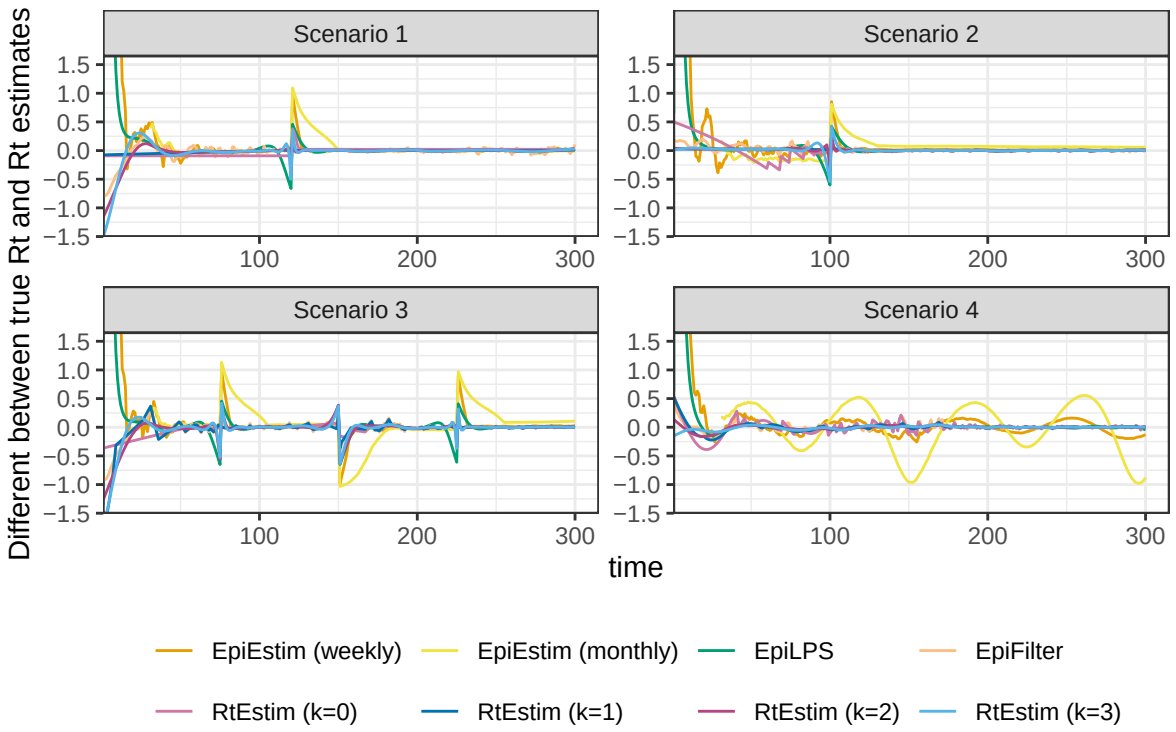

Figure A.7.5: Difference between of the true instantaneous reproduction number and its estimation for SARS epidemics with Poisson observations.

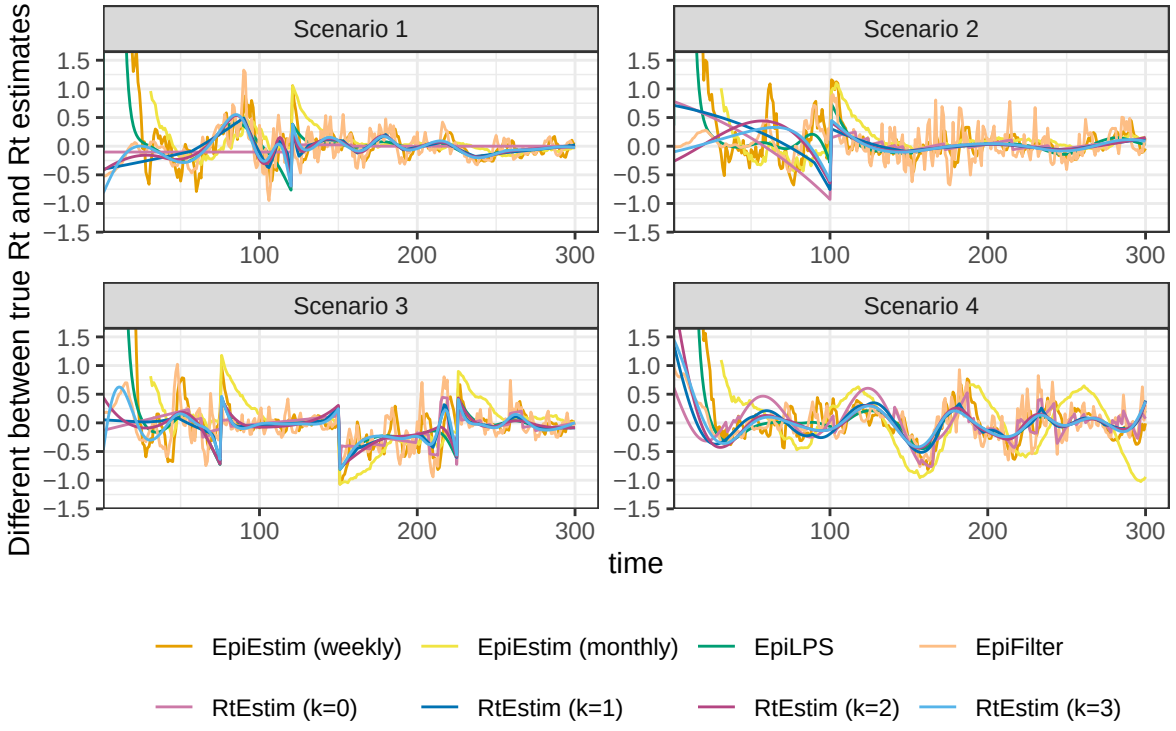

Figure A.7.6: Difference between of the true instantaneous reproduction number and its estimation for measles epidemics with negative binomial observations.

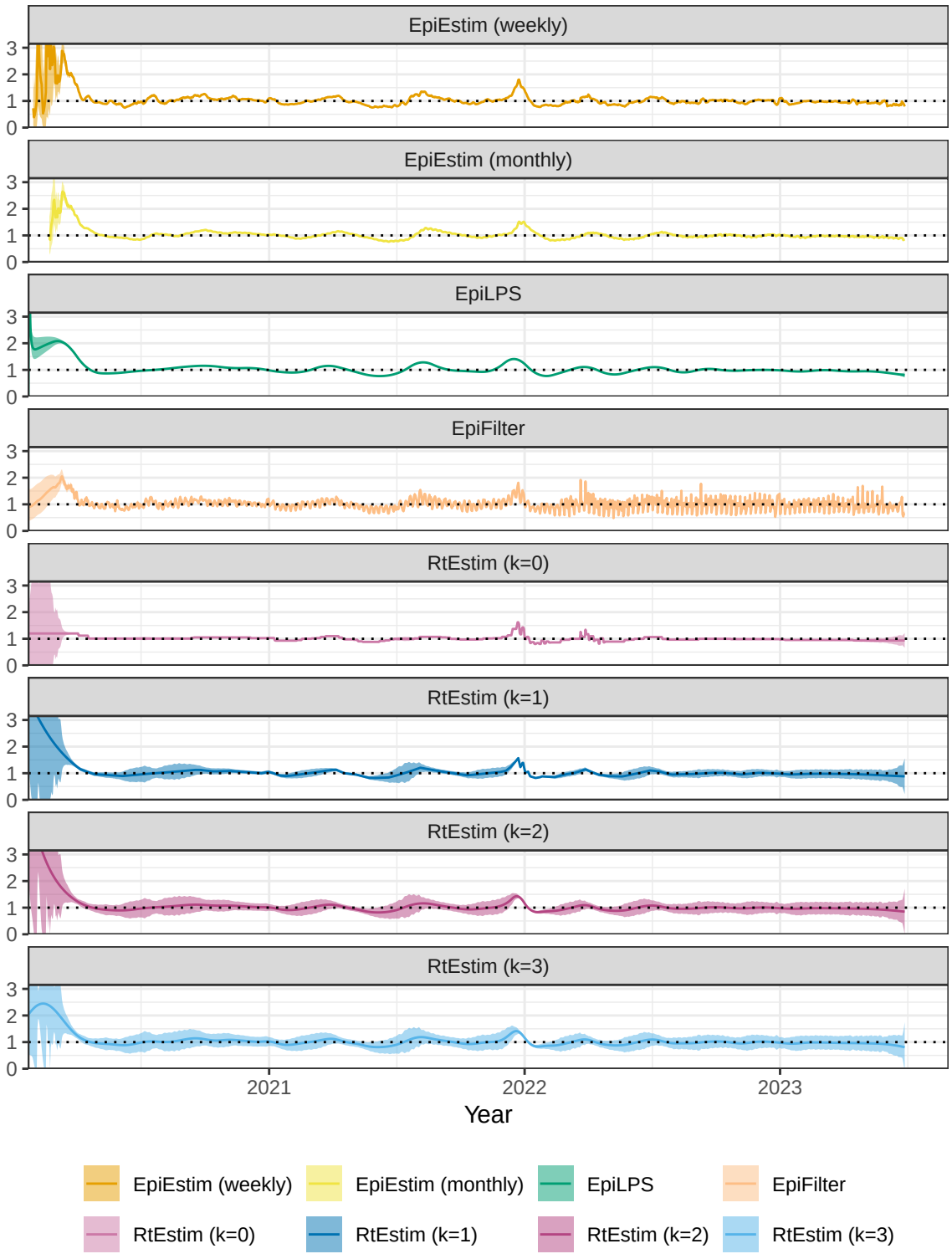

Figure A.8.1: Rt estimates with CIs for Covid19. Y-axes are truncated beyond 3 for a better display of the fluctuation in small values.

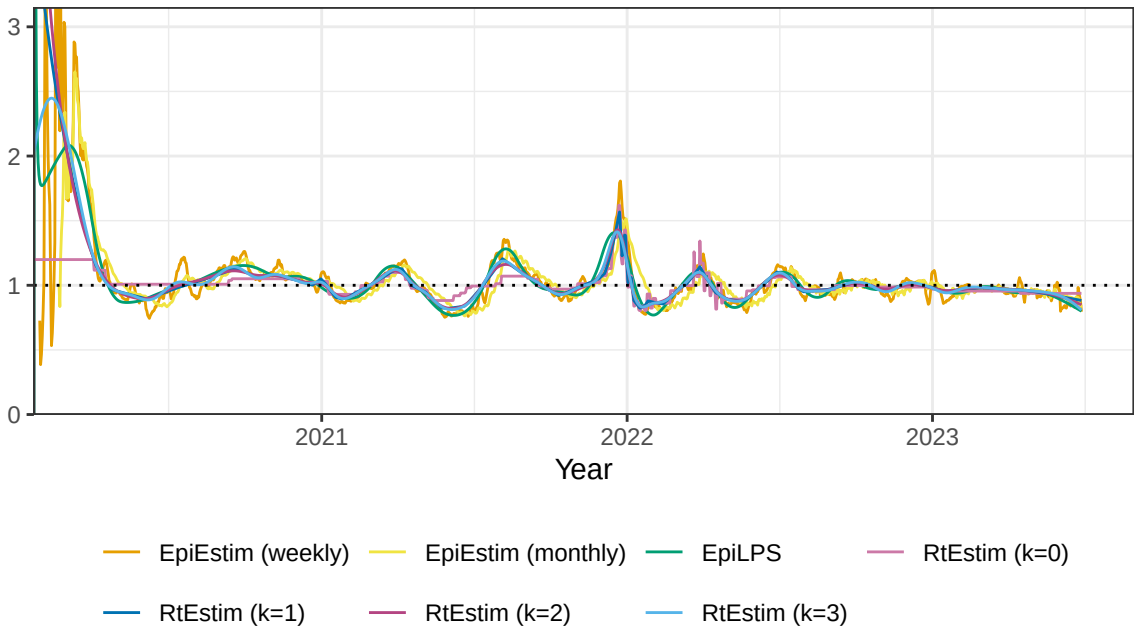

Figure A.8.2: Rt estimates for Covid19. Y-axis is truncated beyond 3 for a better display of the fluctuation in small values. EpiFilter is excluded here, because its estimates are too wiggly and make the plot less readable.

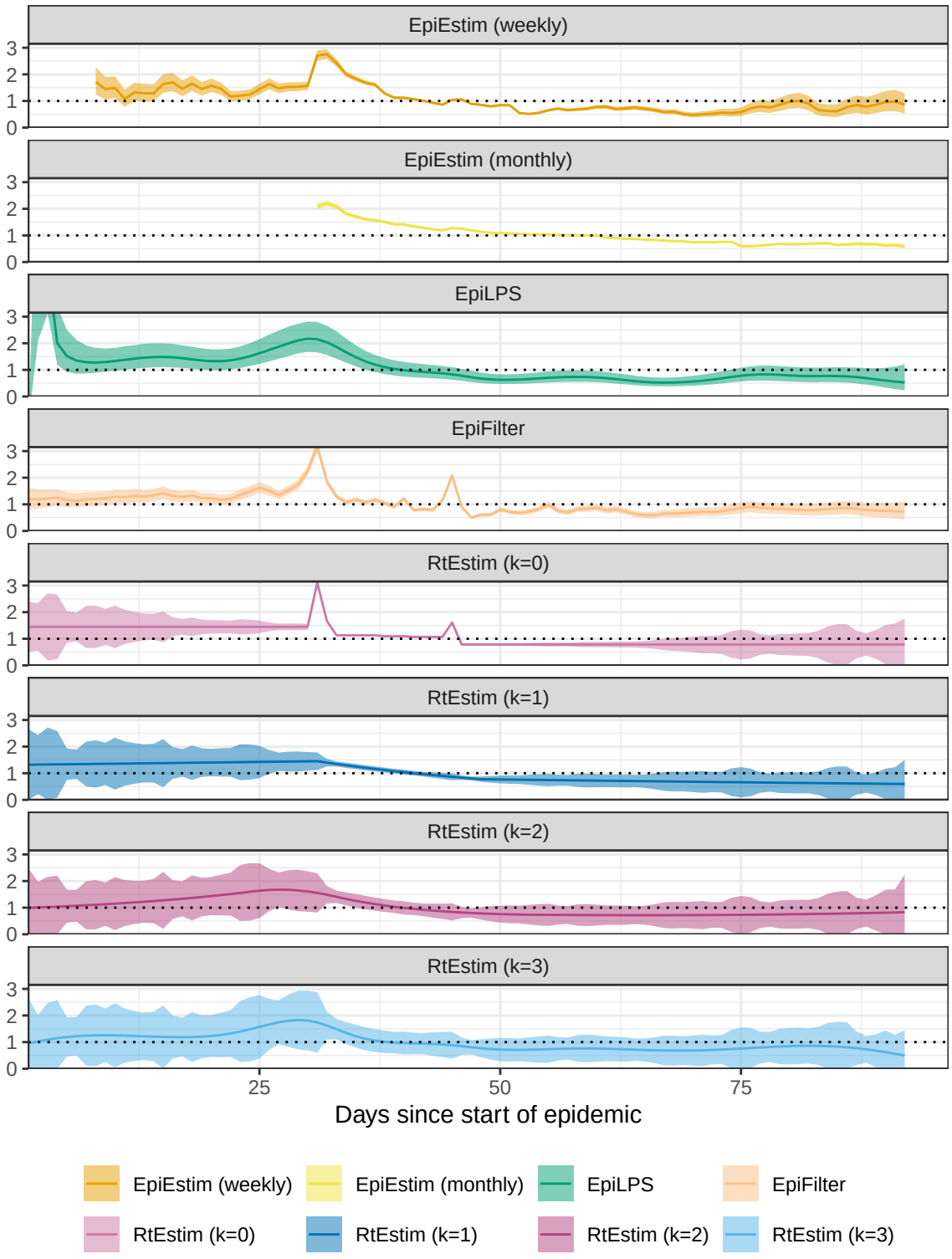

Figure A.8.3: Rt estimates with CIs for Flu 1918. Y-axes are truncated beyond 3 for a better display of the fluctuation in small values.

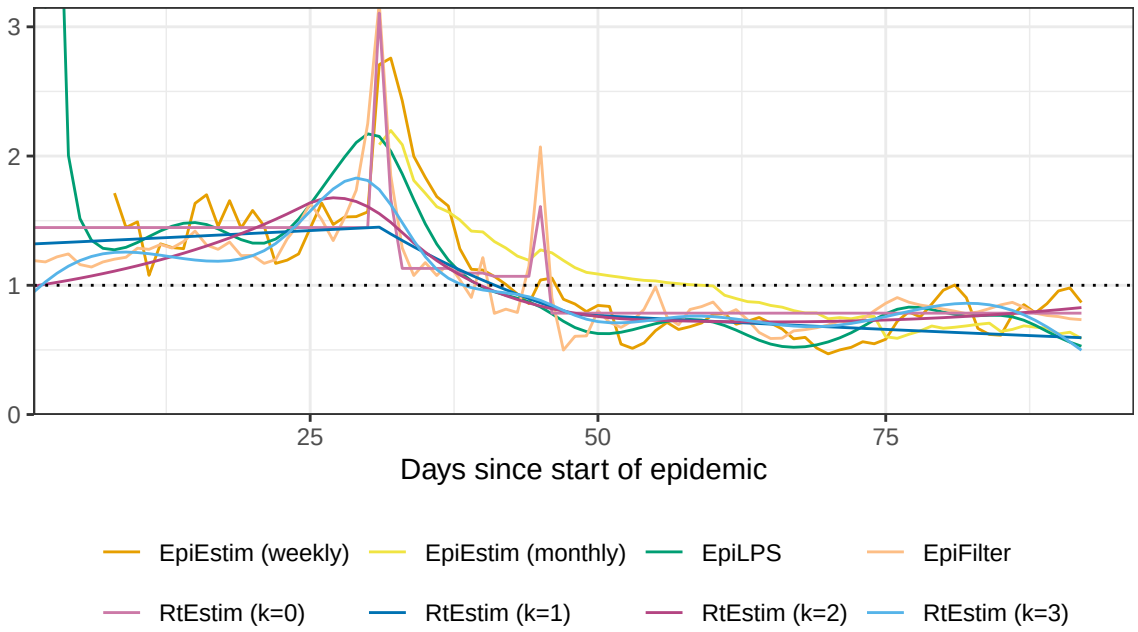

Figure A.8.4: Rt estimates for Flu 1918. Y-axis is truncated beyond 3 for a better display of the fluctuation in small values.