Supplementary details on experiments of effective 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) that is used to compare the accuracy of the estimated effective reproduction number with the true ones. Given the total infectiousness η, 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

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

where

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

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

A.2 Supplmentary 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 runned 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). 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 a shaped (mean increased/decreased by 2) and scaled (standard deviation increased by 10) parameters for both flu/measles epidemics, denoted as `flu_ss` and `measles_ss` respectively. These settings result in 7 pairs of SI distributions (for epidemic generating, and for model fitting), i.e., (`measles`, `measles`), (`SARS`, `SARS`), (`measles`, `measles_ss`), (`measles`, `SARS`) for long epidemics and (`flu`, `flu`), (`flu`, `flu_ss`), (`flu`, `measles`) for short epidemics. Figure A.2.2 displays all SI distributions (`measles`, `measles_ss`, `SARS`, `flu`, and `flu_ss`) 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

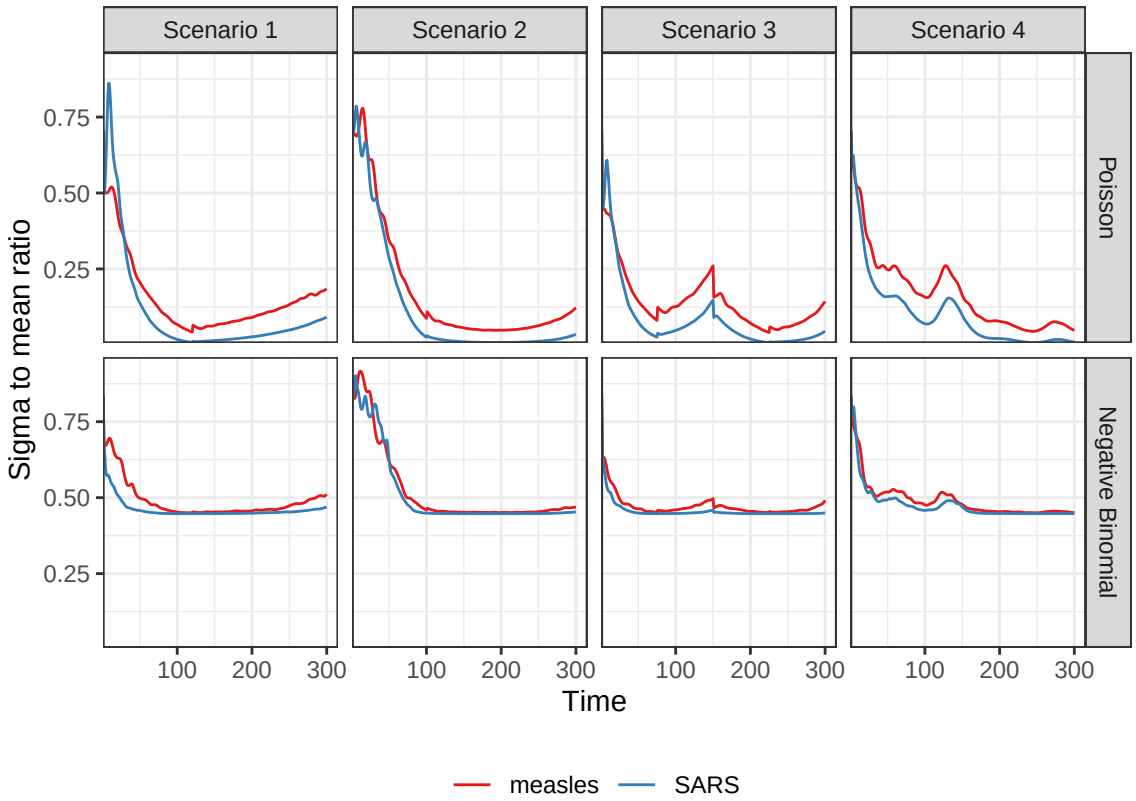

Figure A.2.1: Dispersion level of incidence of sample epidemics

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, measles_ss, SARS	8 methods
300	SARS	1-4	Poisson, NB	SARS	8 methods
50	flu	3	Poisson, NB	flu, flu_ss, measles	9 methods

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.

Here we list the hyperparameters used in modelling for each method. Most of them are the experimental settings used in the papers where they were proposed and deemed as the "best" tuned ones. We consider both weekly and monthly sliding windows in `EpiEstim`, 40 basis functions in `EpiLPS` with the NelderMead method to maximize the hyperparameter posterior distribution. We input 2000 grid size in `EpiFilter` with 0.1 diffusion noise and uniform prior on $\mathcal{R}_t$ with mean 1/2000, and use the smoothed $\mathcal{R}_t$ given all observed incidence as the final estimates. We run 10-fold cross validation (CV) to choose the best tuning parameter from the candidate set of size 50, i.e., $\boldsymbol{\lambda} = \{\lambda_1, \cdots, \lambda_{50}\}$, for long epidemics, and 5-fold CV for short epidemics. Specifically, we divide all samples (except the first and last entries) into, e.g., 10 folds evenly and randomly, and build models on each subset of samples by leaving a fold out using each choice of the tuning parameter. We select the tuning parameter that gives the lowest averaged ***deviance*** between the estimated incidence and the observed samples averaged over all folds.

We visualize the selected key results of the accuracy comparison using long synthetic epidemics in Section 3.2.

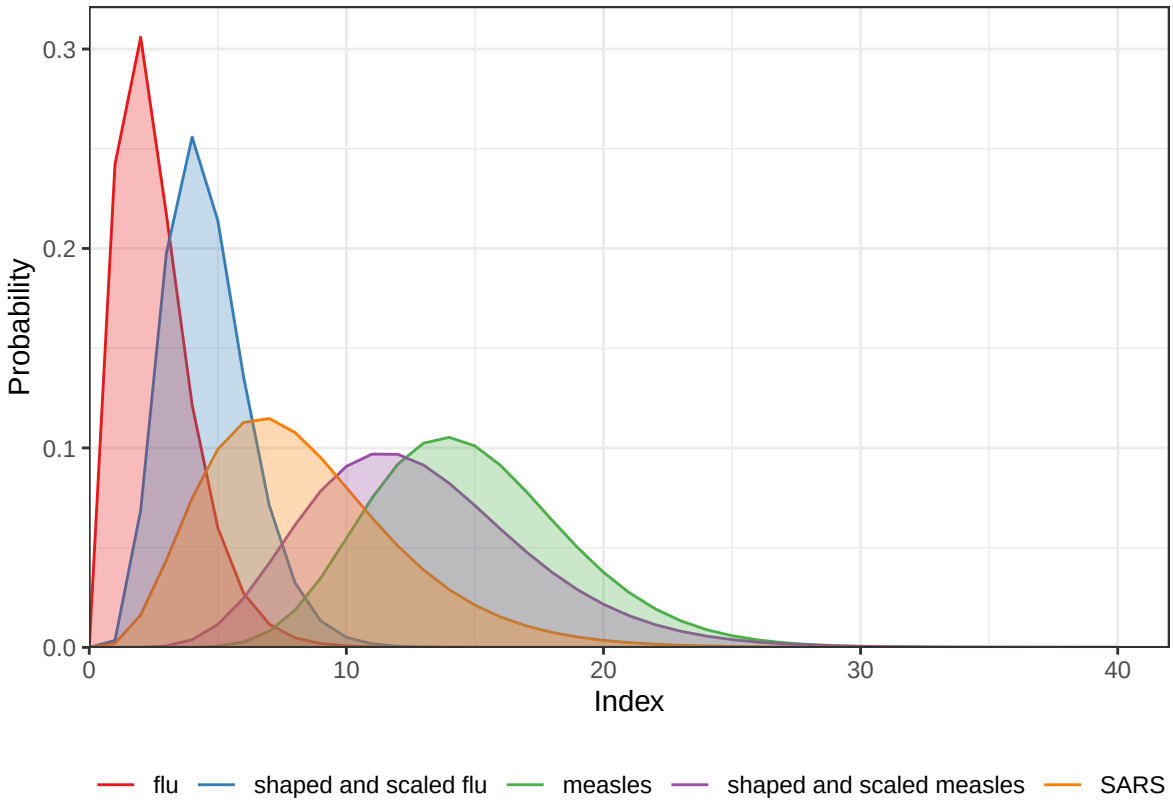

Figure A.2.2: Density curves of serial interval distributions used in the experiments

Other main experimental results are displayed in Section A.3.

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.

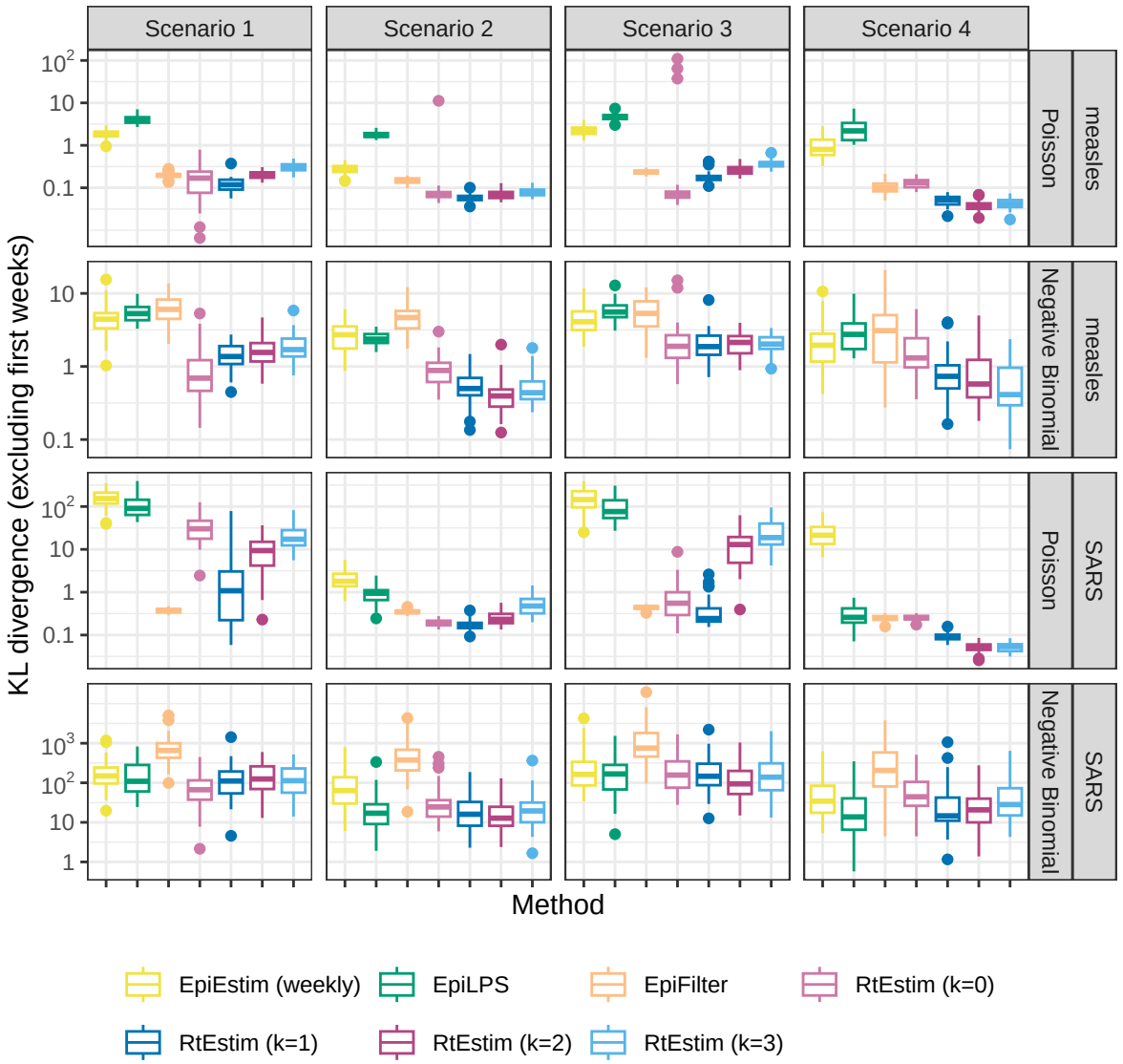

Figure A.3.1: KL divergence excluding the first weeks for measles and SARS epidemics. Y-axis is on a logarithmic scale.

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

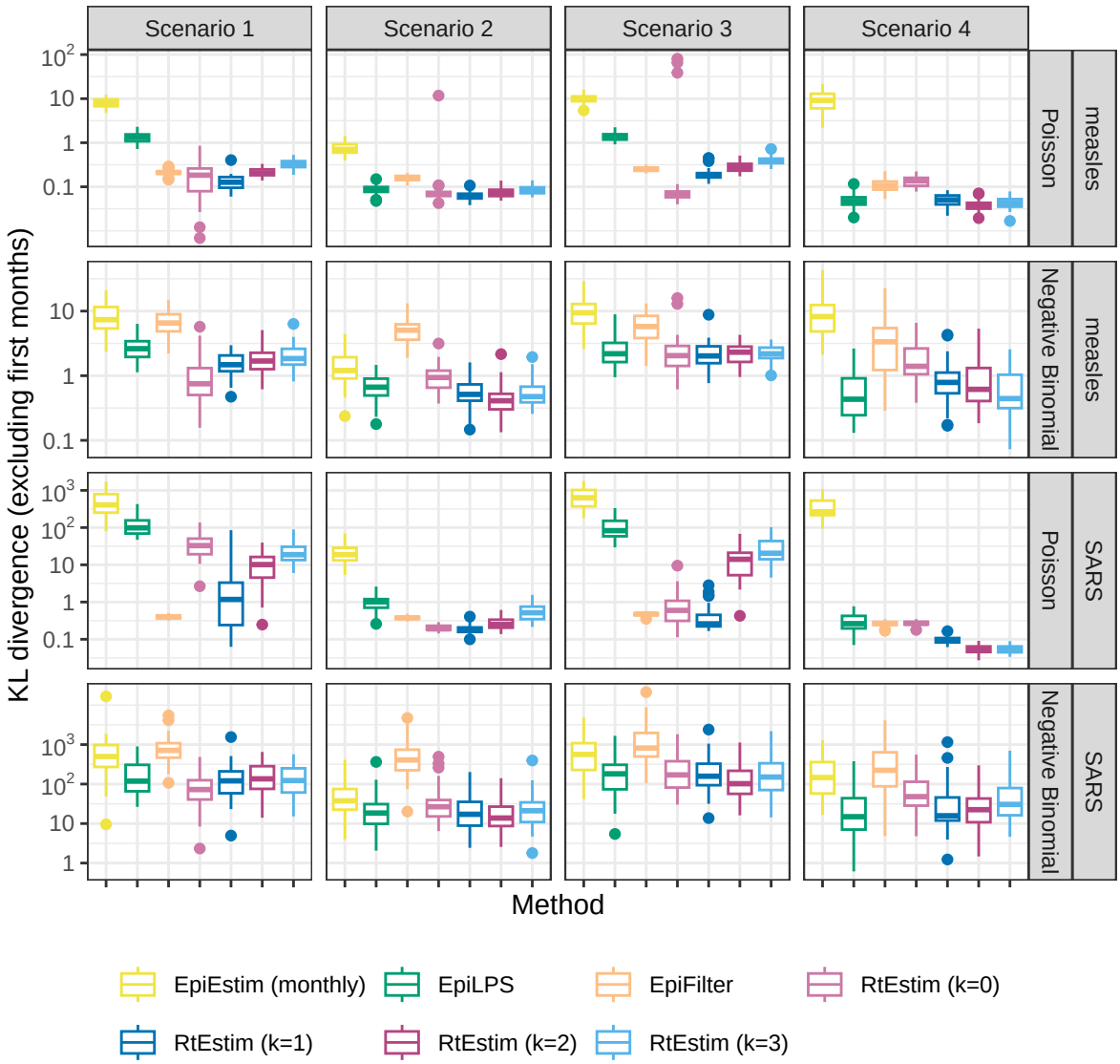

Figure A.3.2: KL divergence excluding the first months for measles and SARS epidemics. Y-axis is on a logarithmic scale.

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

```
## Warning: Removed 400 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```

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.

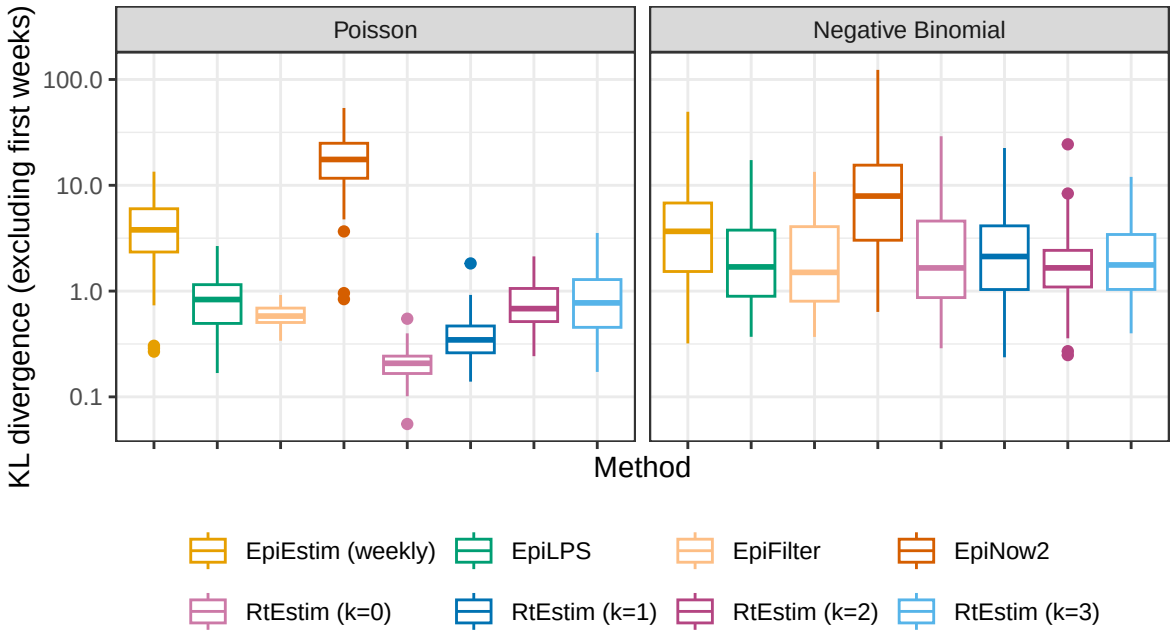

Figure A.3.3: KL divergence excluding the first weeks for flu epidemics. Y-axis is on a logarithmic scale.

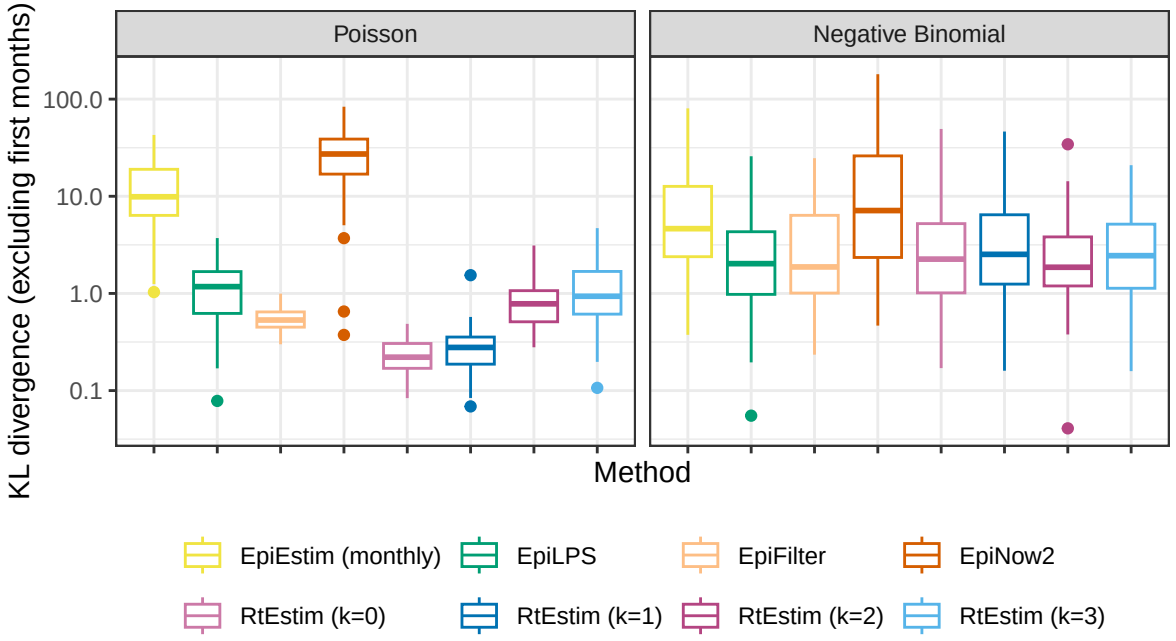

Figure A.3.4: KL divergence excluding the first months for flu epidemics. Y-axis is on a logarithmic scale.

A.5 Time comparisons of all methods

Figures A.5.1 and A.5.2 show the time comparisons across all methods. `EpiEstim` with both weekly and monthly 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. Piecewise linear and cubic `RtEstim` (with $k = 1$ and $k = 3$ respectively) are slower, but mostly within 10 seconds.

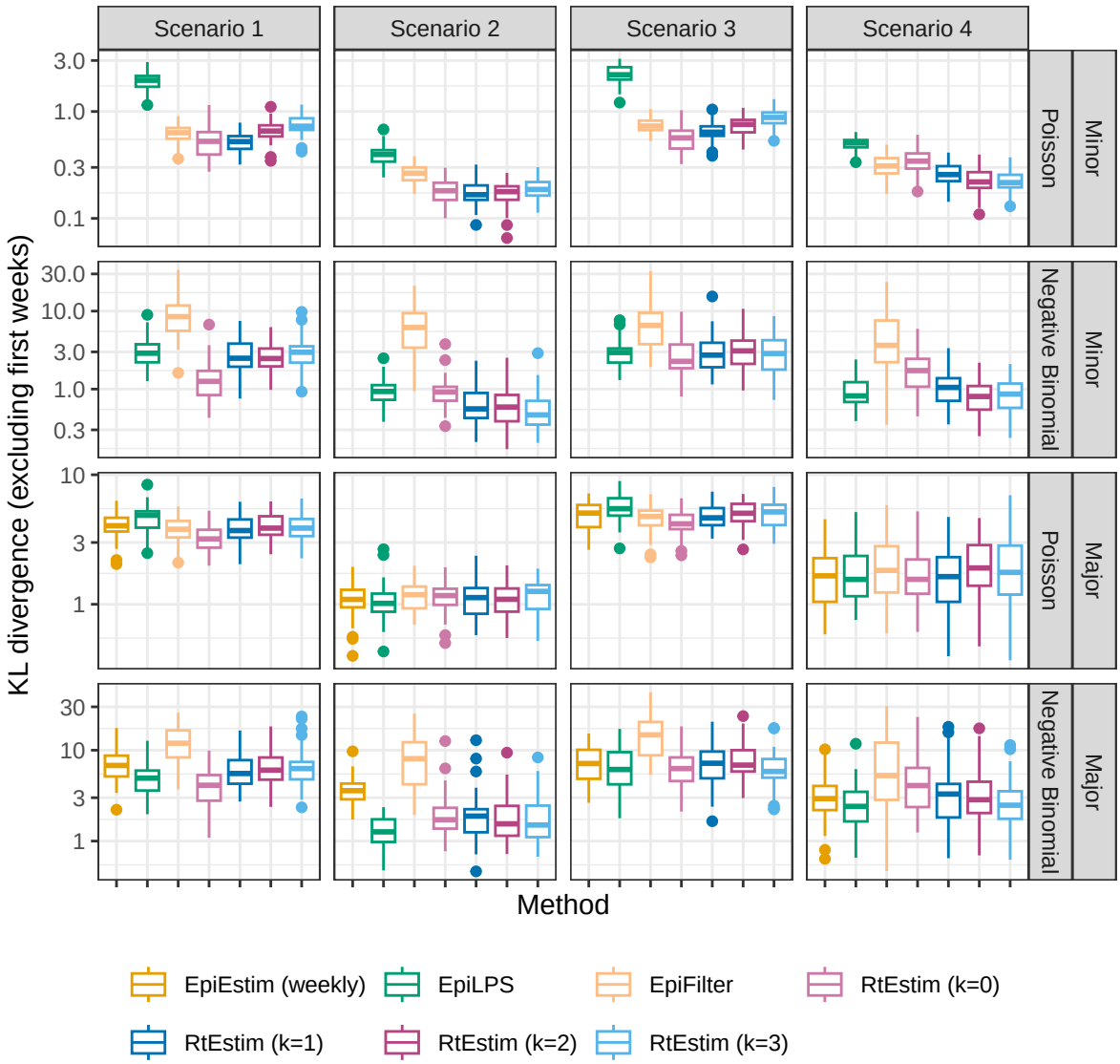

Figure A.4.1: KL divergence excluding the first weeks for measles epidemics with SI misspecification. Y-axis is on a logarithmic scale.

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.

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

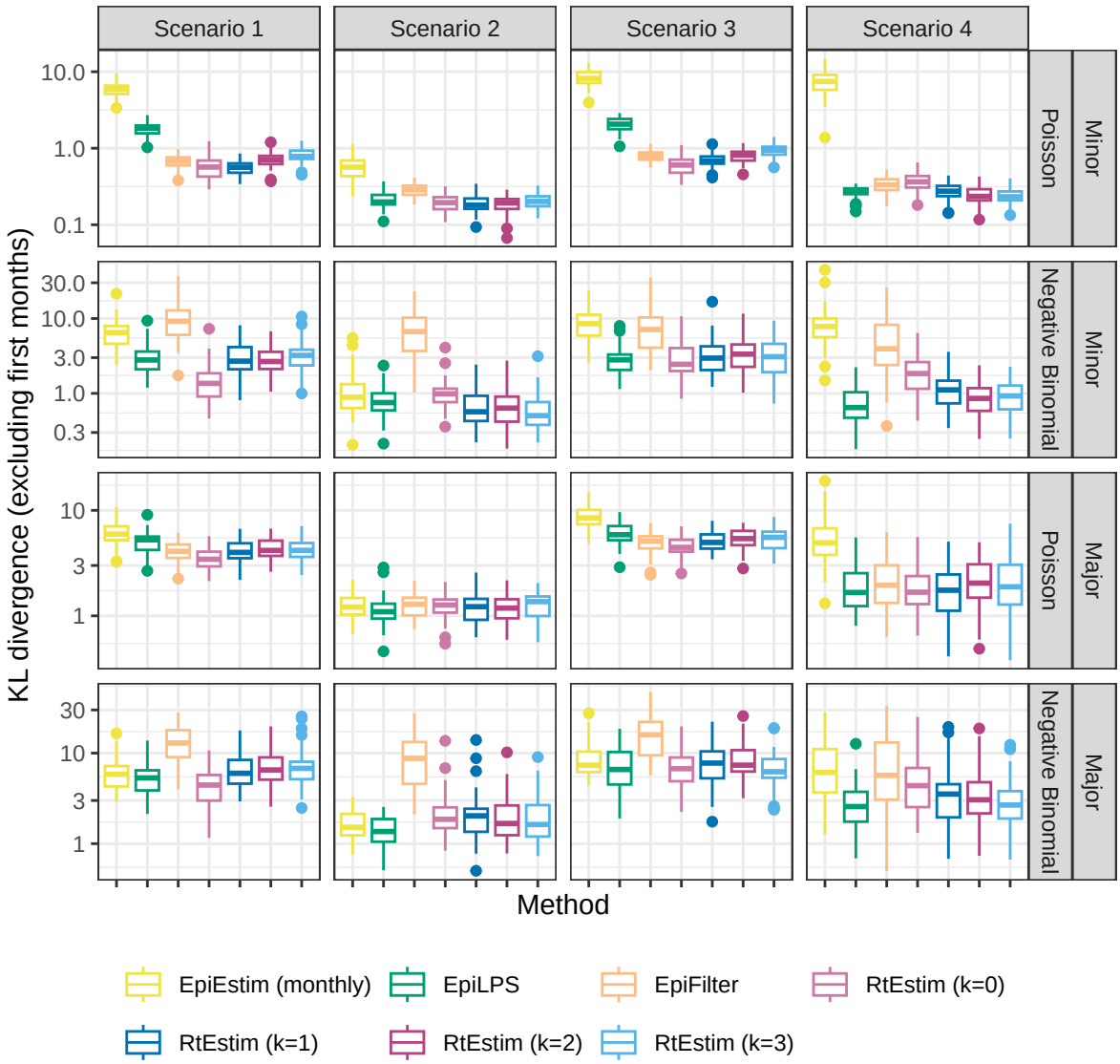

Figure A.4.2: KL divergence excluding the first months for measles epidemics with SI misspecification. Y-axis is on a logarithmic scale.

epidemics are visualized in Figures **??** and **??** as follows.

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.

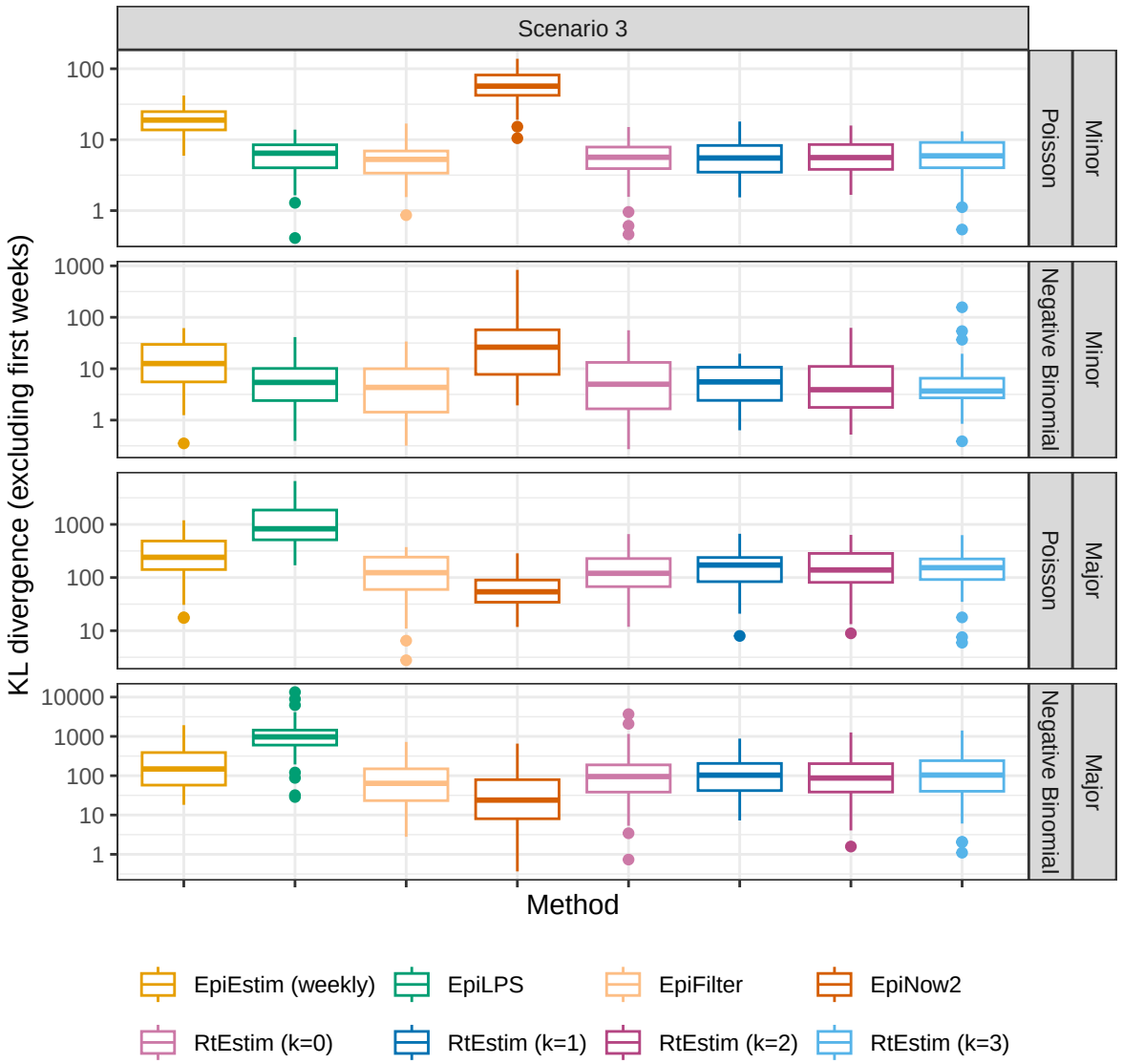

Figure A.4.3: KL divergence excluding the first weeks for flu epidemics with SI misspecification. Y-axis is on a logarithmic scale.

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.

Figures A.6.9 and A.6.10 displays the interval scores of 95% CI averaged over 50 random measles and SARS epidemics respectively.

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

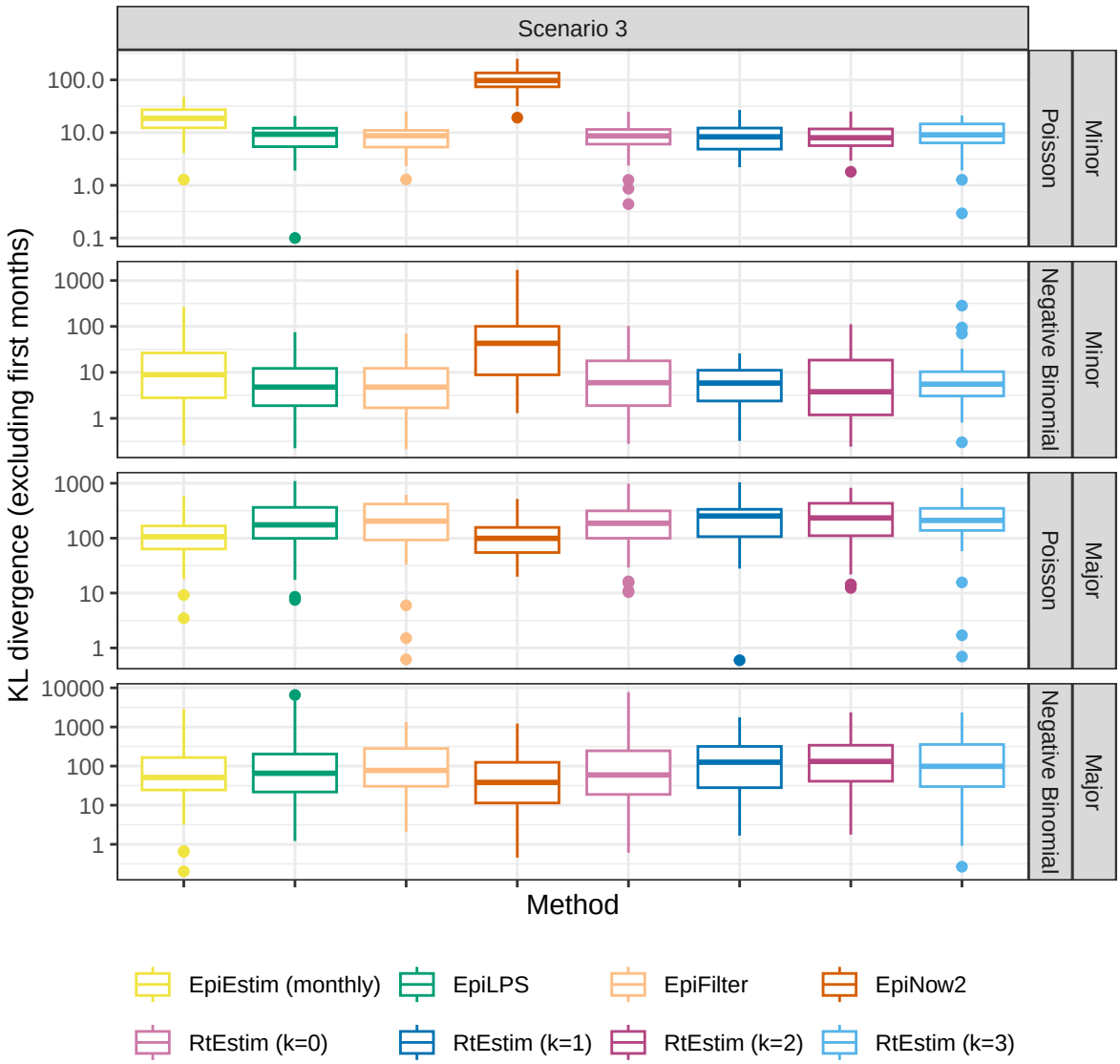

Figure A.4.4: KL divergence excluding the first weeks for flu epidemics with SI misspecification. Y-axis is on a logarithmic scale.

in Section A.6.1. The condensed visualization of other cases is provided below in Figures A.7.1 and A.7.2.

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.

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 A.8.1. An alternative display which plot all estimated curves in one panel for an easier comparison is provided in A.8.2.

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

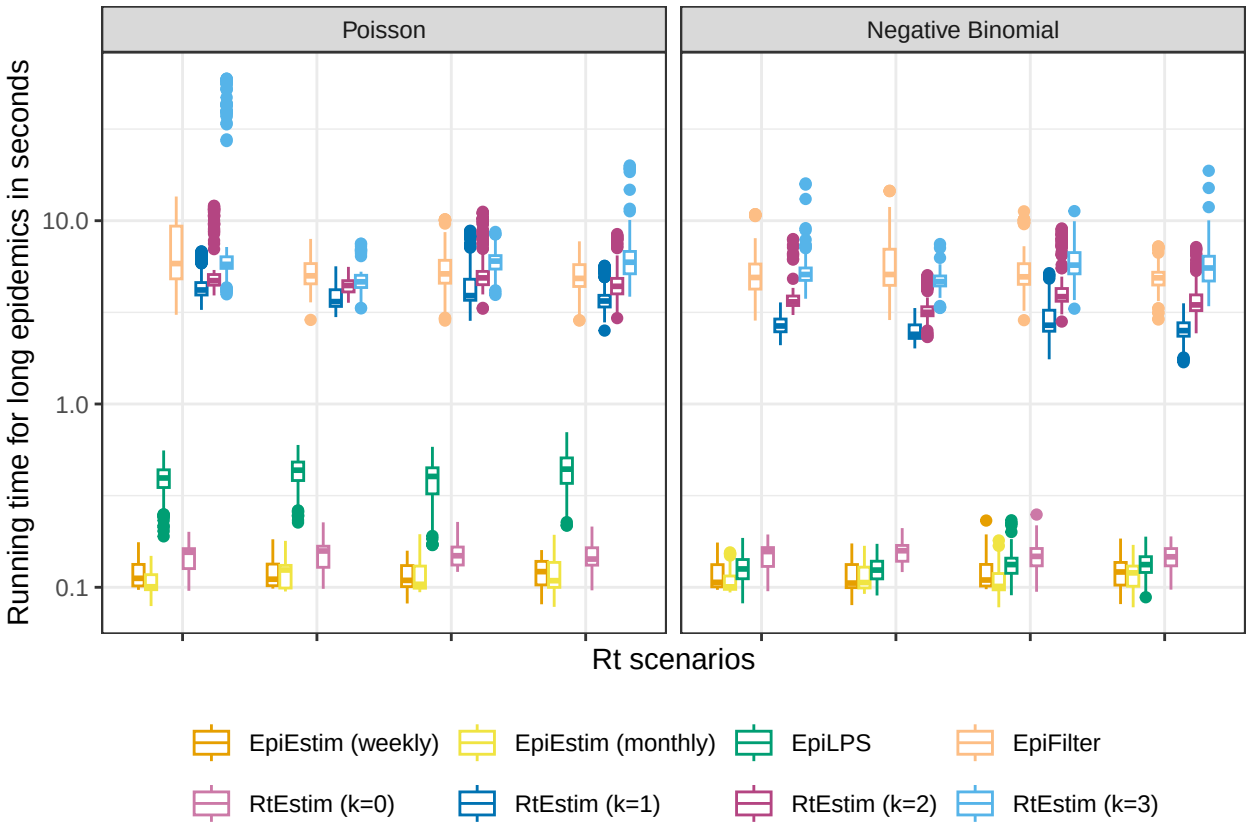

Figure A.5.1: Time comparisons of methods (excluding one outlier of 'RtEstim (k=1)' in Scenario 2 with negative Binomial incidence). Y-axis is on a logarithmic scale.

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

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.

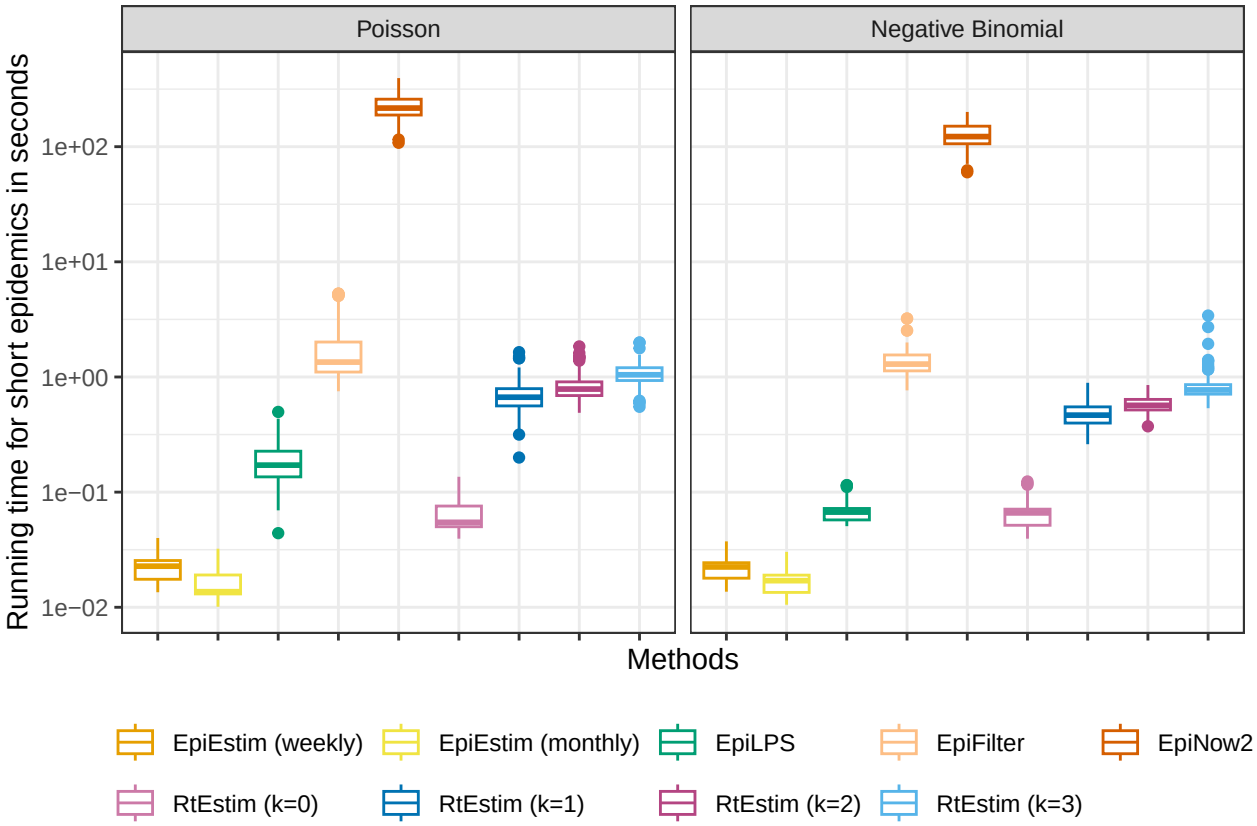

Figure A.5.2: Time comparisons of methods (excluding one outlier of 'RtEstim (k=1)' in Scenario 2 with negative Binomial incidence). Y-axis is on a logarithmic scale.

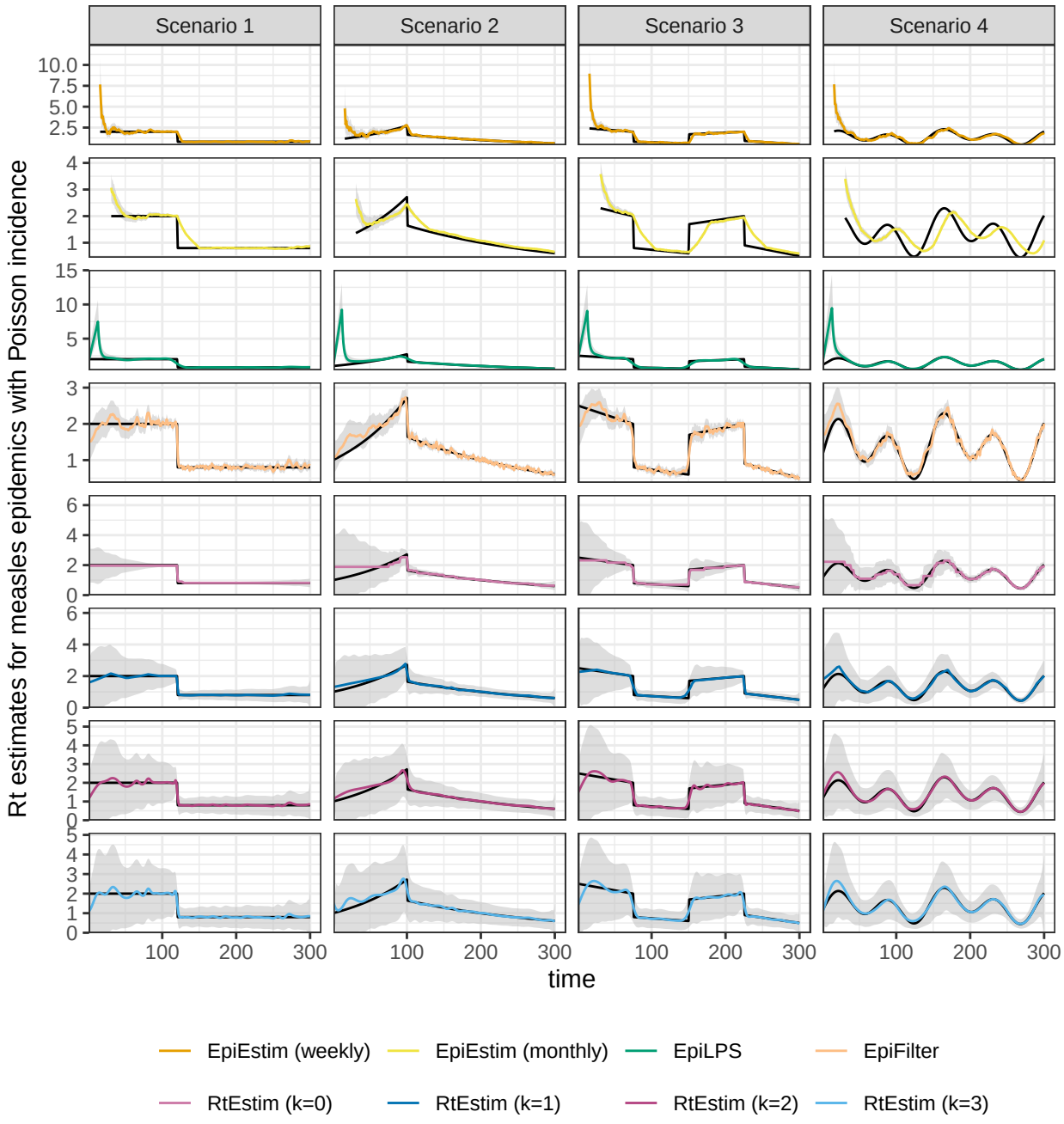

Figure A.6.1: Example measles epidemics with Poisson incidence.

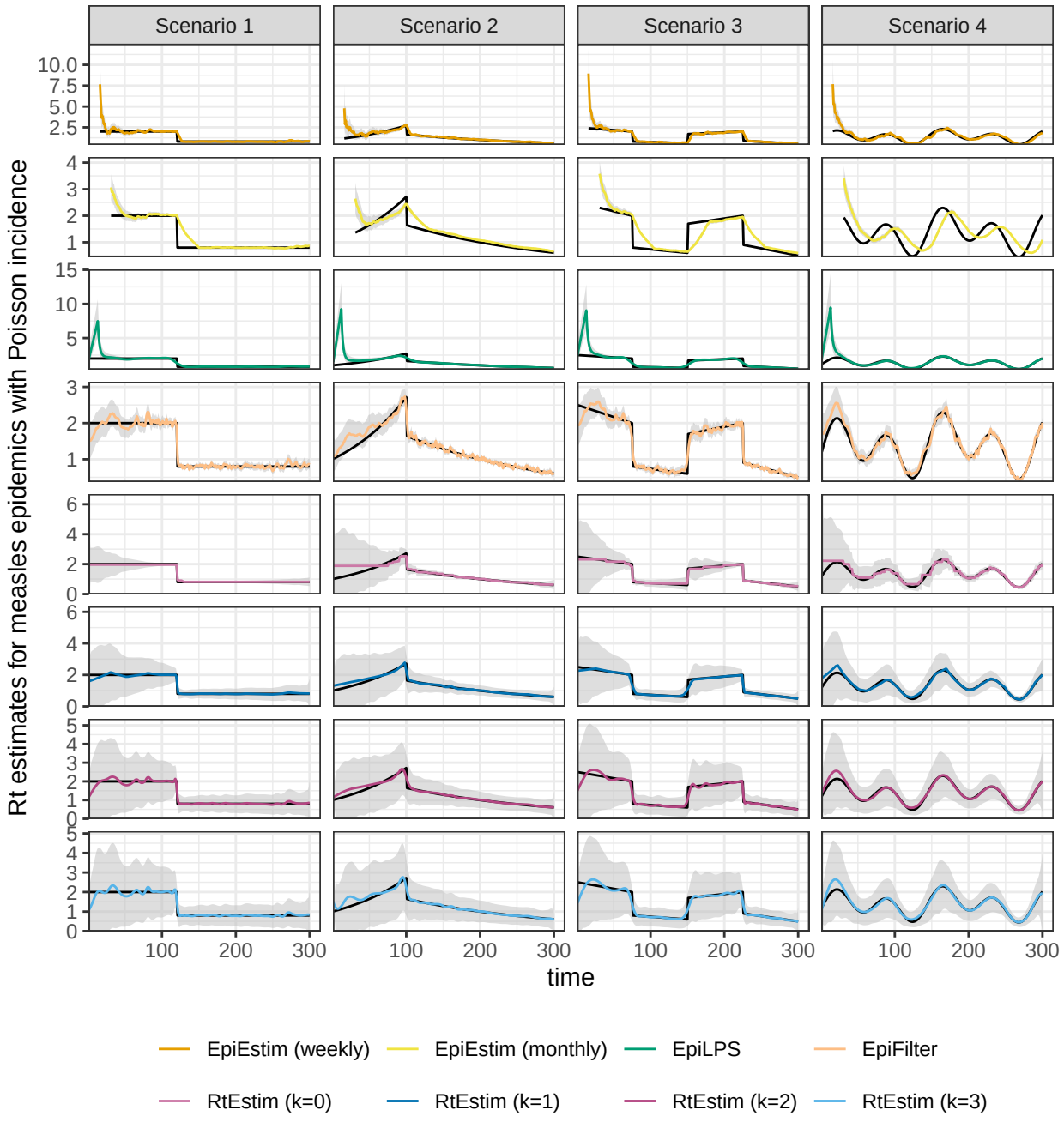

Figure A.6.2: Example measles epidemics with negative Binomial incidence.

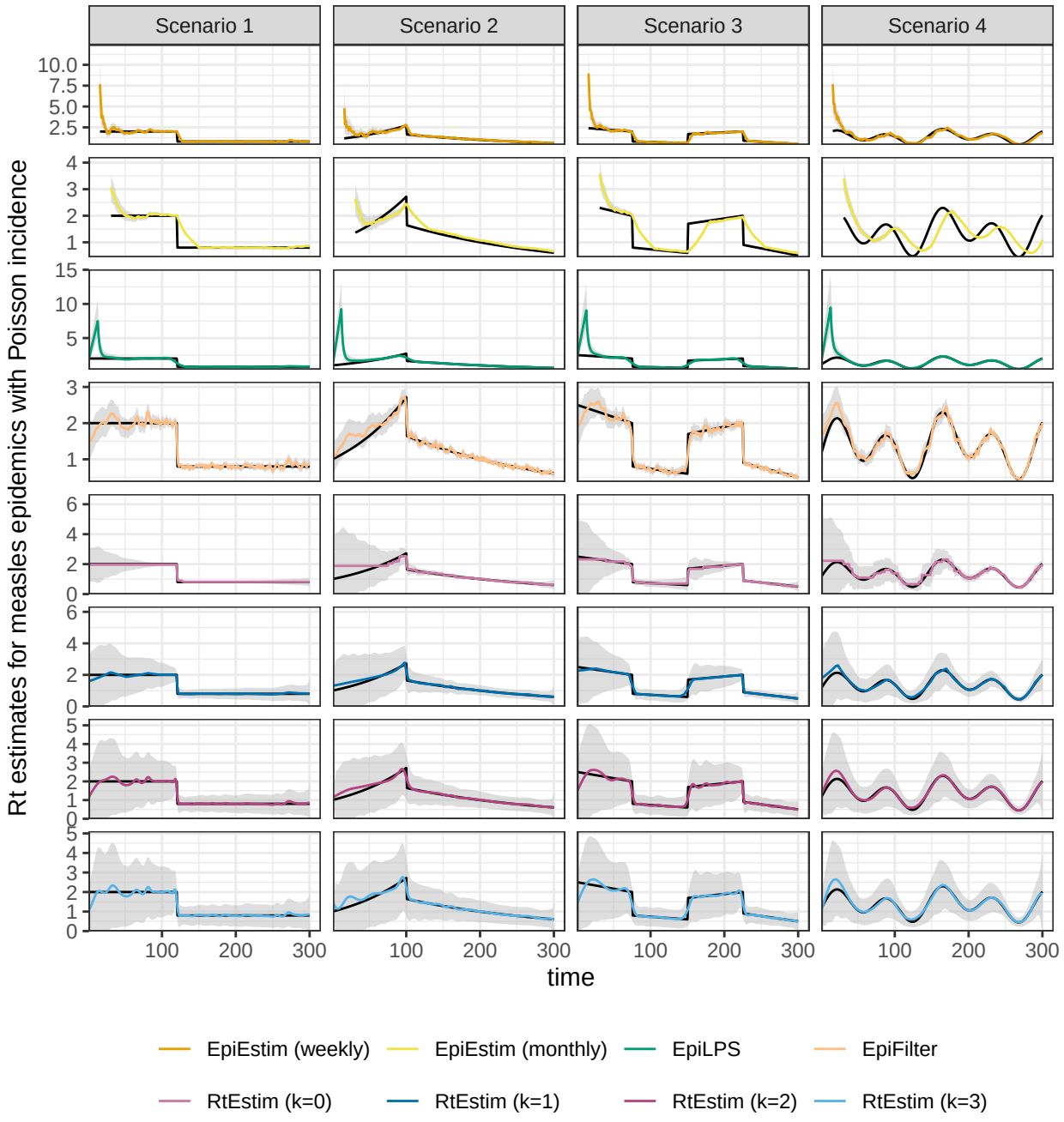

Figure A.6.3: Example SARS epidemics with Poisson incidence.

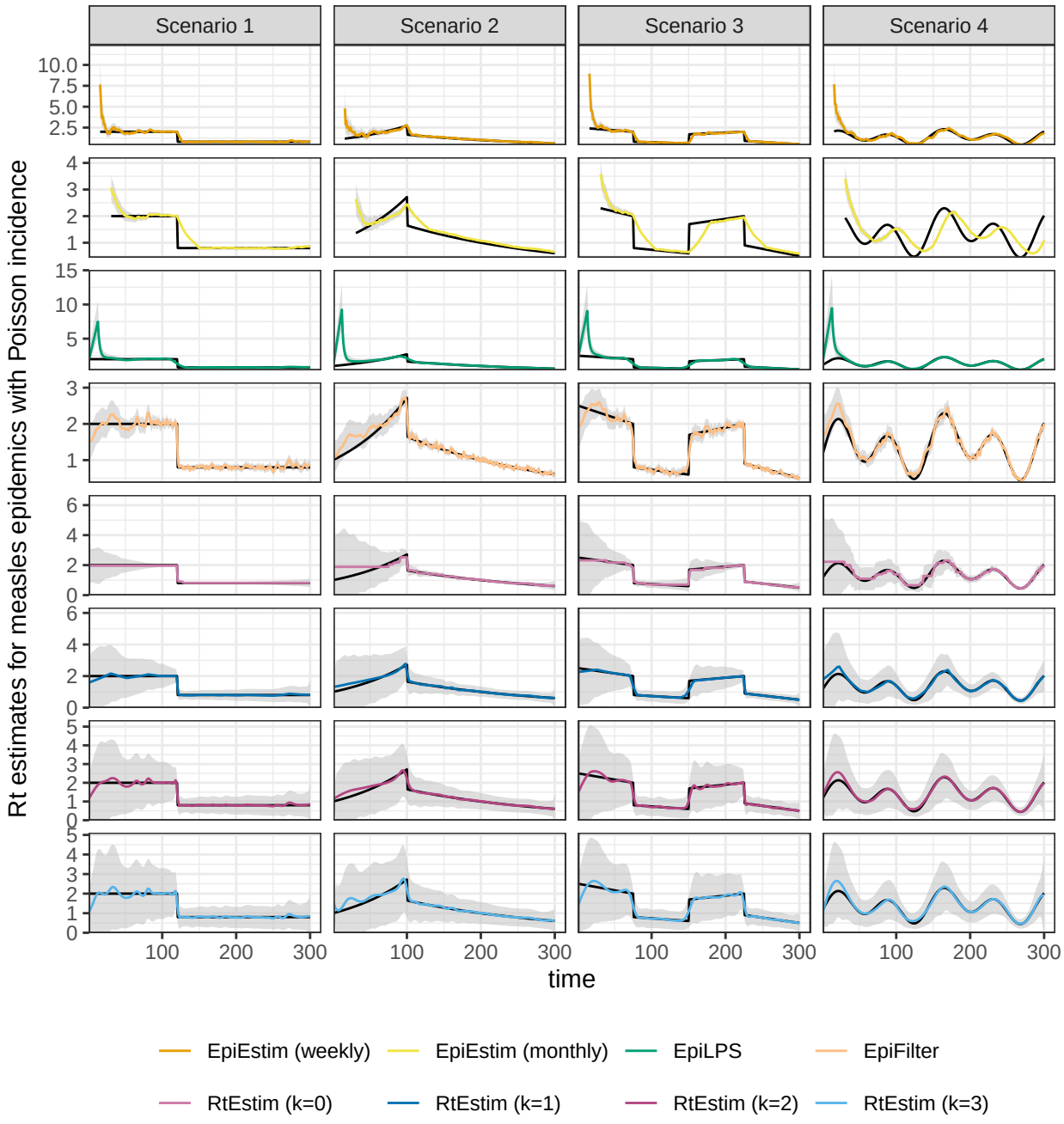

Figure A.6.4: Example SARS epidemics with negative Binomial incidence.

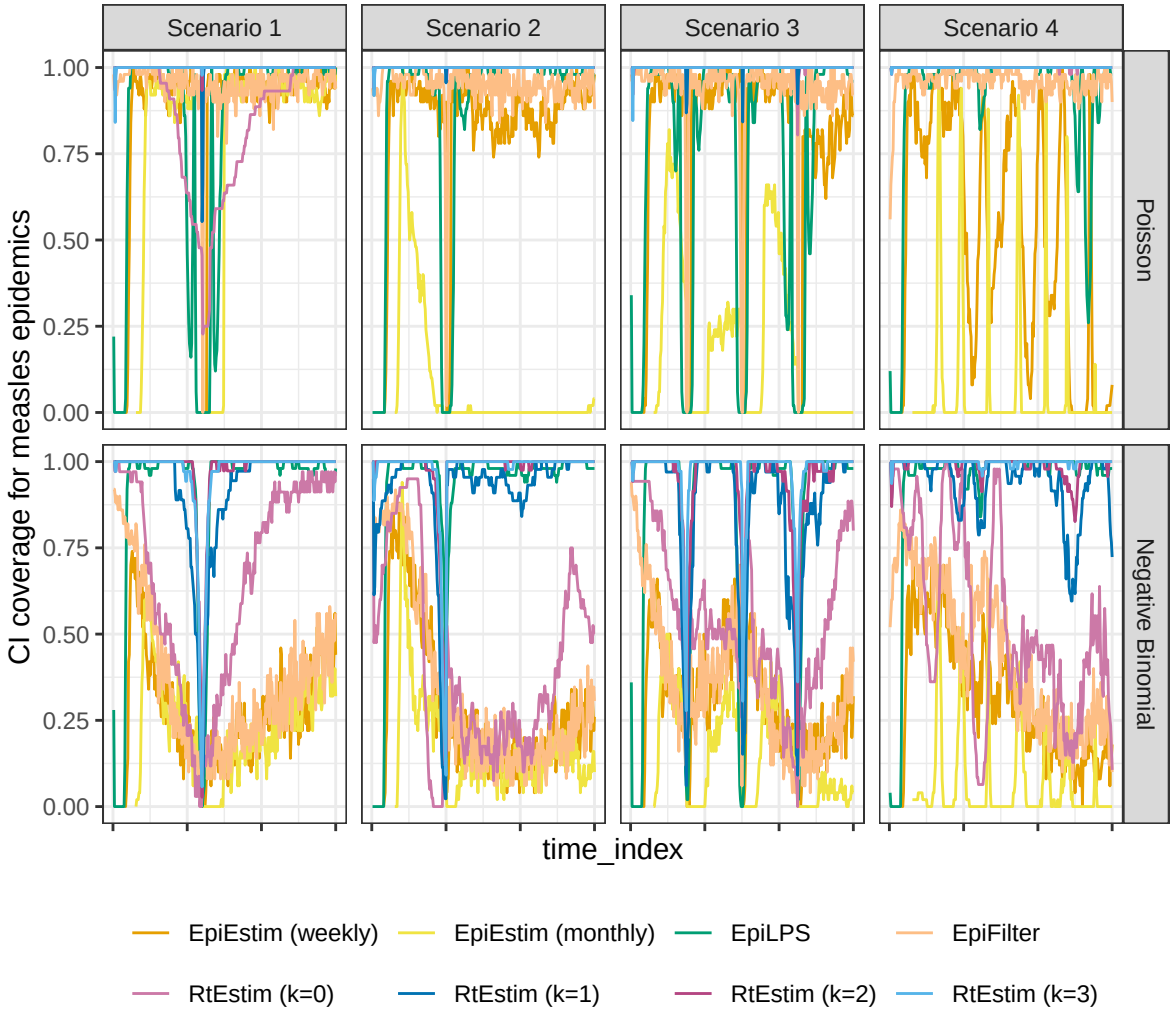

Figure A.6.5: Averaged coverage of CI per coordinate with measles epidemics.

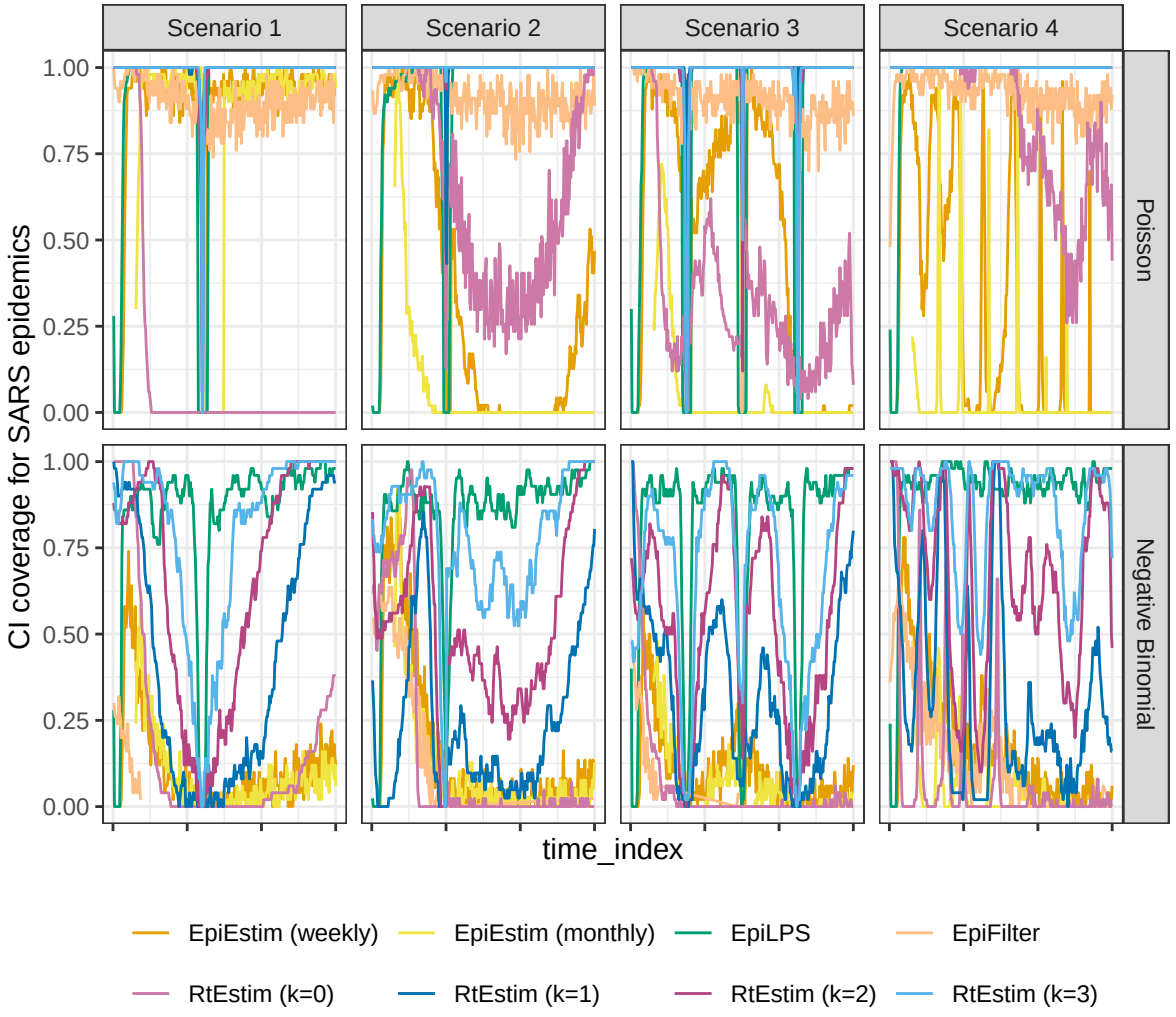

Figure A.6.6: Averaged coverage of CI per coordinate with SARS epidemics.

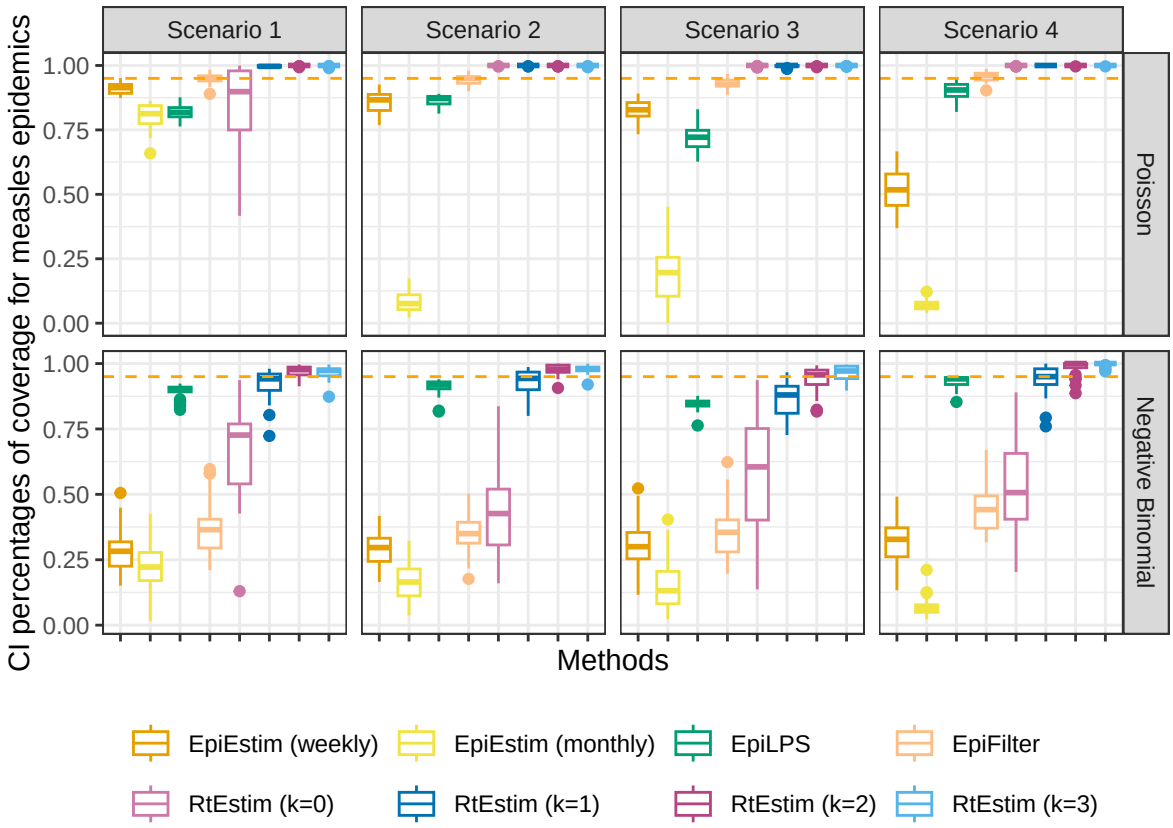

Figure A.6.7: Averaged percentages of CI coverage with measles epidemics.

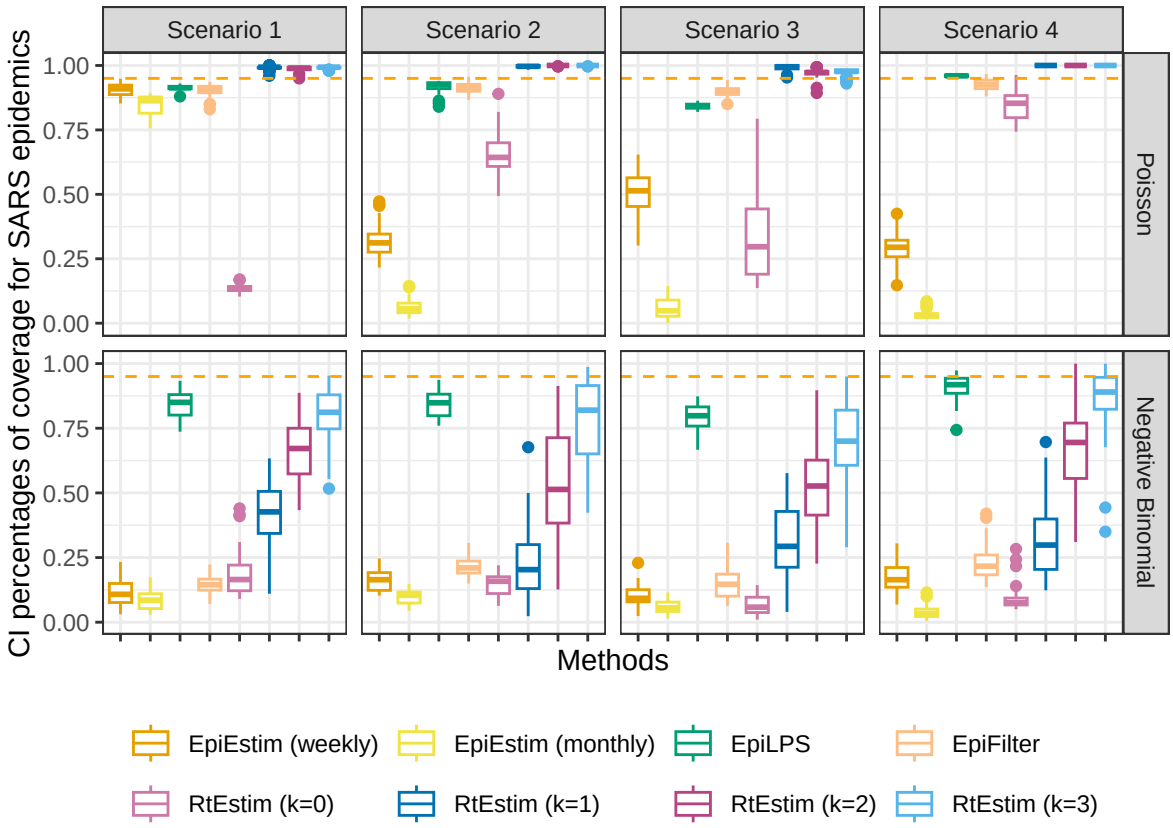

Figure A.6.8: Averaged percentages of CI coverage with SARS epidemics.

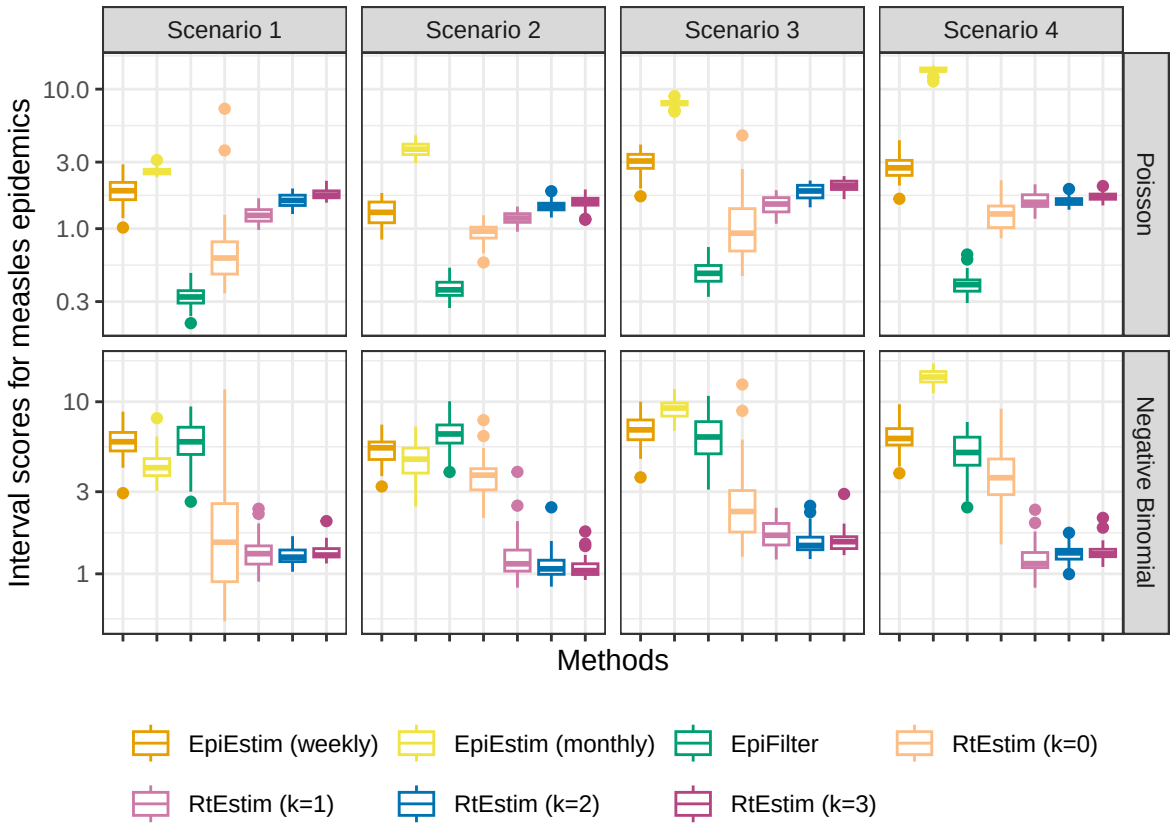

Figure A.6.9: Averaged interval scores with measles epidemics.

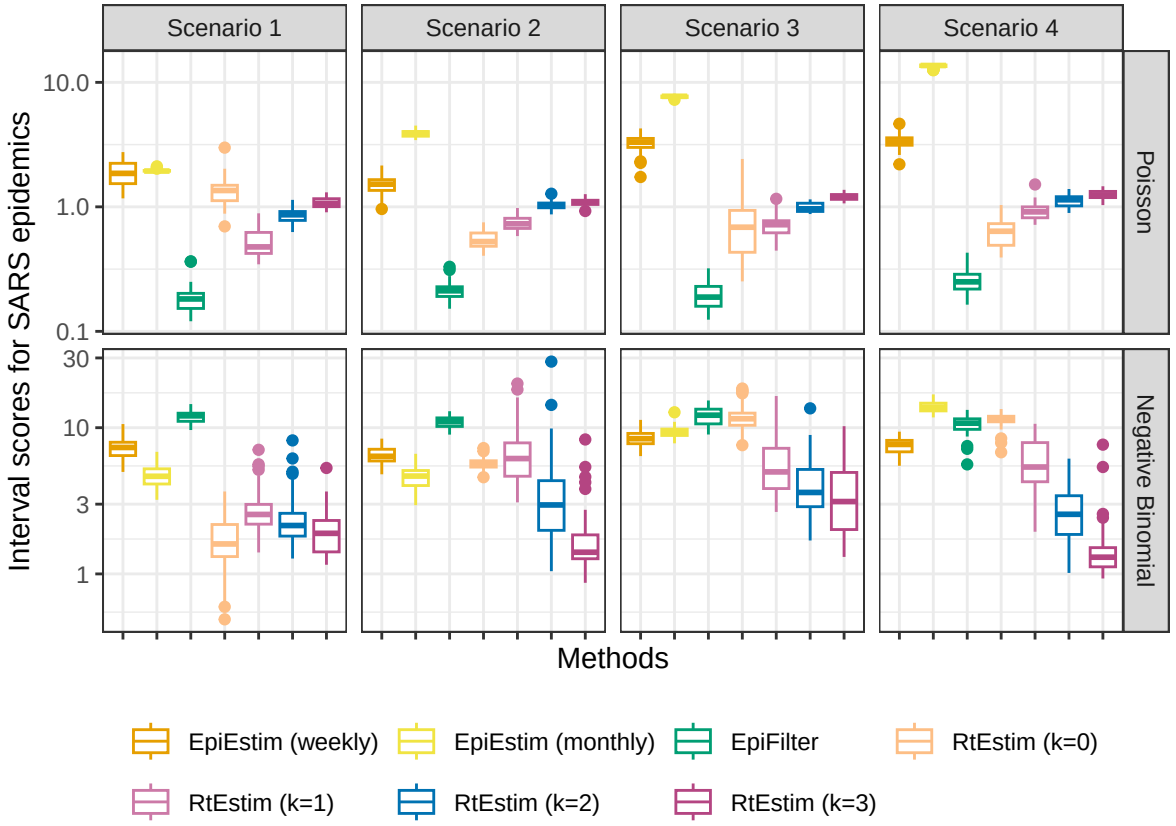

Figure A.6.10: Averaged interval scores with SARS epidemics.

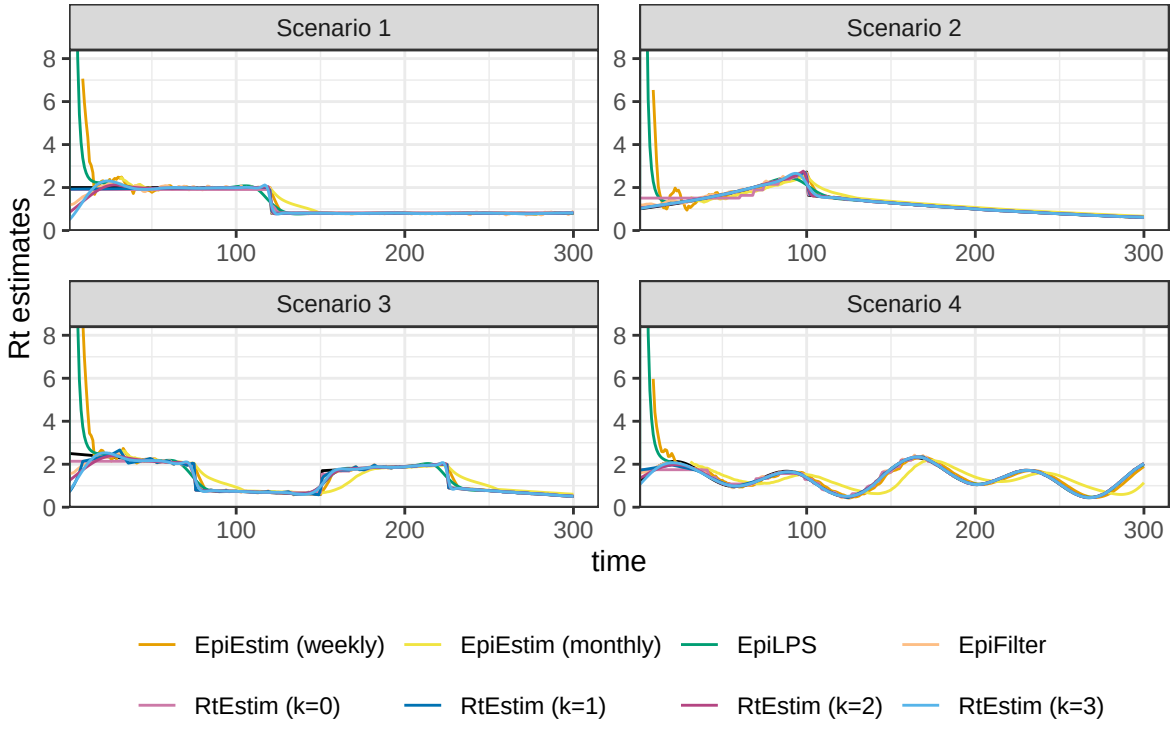

Figure A.7.1: Example of effective reproduction number estimation for SARS epidemics with Poisson observations.

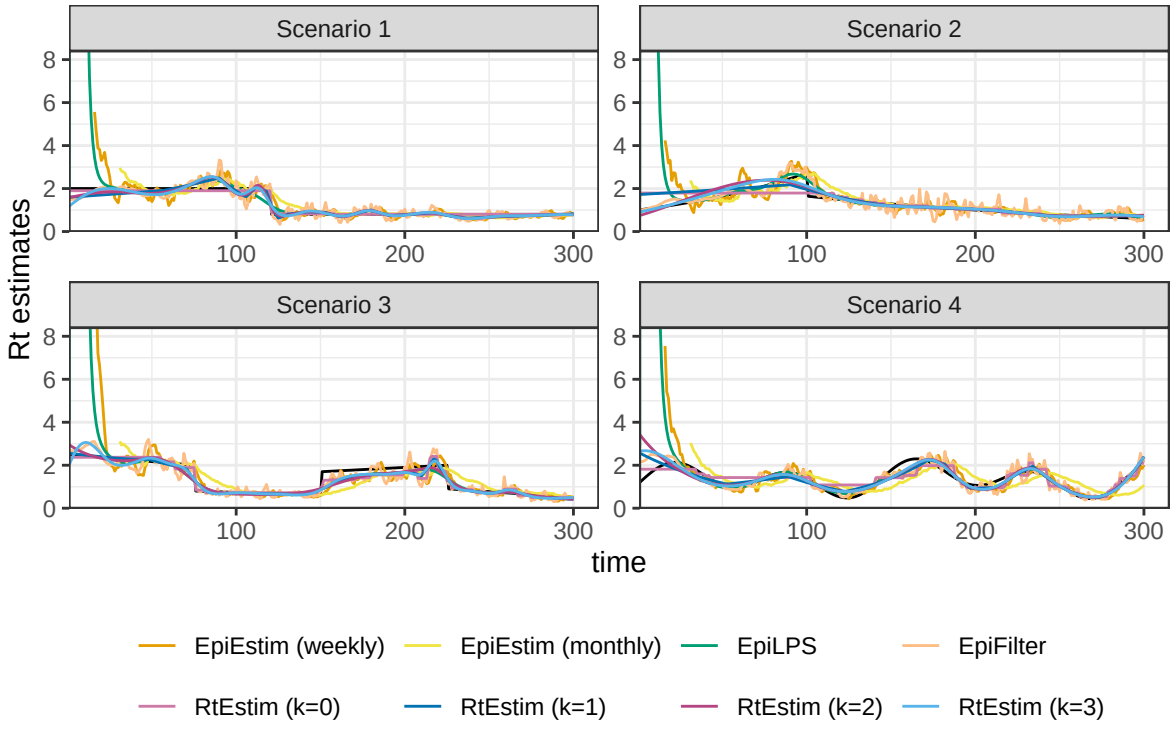

Figure A.7.2: Example of effective reproduction number estimation for measles epidemics with negative Binomial observations.

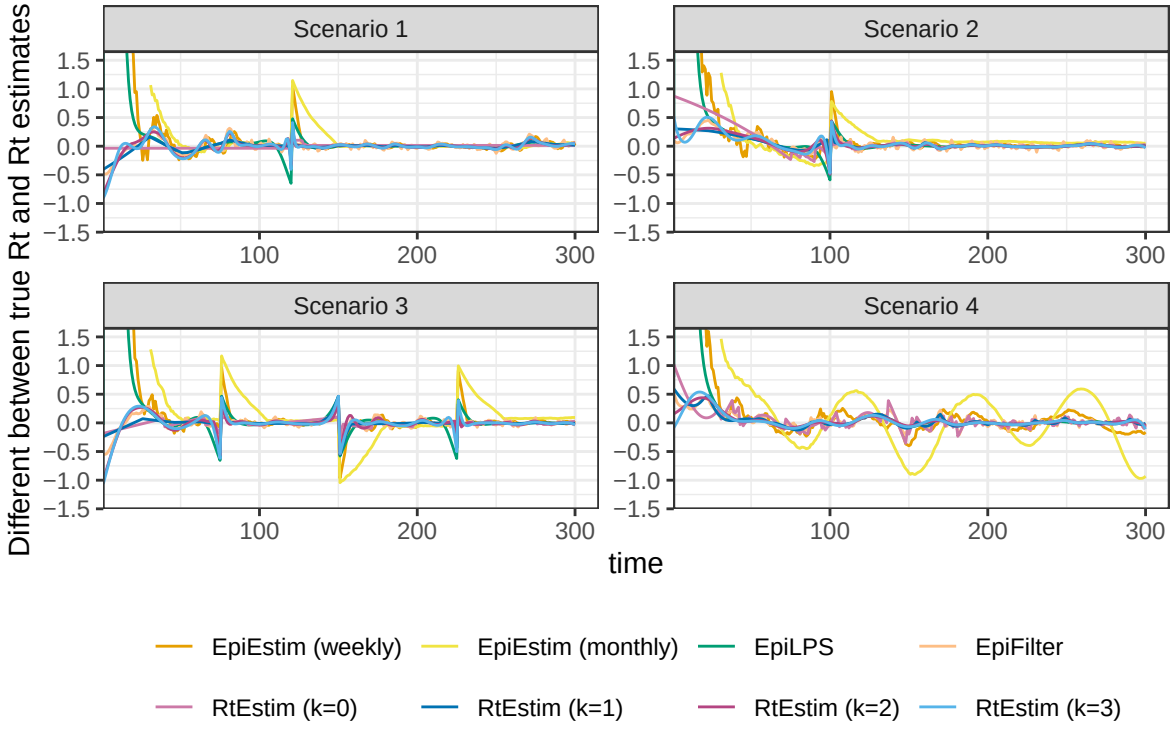

Figure A.7.3: Difference between of the true effective reproduction number and its estimation for measles epidemics with Poisson observations.

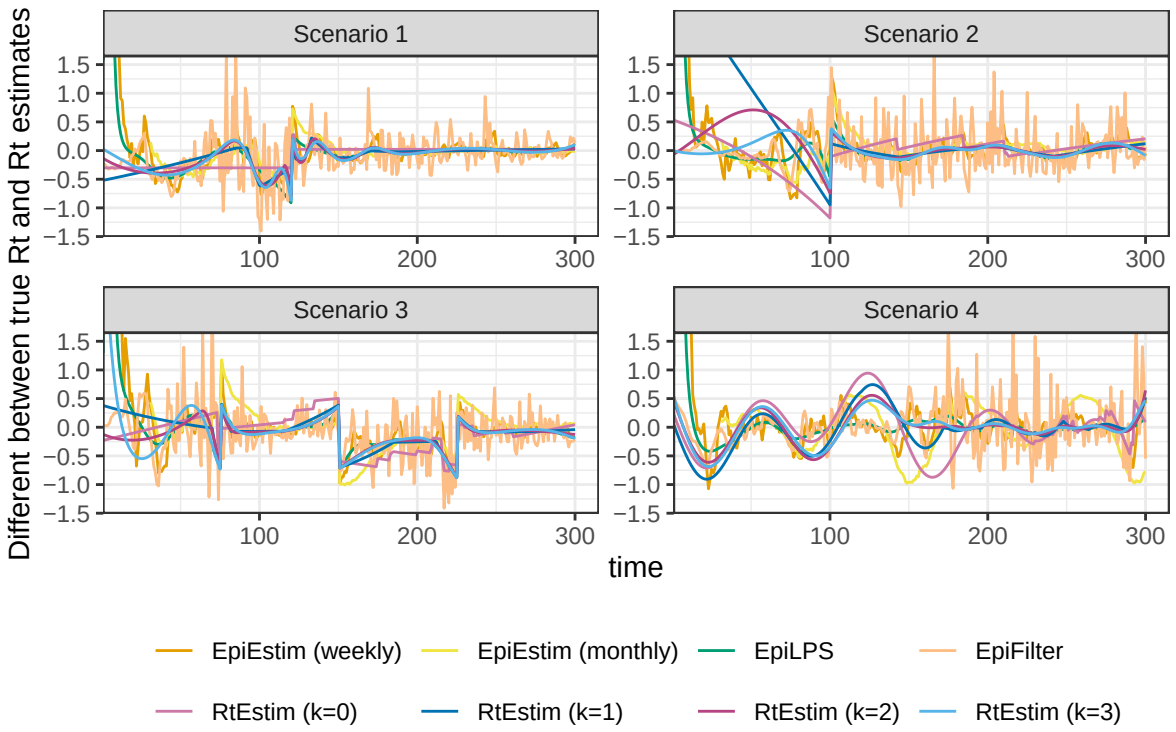

Figure A.7.4: Difference between of the true effective reproduction number and its estimation for SARS epidemics with negative Binomial observations.

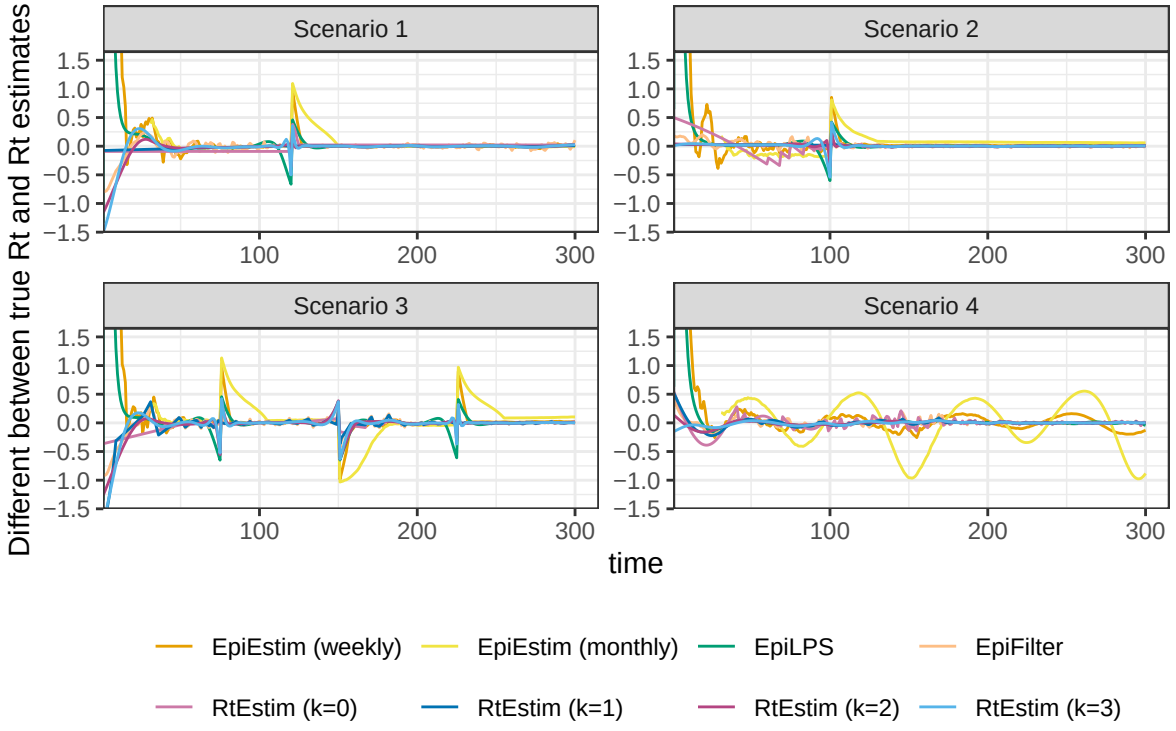

Figure A.7.5: Difference between of the true effective reproduction number and its estimation for SARS epidemics with Poisson observations.

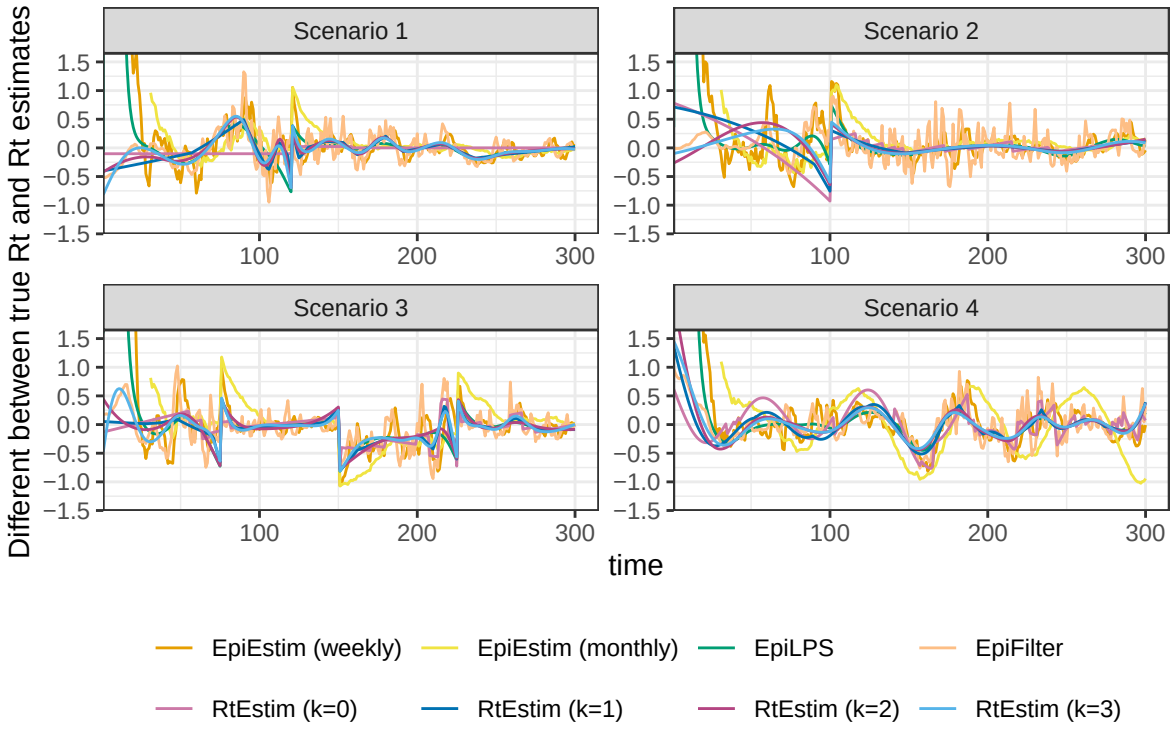

Figure A.7.6: Difference between of the true effective reproduction number and its estimation for measles epidemics with negative Binomial observations.

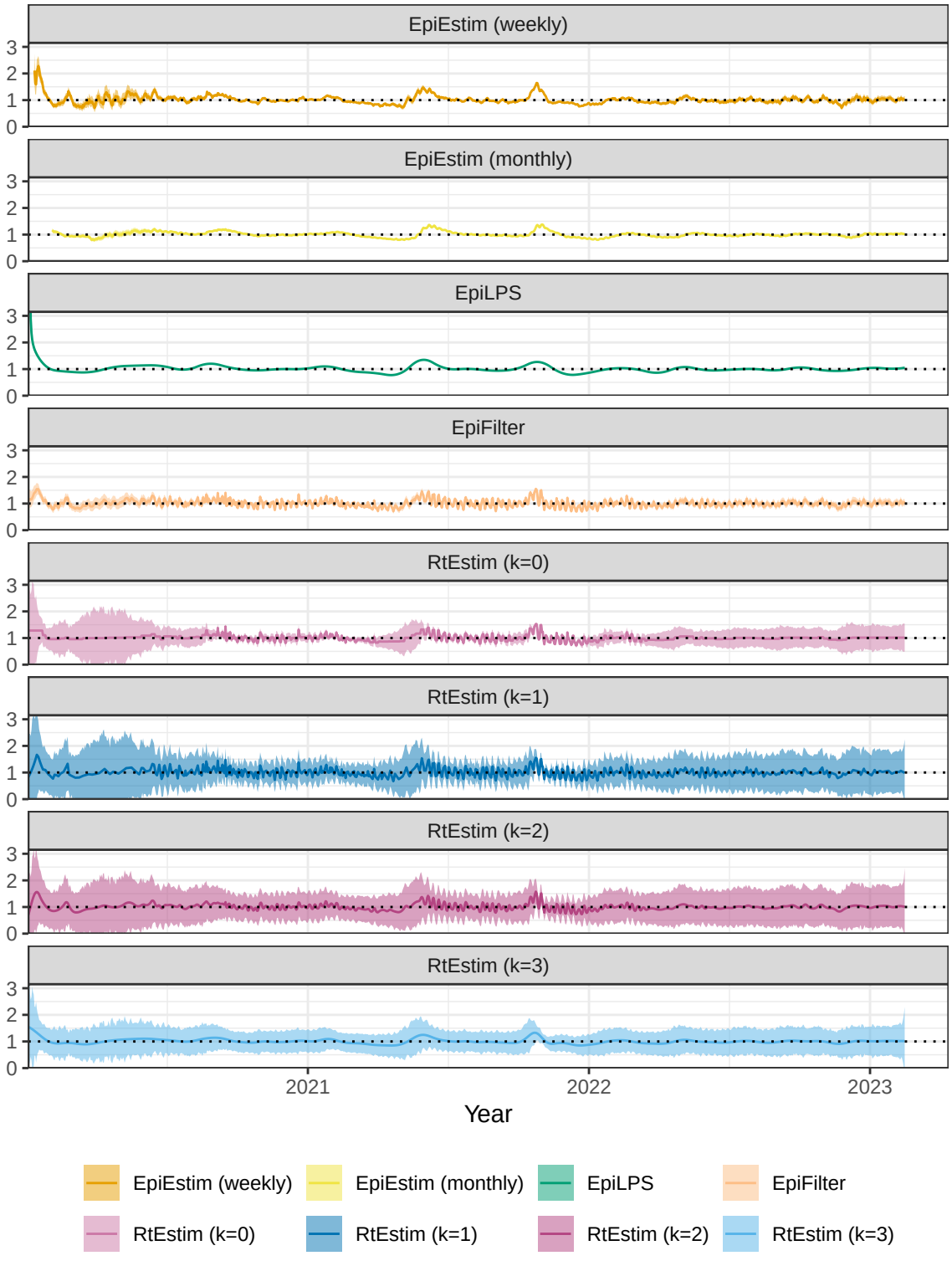

Figure A.8.1: Rt estimates with CIs for Covid19.

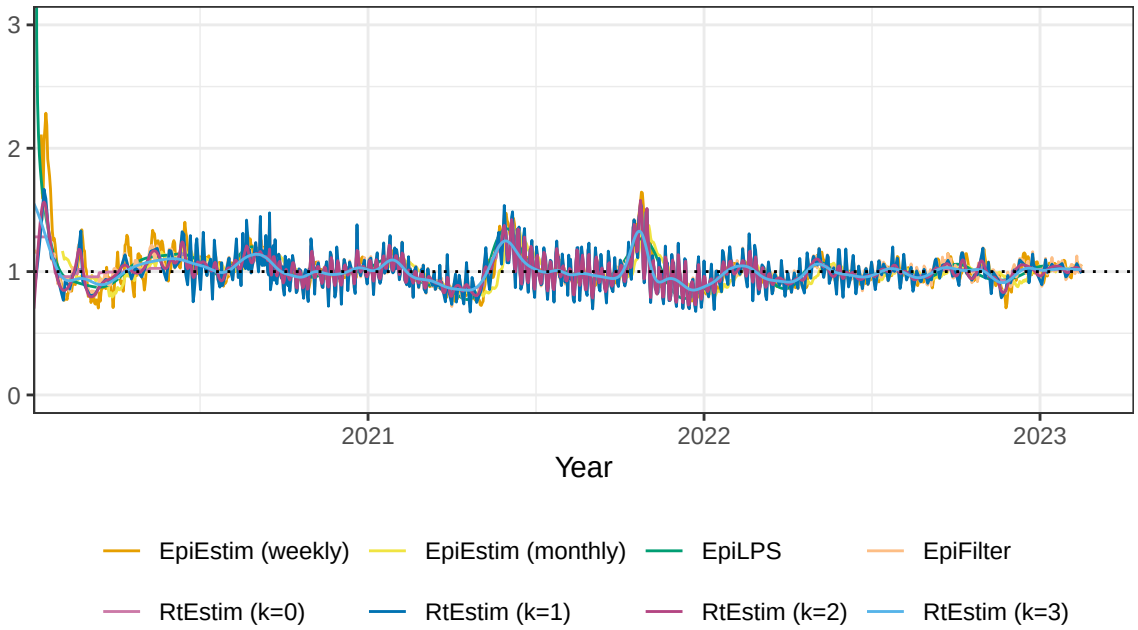

Figure A.8.2: Rt estimates for Covid19.

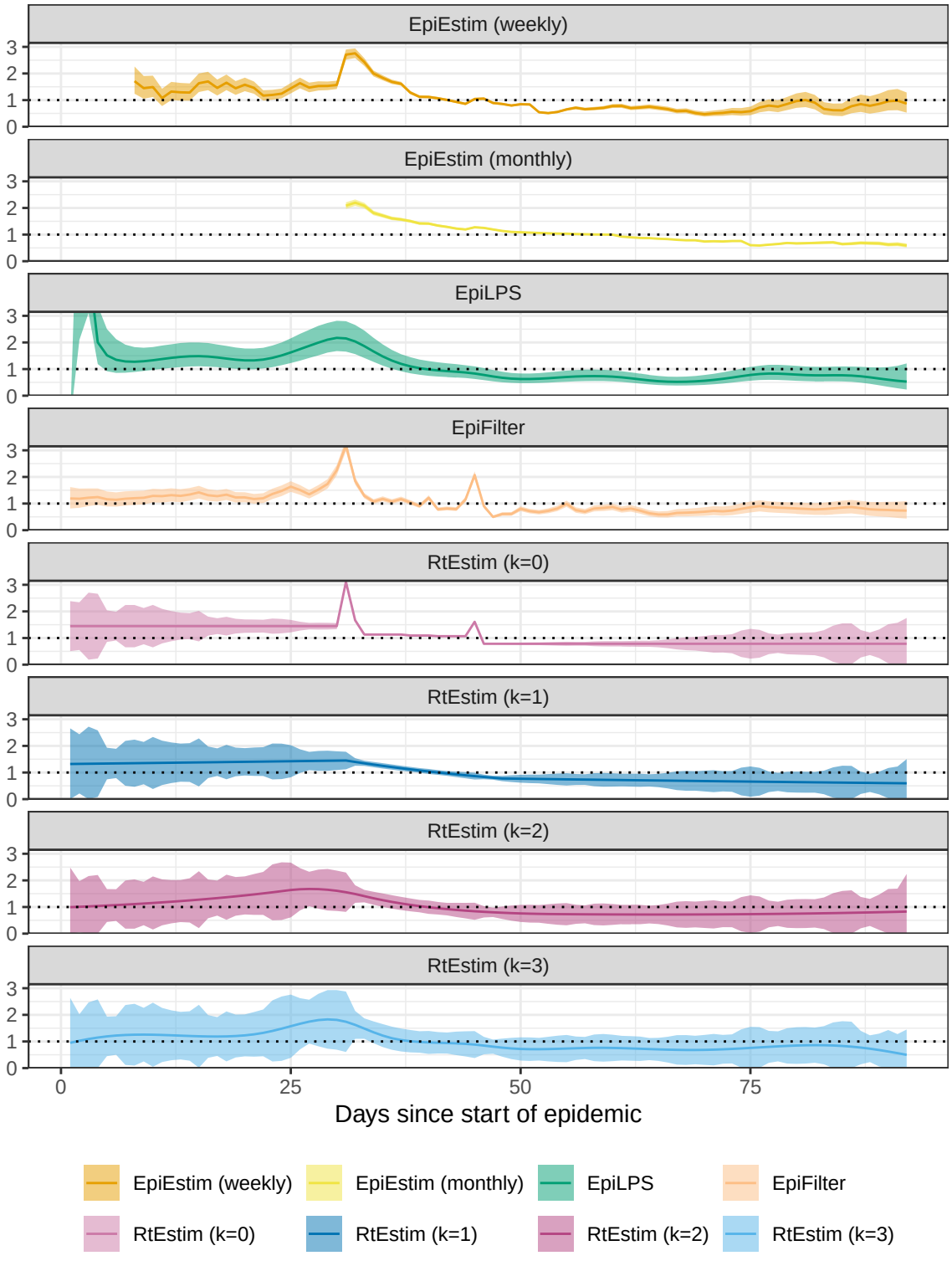

Figure A.8.3: Rt estimates with CIs for Flu 1918.

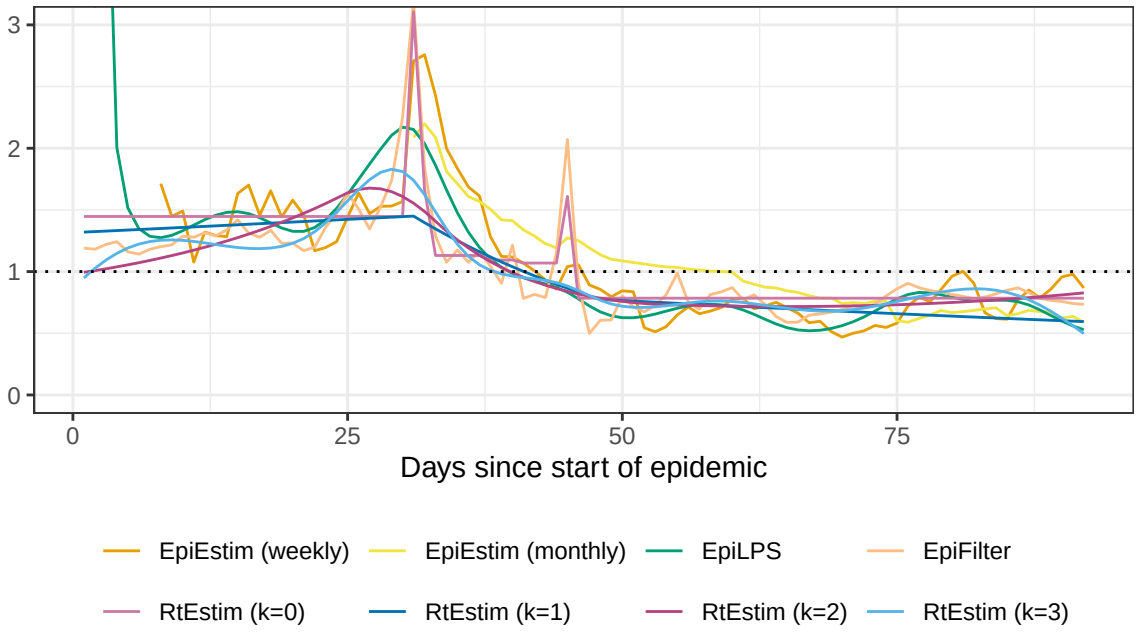

Figure A.8.4: Rt estimates for Flu 1918.