RtEstim: Effective reproduction number estimation with trend filtering

Jiaping Liu
Department of Statistics
The University of British Columbia
Vancouver, BC V6T 1Z4
jiaping.liu@stat.ubc.ca

Zhenglun Cai
Centre for Health Evaluation and Outcome Sciences
The University of British Columbia
Vancouver, BC V6T 1Z4
ecai@cheos.ubc.ca

Paul Gustafson
Department of Statistics
The University of British Columbia
Vancouver, BC V6T 1Z4
gustaf@stat.ubc.ca

Daniel J. McDonald*
Department of Statistics
The University of British Columbia
Vancouver, BC V6T 1Z4
daniel@stat.ubc.ca

October 13, 2023

Abstract

*Corresponding author.

1 Introduction

Effective reproduction numbers (also called, instantaneous reproduction numbers) are a key to understand infectious disease dynamics including the potential sizes of a pandemic, the scale of epidemic prevention measures, and the effectiveness of control effects. They are expected numbers of secondary infections caused by an infected individual in a population, and a time series that can reflect the effect of time-varying factors, such as intervention policy and population immunity. Let $\mathcal{R}_t$ be the effective reproduction number at time t. A practical interpretation is that $\mathcal{R}_t < 1$ represents a circumstance when the infection dies out gradually and achieves a *disease-free equilibrium*, whereas when $\mathcal{R}_t > 1$, the infection is always present, which leads to the *endemic equilibrium*. Effective reproduction numbers reveal an unobservable biological reality. There exist a number of models to uncover this reality relying on various domain-specific assumptions and using different types of observed data such as incidence data and death rates. Estimation of effective reproduction numbers relies heavily on the quality of the available data. Due to the limitations of data collection, such as underreporting and lack of standardization, epidemiological models, however, need to make *accurate* estimations given the poor quality of data. Since model assumptions may not be verified in practice, it is also critical for an estimator to be *robust* to model misspecification.

Many existing approaches for effective reproduction number estimation are Bayesian approaches that estimate posterior distributions of $\mathcal{R}_t$. `EpiEstim`(Cori et al., 2020), proposed by Cori et al. (2013), is a Bayesian approach solving the posterior distribution of $\mathcal{R}_t$ given incidence data prior to time t. An advantage of `EpiEstim` is that it only depends on limited assumptions (the Poisson distributed incidences, the prior distribution of effective reproduction number, and the serial interval distribution) and only requires incidence data which is easily obtainable. For these reasons, `EpiEstim` does not require much domain expertise in implementation. It is one of the earliest approaches that are both succinct and accurate in $\mathcal{R}_t$ estimation. They proposed an upgraded version `EpiEstim(2.2)` in Thompson et al. (2019), which distinguished imported cases from local transmission and simultaneously estimated the serial interval. They further extended `EpiEstim` by using "reconstructed" daily incidence data to overcome the issue when incidences were not always daily records in Nash et al. (2023). Abbott et al. (2020) proposed a Bayesian latent variable framework, `EpiNow2` (Sam Abbott et al., 2020), which uses both incidence and death counts for a precise $\mathcal{R}_t$ estimation. They further proposed a generative Bayesian model to handle missing data by imputation followed by truncation adjustment in Lison et al. (2023). Parag (2021) proposed an alternative Bayesian approach, `EpiFilter`, that is a recursive Bayesian smoother based on Kalman Filter. `EpiFilter` also solves the posteriori of $\mathcal{R}_t$ given a Gamma prior and Poisson distributed infection counts. Compared to `EpiEstim`, `EpiFilter` estimates $\mathcal{R}_t$ retrospectively

using all available incidences both prior and subsequent to time t, and provides robust estimation in low incidence cases. Gressani et al. (2022) proposed a Bayesian P-splines approach, `EpiLPS`, that assumes negative Binomial distributed incidence. Trevisin et al. (2023) proposed a Bayesian model based on particle filtering to estimate spatially explicit effective reproduction numbers. Bayesian approaches estimate the posterior distribution of the effective reproduction numbers with the advantage that the credible intervals can be easily computed. A limitation of many Bayesian approaches is that they usually require heavy computational workload, especially when data sequences are long or hierarchical structures are complex.

There are also frequentist approaches. Abry et al. (2020) proposed to regularize the smoothness of $\mathcal{R}_t$ regarding its temporal and spatial evolution. They considered a penalized regression with a second-order temporal regularization and a spatial regularization on $\mathcal{R}_t$ and with Poisson loss. They further extended it by introducing another penalty on outliers for robustness in Pascal et al. (2022). Pircalabelu (2023a) is a spline-based model relying on the assumption of exponential-family distributed incidences. Ho et al. (2023) estimates $\mathcal{R}_t$ while monitoring the time-varying level of overdispersion. There are other spline-based approaches such as Azmon et al. (2014); Gressani et al. (2021); Pircalabelu (2023b), regressive models with random effects (Jin et al., 2023) that is robust to low incidence cases, and generalized autoregresive moving average (GARMA) model (Hettinger et al., 2023) that is robust to measurement errors in incidences.

We propose a retrospective effective reproduction number estimation approach, called `RtEstim`, that requires only the daily incidence data based on the assumptions of Poisson incidences and Gamma serial interval functions. It makes our approach straightforward and depending on little expertise in domain knowledge for implementation. `RtEstim` produces accurate estimations that are empirically robust in model misspecification, i.e., the violation of distributional assumption of incidences. It is a convex optimization problem with Poisson loss and ℓ_1 penalty on the temporal evolution of $\mathcal{R}_t$, which is known as the trend filtering penalty (Kim et al., 2009; Sadhanala et al., 2022; Tibshirani, 2014). Thus, `RtEstim` is a *Poisson trend filtering* problem. Sadhanala et al. (2022) proposed trend filtering with exponential family loss on lattices. Here, we focus on Poisson loss on univariate data that is modified to integrate the serial interval functions. Our `RtEstim` generates discrete splines, and the estimated curves appear to be piecewise polynomials. The estimators have the property of local adaptivity, i.e., heterogeneous smoothness throughout the range of time, and facilitate the computational efficiency. We propose a proximal Newton method to solve the convex optimization problem. Our approach takes the advantage of convex optimization and is solved by Newton's method, which is known to converge rapidly. Moreover, the sparse structure of the divided difference matrix used in the trend filtering penalty allows further efficiency in computation.

We show empirically that our approach is more accurate than existing methods given the specific structures of the effective reproduction numbers.

The manuscript unfolds as follows. We first introduce the methodology of `RtEstim` including the usage of renewal equation, the development of Poisson trend filtering estimator, and the proximal Newton algorithm. The methodology is followed by the interpretation from an alternative Bayesian perspective. We run experiments to compare our estimator to `EpiEstim` and `EpiLPS`, which are both Bayesian competitors that are both accurate and computationally efficient. We then apply our `RtEstim` on the Covid-19 pandemic incidence in British Columbia and the 1918 influenza pandemic in the united states. More discussion on advantages and limitations of our approach and more practical considerations in the effective reproduction number estimation follows in the end.

2 Methods

2.1 Renewal model for incidence data

Effective reproduction number $\mathcal{R}_t$, the expected secondary infection by a primary infection in a population at time t, is inherently a ratio of new infections at t over the total primary infectious cases until t. Here, we assume an ideal scenario of homogenous population that individuals follow similar social behaviors, exposure risks and random mixing patterns such as having similar contact rates to each other or similar susceptibility and infectiousness. Denote the new expected infections at time t as N_t. Given an infectious period τ_t at t, we construct the total primary infectiousness as $\Lambda_t := \sum_{i=1}^{\tau_t} p_i N_{t-i}$, where p_i is the infectious probability that a secondary case is infected by a primary case which is infected i timepoints ago. The reproduction number can then be constructed based on the ratio $\mathcal{R}_t := N_t/\Lambda_t$. We assume a constant observable proportion of infections $c \in (0, 1)$, which is applied to all incidence counts and gets cancelled in the $\mathcal{R}_t$ ratio. Rearranging terms of the ratio yields the widely used renewal equation

$$N_t = \sum_{i=1}^{\tau_t} p_i \mathcal{R}_t N_{t-i} \tag{1}$$

for the analysis of transmission dynamics of infectious diseases. Other approaches, that are based on the renewal equation, include `EpiEstim` (Cori et al., 2013) and `EpiFilter` (Parag, 2021).

The sequence of probabilities $p_{1:\tau_t}$ gives the probabilities that a secondary infection at t is infected by a primary infection that is infected 1 to τ_t timepoints ago. The period between primary and secondary infections is exactly the generation time. Theoretically, $p_{1:\tau_t}$ are probabilities of the generation time in discretized, contiguous time intervals, i.e., $(0, 1], (1, 2], \ldots, (\tau_{t-1}, \tau_t]$. We assume that the infectiousness disappears beyond τ_t timepoints, so that the sequence $p_{1:\tau_t}$ has a sum of 1.

Generation time, however, is usually unobservable and tricky to estimate. We take a common strategy that approximating it by serial interval which is the period between the symptom onsets of primary and secondary infections. When the infectiousness profile after symptoms is independent of the incubation period (i.e., the period from the time of infections to the time of symptom onsets when the cases are confirmed), the serial interval is regarded identical as the generation time (Cori et al., 2013). We assume the distribution of generation time, and correspondingly serial interval, to be independent to time t, i.e., the probability p_i is only relevant to the relative time range i between primary and secondary infections. The sequence $p_{1:\tau_t}$ only depends on t by a chosen length τ_t; if τ_t is given, the probability sequence will be right-truncated at τ_t and rescaled to have a sum of 1. We assume that the serial interval follows Gamma distribution, which is demonstrated to be a reasonable choice in many existing studies, e.g., Abry et al. (2020); Cori et al. (2013); Pascal et al. (2022).

The renewal equation in Equation (1) quantifies the transmission dynamic that primary incidences result in new incidences by effective reproduction numbers given serial interval. This dynamic is straightforward and efficient in data usage, since it only depends on the observed incidence counts (which can be easily obtainable, usually sufficient and of good quality such as completeness, accuracy, and timeliness) and the specification of serial interval distribution.

2.2 Poisson trend filtering estimator

We use the daily confirmed cases y_t on day t to estimate the observed infectious cases by assuming a consistent incubation period and further assume y_t to be Poisson distributed with mean N_t, i.e.,

$$y_t \sim \text{Poisson}(N_t),$$

where $N_t = \Lambda_t \mathcal{R}_t$. Considering retrospectively a past period of n days, our interest is to estimate the Poisson parameter $\mathcal{R}_t$ given confirmed incidence counts $y_{1:n} := \{y_1, \ldots, y_n\}$ and the total infectiousness based on the confirmed cases $\Lambda_t^* := \sum_{i=1}^{\tau_t} p_i y_{t-i}$. A natural approach is to solve the maximum likelihood estimates (MLEs), i.e.,

$$\begin{aligned} \hat{\mathcal{R}}_t :&= \underset{\mathcal{R}_t \in \mathbb{R}_+}{\text{argmax}}\, \mathbb{P}(\mathcal{R}_{1:n} \mid y_{1:n}, p_{1:\tau_t}) \\ &= \underset{\mathcal{R}_t \in \mathbb{R}_+}{\text{argmax}} \prod_{t=1,\ldots,n} \frac{e^{-\mathcal{R}_t \Lambda_t^*} (\mathcal{R}_t \Lambda_t^*)^{y_t}}{y_t!}. \end{aligned} \tag{2}$$

This maximization problem, however, yields a one-to-one correspondence between the confirmed case (or the total infectiousness) and the effective reproduction number per day, i.e. $\hat{\mathcal{R}}_t = y_t / \Lambda_t^*$, so that the estimated curves have no significant graphical smoothness.

Smoothness of the effective reproduction numbers is a key to understand the trend of transmissibility of infectious diseases in retrospective studies. Smoother estimated curves give more high-level ideas with less changing points and hide minor details, and vice versa. We assume the effective reproduction numbers to appear as piecewise polynomials with multiple knots (i.e., changing points of graphical curvature) with varying degrees. We specifically consider discrete splines with various degrees of continuity. For instance, the 0th degree discrete splines are piecewise constant, the 1st degree curves are piecewise linear, and the 2nd degree curves are piecewise quadratic. For $k \geq 1$, the kth degree discrete splines are continuous and have continuous discrete differences up to degree $k-1$ at the knots.

To achieve such smoothness, we regularize the distance between adjacent effective reproduction numbers. Since $\mathcal{R}_t > 0$, penalizing the distance between $\mathcal{R}_t$s directly may cause numerical issues such that there may be negative estimates generated in computation. Therefore, we equivalently penalize the distance between natural logarithms of neighboring $\mathcal{R}_t$s through divided differences (i.e., discrete derivatives) with various orders. Compared to splines, discrete splines introduce computational efficiency without loss of numerical accuracy. We penalize ℓ_1 norm of the distance, which introduces sparsity into the curvature, so that the estimates have heterogeneous smoothness in different subregions of the entire domain. It is a more realistic setting compared to homogeneous smoothness in the squared ℓ_2 norm. The divided differences with various orders realize the temporal evolution of effective reproduction numbers with various degrees.

We define a penalized regression to solve the MLE problem with the smoothness regularization in Equation (3). It is a minimization problem with Poisson loss (which is the negative log-likelihood of Poisson distributions) to control the data fidelity and the trend filtering penalty to control the graphical smoothness (Kim et al., 2009; Tibshirani, 2014; Tibshirani et al., 2022). The problem solves a Poisson trend filtering (PTF) estimator on univariate cases. Let $\theta := \log(\mathcal{R}) \in \mathbb{R}^n$, and then $\Lambda \circ \mathcal{R} = \Lambda \circ e^{\theta}$, $\log(\Lambda \circ \mathcal{R}) = \log(\Lambda) + \theta$, where $\circ$ is elementwise product, $e^a, \log(a)$ apply to vector a elementwise. Let $w := \Lambda$ to represent weights in the objective. Define the problem with evenly spaced incidences as:

$$\hat{\theta} := \underset{\theta \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} -y_i \theta_i + w_i e^{\theta_i} + \lambda \left\| D^{(k+1)} \theta \right\|_1, \tag{3}$$

where $D^{k+1} \in \mathbb{Z}^{(n-k-1) \times n}$ is a $(k+1)$st order divided difference matrix with $k = 0, 1, 2, \ldots, n-2$, and $\hat{\mathcal{R}} := e^{\hat{\theta}}$ solves the estimated effective reproduction numbers. Define $D^{(k+1)}$ recursively as $D^{(k+1)} := D^{(1)} D^{(k)}$, where $D^{(1)} \in \{-1, 0, 1\}^{(n-k-1) \times (n-k)}$ is a banded matrix of dynamic

dimensions with diagonal band $(-1, 1)$ and off-band components 0s:

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

Define $D^{(0)} := I_n$, which is an identity matrix with size n.

The tuning parameter λ balances the contributions between data fidelity and smoothness. When $\lambda = 0$, the problem in Equation (3) reduces to the regular least squares problem. A larger tuning parameter gives a higher importance on the regularization term and yields a smoother curve until the divided differences are all zeros, i.e., all parameters are projected onto the null space of the corresponding divided difference matrix(, since the tuning parameter is large enough to make the penalty term dominate the objective).

For unevenly spaced observations, the distances between neighboring parameters vary by the time lengths between observation times, and thus, the divided differences should be adjusted by the days that the incidences are confirmed (i.e., data locations). Given the data locations $x_{1:n} = \{x_1, \ldots, x_n\}$, for $k \geq 1$, define a kth order diagonal matrix

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

Let $D^{(x,1)} := D^{(1)}$. For $k \geq 1$, define the $(k+1)$st order divided difference matrix for unevenly spaced incidences recursively as

$$D^{(x,k+1)} := D^{(1)} \cdot X^k \cdot D^{(x,k)}.$$

Our estimator is locally adaptive so that it captures the local changes such as the initiation of effective control measures. More specifically, it regularizes the similarity among reproduction numbers across a chosen number of neighboring time points and segments the curvature of the reproduction numbers such that there are more jumpiness in some subregions and more smoothness in others. Abry et al. (2020); Pascal et al. (2022) considered the second-order divided difference of effective reproduction numbers. In comparison to their studies, our estimator is more flexible in the degree of temporal evolution of the effective reproduction numbers and also avoids the potential numerical issues of penalizing/estimating positive real values.

2.3 Proximal Newton solver

The proximal Newton method is a second-order algorithm solving a proximal optimization iteratively followed by a line search algorithm adjusting the step size at each iteration for faster convergence. The proximal Newton method for Poisson trend filtering in Equation (3) solves an approximate problem iteratively — specifically, it takes a second-order Taylor expansion of the Poisson loss, which results in a proximal optimization, i.e., trend filtering with squared ℓ_2 loss, with dynamic weights during iteration, and solves it iteratively until convergence to the objective.

Let $g(\theta) := \frac{1}{n}\sum_{i=1}^{n} -y_i\theta_i + w_i e^{\theta_i}$ be the Poisson loss and $h(\theta) := \lambda \left\| D^{(k+1)}\theta \right\|_1$ be the regularization in Equation (3). At iterate $t+1$, consider the following approximation of $g(\theta)$ using the second-order Taylor expansion around θ^t,

$$g(\theta) = g(\theta^t) + (\theta - \theta^t)^\top \nabla_\theta^{(1)} g(\theta^t) + \frac{1}{2}(\theta - \theta^t)^\top \nabla_\theta^{(2)} g(\theta^t)(\theta - \theta^t),$$

where $\nabla_\theta^{(1)} g(\theta^t) = \frac{1}{n}\left(-y + w \circ e^{\theta^t}\right) \in \mathbb{R}^n$ is the gradient of $g(\theta)$ at θ^t and $\nabla_\theta^{(2)} g(\theta^t) = \frac{1}{n}\operatorname{diag}\left(w \circ e^{\theta^t}\right) \in \mathbb{R}^{n\times n}$ is the Hessian matrix of $g(\theta)$ at θ^t.

Define the proximal operator as $\operatorname{prox}_{W,D}(x) := \underset{z\in\mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2n}\|z - x\|_W^2 + \lambda \|D\theta\|_1$, where $\|a\|_W^2 := a^\top W a$. The proximal optimization problem at iterate $t+1$ can be further written as, given θ^t,

$$\begin{aligned}
\theta^{t+} :&= \underset{\theta\in\mathbb{R}^n}{\operatorname{argmin}} \; (\theta - \theta^t)^\top \nabla_\theta^{(1)} g(\theta^t) + \frac{1}{2}(\theta - \theta^t)^\top \nabla_\theta^{(2)} g(\theta^t)(\theta - \theta^t) + h(\theta), \\
&= \underset{\theta\in\mathbb{R}^n}{\operatorname{argmin}} \; \frac{1}{2n}\left\|\theta - c^t\right\|_{W^t}^2 + \lambda \left\| D^{(k+1)}\theta \right\|_1, \\
&= \operatorname{prox}_{W^t, D^{(k+1)}}(c^t),
\end{aligned} \tag{4}$$

where $W^t := \operatorname{diag}\left(w \circ e^{\theta^t}\right)$ is the weighted (Hessian) matrix multiplied by n and $c^t := \theta^t - n\left(W^t\right)^{-1}\nabla_\theta^{(1)} g(\theta^t) = y \circ w^{-1} \circ e^{-\theta^t} - \mathbf{1} + \theta^t \circ w^{-1}$, where $\{e^{\theta^t}\}_{i\in[n]} > 0, [n] := 1, 2, \ldots, n$. This is just univariate trend filtering with weights W^t (Tibshirani, 2014).

We solve the trend filtering problem in Equation (4) using the specialized ADMM, proposed by Ramdas and Tibshirani (2016), with the primal θ step solved in closed-form and the auxiliary step solved by the dynamic programming algorithm for fused lasso proposed by Johnson (2013). Let the auxiliary variable $z := D^{(k)}\theta$. The scaled augmented Lagrangian is

$$\mathcal{L}_{\lambda,\rho}(\theta, z, u) = \frac{1}{2n}\left\|\theta - c^t\right\|_{W^t}^2 + \lambda \left\| D^{(1)} z \right\|_1 + \frac{\rho}{2}\left\| D^{(k)}\theta - z + u \right\|^2 - \frac{\rho}{2}\|u\|^2,$$

where ρ is a scaled dual parameter and u is a dual variable. At Newton's iteration $t+1$, the specialized ADMM solves the following subproblems, at ADMM iteration $l+1$:

$$\begin{aligned}
\theta^{l+1} &:= \underset{\theta}{\operatorname{argmin}} \frac{1}{2n} \left\| \theta - c^t \right\|_{W^t}^2 + \frac{\rho}{2} \left\| D^{(k+1)}\theta - z^l + u^l \right\|_2^2, \\
z^{l+1} &:= \underset{z}{\operatorname{argmin}} \frac{\lambda}{\rho} \left\| D^{(1)} z \right\|_1 + \frac{1}{2} \left\| D^{(k+1)}\theta^{l+1} - z + u^l \right\|_2^2, \\
u^{l+1} &\leftarrow u^l + D^{(k+1)}\theta^{l+1} - z^{l+1}.
\end{aligned} \tag{5}$$

We further adjust the step size $\gamma^{t+1} \in (0, 1]$ at iterate $t+1$ by a backtracking line search algorithm to solve for θ^{t+1}, i.e.,

$$\theta^{t+1} \leftarrow \theta^t + \gamma^{t+1}(\theta^{t+} - \theta^t).$$

The proximal Newton algorithm iterates until convergence of the objective.

2.4 Bayesian perspective

Our approach can be interpreted as a state-space model of Poisson observational noises and Laplace transition noises with certain degree $k \geq 0$, e.g., $\theta_{t+1} = 2\theta_t - \theta_{t-1} + \varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim \text{Laplace}(0, 1/\lambda)$ for $k = 1$. Compared to EpiFilter (Parag, 2021), another retrospective study of $\mathcal{R}_t$, we share same observational assumptions, but our approach has a different transition noises. EpiFilter estimates the posterior distribution of $\mathcal{R}_t$, and thus it can provide the credible interval estimation with various credible levels. Our approach solves the point estimation using optimization problem, which has the advantage of computational efficiency.

3 Results

Implementation of the our approach is provided in the `R` package `rtestim`.

3.1 Experimental settings

We consider four scenarios of the time-varying effective reproduction numbers to simulate different epidemics. The first two scenarios are simple cases that are rapidly controlled by intervention, where the graphical curves consist of one knot and two segments. Scenario 1 is instantaneous prior and post-intervention, and Scenario 2 is exponentially grow and decay. The last two scenarios are more complicated, where more waves in the epidemics are involved. Scenario 3 has four linear segments with three knots, which reflect the effect of intervention, the resurgence to large epidemics, and the suppression of pandemic respectively. Scenario 4 involves more complicated waves and curvatures of the epidemic. Effective reproduction numbers across all scenarios are evenly spaced. The first three scenarios and the last scenario are motivated by Gressani et al. (2022); Parag (2021)

respectively. We name the four scenarios as (1) 2-segment constant line, (2) 2-segment exponential curve, (3) 4-segment linear line, and (4) periodic curve respectively.

We consider epidemics of length $n = 300$. Specifically, in Scenario 1, $\mathcal{R}_t = 2, 0.8$ before and after $t = 70$. In Scenario 2, $\mathcal{R}_t$ increases and decreases exponentially with rates $0.015, 0.005$ pre and post $t = 50$. In Scenario 3, $\mathcal{R}_t$ reduces from 2.5 to 2 linearly between $t \in [1, 60]$, falls to 0.8 at $t = 61$ and goes linearly down to 0.6 until $t = 110$, resurges to 1.7 at $t = 111$ and grows linearly back to 2 until $t = 150$, and then drops to 0.9 at $t = 151$ and descends to 0.5 until the end. In Scenario 4, $\mathcal{R}_t$ is a continuous, periodic curve generated by the function $f(x) = 0.2\left(\left(\sin(\frac{\pi x}{12}) + 1\right) + \left(2\sin\left(\frac{\pi x}{6}\right) + 2\right) + \left(3\sin(\frac{\pi x}{1.2}) + 3\right)\right)$ at equally spaced points $x \in [0, 10]$.

We assume that the serial interval follows Gamma distribution with fixed shapes and scales $(3, 3)$, $(2.5, 2.5)$, $(3.5, 3.5)$ and $(3.5, 3.5)$ for Scenarios $1 - 4$ respectively. We consider all epidemics starting from $N_1 = 2$ incidences and generating until timepoints $t = 300$. We compute the expected incidence N_t use renewal equation, and generate the incidence samples from the Poisson distribution $y_t \sim Pois(N_t)$. To verify the performance of our model under the violation of distributional assumption of incidence, we generate incidence samples using negative Binomial distribution with dispersion size 5, i.e., $y_t \sim NB(N_t, size = 5)$. We generate 50 random samples for each setting of experiments. It results in 4 $\mathcal{R}_t$ scenarios $\times 2$ Incidences distributions $\times 50$ random samples $= 400$ problems in total. Examples of each effective reproduction number scenario with corresponding Poisson and negative Binomial incidences are displayed in Figure 1.

We compare our `RtEstim` to `EpiEstim` and `EpiLPS`. `EpiEstim` is a widely used Bayesian method that estimates the posterior distribution of effective reproduction numbers given the Gamma prior and Poisson distributed incidences. They estimate the reproduction number over a sliding window by assuming the reproduction number is constant during the specific time window. A longer sliding window averages out more fluctuations and noises, and leads to smoother estimation over time; whereas, a shorter sliding window is more responsive to sudden spikes or declines in a shorter period. We use the default weekly sliding window for the experimental study. Monthly sliding window is also considered in simulation. Since neither of the two sliding windows considerably outperforms the other across all scenarios, we defer the estimation of monthly sliding window to the supplementary document. `EpiLPS` is another Bayesian approach that estimates P-splines coupled with Laplace approximations of the conditional posterior of the spline vector based on negative Binomial distributed incidences. We tune the model over the candidate set of size 50 using cross validation. For four scenarios, we estimate piecewise constant $k = 0$, piecewise cubic $k = 3$, piecewise linear $k = 1$ and piecewise cubic polynomials respectively. We assume the serial intervals are known and use same serial intervals across all models for each problem.

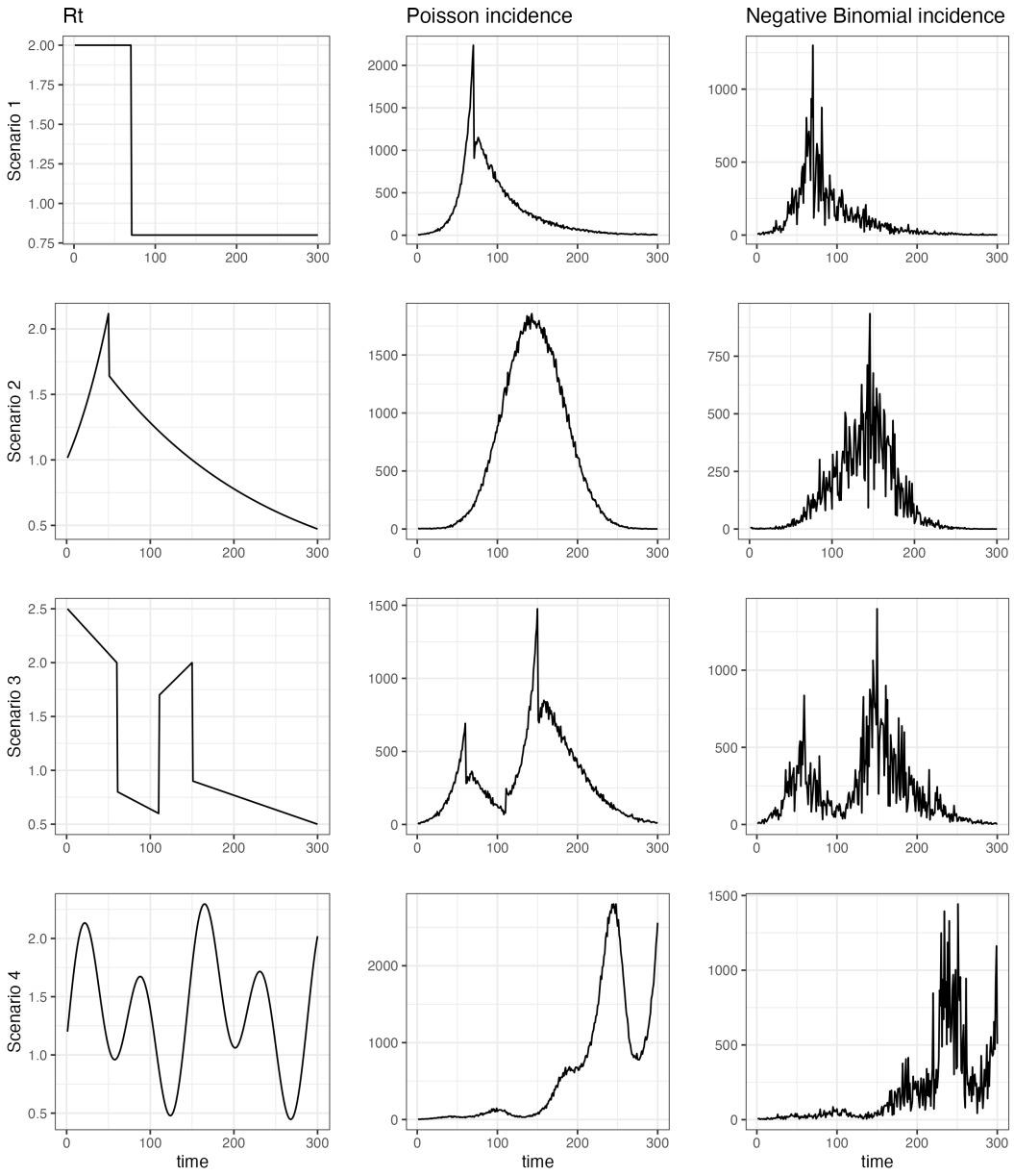

Figure 1: An example of effective reproduction numbers and corresponding incidences following Poisson or negative Binomial distribution. Three columns illustrate four $\mathcal{R}_t$ cases, Poisson incidences, and negative Binomial distributed incidences for each $\mathcal{R}_t$ case respectively. Four rows correspond to four $\mathcal{R}_t$ cases respectively.

The accuracy of $\mathcal{R}_t$ estimates is measured by the mean Kullback-Leibler (KL) divergence for Poisson distributions

$$\frac{1}{n} D_{KL}(\hat{\mathcal{R}}||\mathcal{R}) = \frac{1}{n} \sum_{t=1}^{n} \hat{\mathcal{R}}_t \log\left(\frac{\hat{\mathcal{R}}_t}{\mathcal{R}_t}\right) + \mathcal{R}_t - \hat{\mathcal{R}}_t,$$

where $\mathcal{R} := \{\mathcal{R}_t\}_{t=1}^n$. In comparison of the accuracy across methods, we drop the estimates during the first week as the $\mathcal{R}_t$ estimates of `EpiEstim` starts at $t = 8$. We avoid using the Euclidean (ℓ_2) norm in favor of KL divergence because Poisson distribution corresponds to points on a discrete lattice, which can be regarded as lying on a curved (non-Euclidean) space in the context of geometry. Other details of the experimental settings are deferred to the supplementary document.

We run leave-third-out cross validation (CV) to choose the best tuning parameter from the candidate set of size 50, i.e., $\boldsymbol{\lambda} = \{\lambda_1, \cdots, \lambda_{50}\}$. Specifically, we divide the all samples into three folds and build models on each sample set which excludes one fold of the samples across all hyperparameters. Every third samples are placed into the same fold by excluding the first and last samples. We select the tuning parameter that gives the lowest averaged mean squared errors (MSEs) of the estimated reproduction numbers from the observed samples across all folds. The experiments are run in `R` with version 4.3.1 on a MacBook with an Apple M1 Pro chip and RAM 32GB running under macOS Sonoma 14.0. The `R` packages are of versions `EpiEstim_2.2-4`, `EpiLPS_1.2.0` and `rtestim_0.0.3`.

3.2 Experimental results

Figure 2 illustrates the estimated reproduction numbers by three models for the Poisson incidence cases. Compared to `EpiEstim` and `EpiLPS`, which have an edge problem at the beginning of the time series, our `RtEstim` estimates are more accurate — almost overlap with the true values — without suffering from the edge problem. Scenario 2 is a difficult problem for all methods; the immediate drop from the end of the exponential growth to the start of the exponential decay is hard to capture for all models. Since we fit a cubic Poisson trend filtering problem for Scenario 2, our estimated $\hat{\mathcal{R}}_t$ curve is continuous at the knot, which hinders the estimates from fitting the steep decline. 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 at the knot, which results in divergence from the true values in the first segment in Scenario 1. Since the piecewise constant `RtEstim` estimator does not require the smoothness in $\mathcal{R}_t$, it captures the sharp decrease in Scenario 1.

Under the violation of distributional assumption of incidences, we estimate $\mathcal{R}_t$s using negative Binomial incidences. Figure 3 displays the estimates across all methods. `RtEstim` estimates overall do not perform as remarkably accurate as in the Poisson incidence cases. Especially for Scenario 4, `RtEstim` fails to recover the wiggly curvature during the first few waves. In Scenario 2, `RtEstim` succeeds to capture the knot, but suffers from the same problem as in the Poisson cases. In Scenario 3, piecewise linear `RtEstim` estimates fail to capture the knots and do not fit well during the first half of the time period.

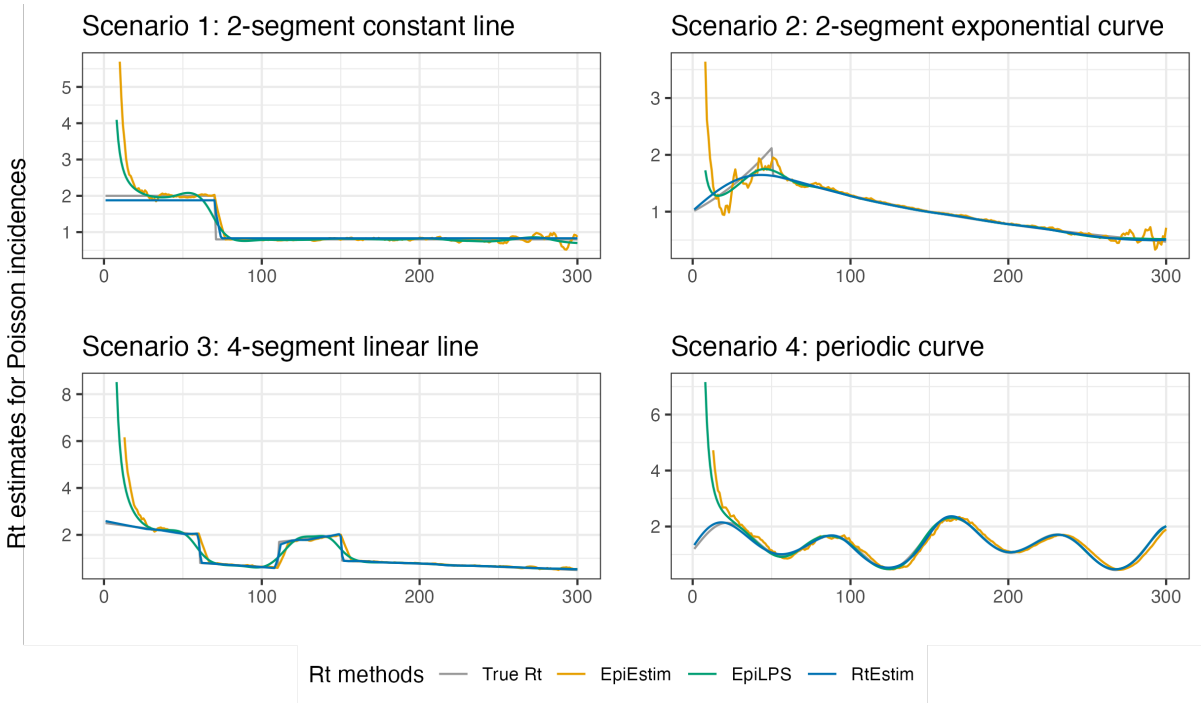

Figure 2: Effective reproduction number estimation for Poisson incidences.

`RtEstim` overall outperforms `EpiEstim` and `EpiLPS`in the experimental study. Figure 4 visualizes the KL divergences across three models in boxplots. In Scenario 2, the KL divergence boxes of `EpiLPS` is slightly lower than `RtEstim`'s boxes, but `EpiLPS` has large outliers for both Poisson and negative Binomial incidences cases. In general, given the Poisson incidences, `RtEstim` is more accurate than `EpiEstim` and `EpiLPS` across Scenarios 1, 3, 4, and slightly less accurate but more stable than `EpiLPS` in Scenario 2. Given negative Binomial incidences, `RtEstim` is still the most accurate in Scenarios 1 and 3, and achieves similar levels of accuracy as `EpiLPS` (which is based on the negative Binomial distributional assumption on the incidences) in Scenarios 2 and 4. `RtEstim` has larger outliers than `EpiLPS` for Scenarios 1, 3, and 4 given negative Binomial incidences, but the difference is no more than 0.1.

We also compare the running times of three models across 8 experimental settings. We find that all models across all experiments takes less than 3 seconds to converge. Our model runs longer generally, which is likely due to a relatively large candidate set (of size 50), other models only run a single time for a fixed set of hyperparameters per experiment. More experimental results on time comparisons are deferred to the supplementary document.

3.3 Covid-19 incidences in British Columbia

We implement our `RtEstim` on the Covid-19 confirmed cases in British Columbia (B.C.) as of May 18, 2023 (visualized in Figure 5) reported by B.C. Centre for Disease Control. We choose the

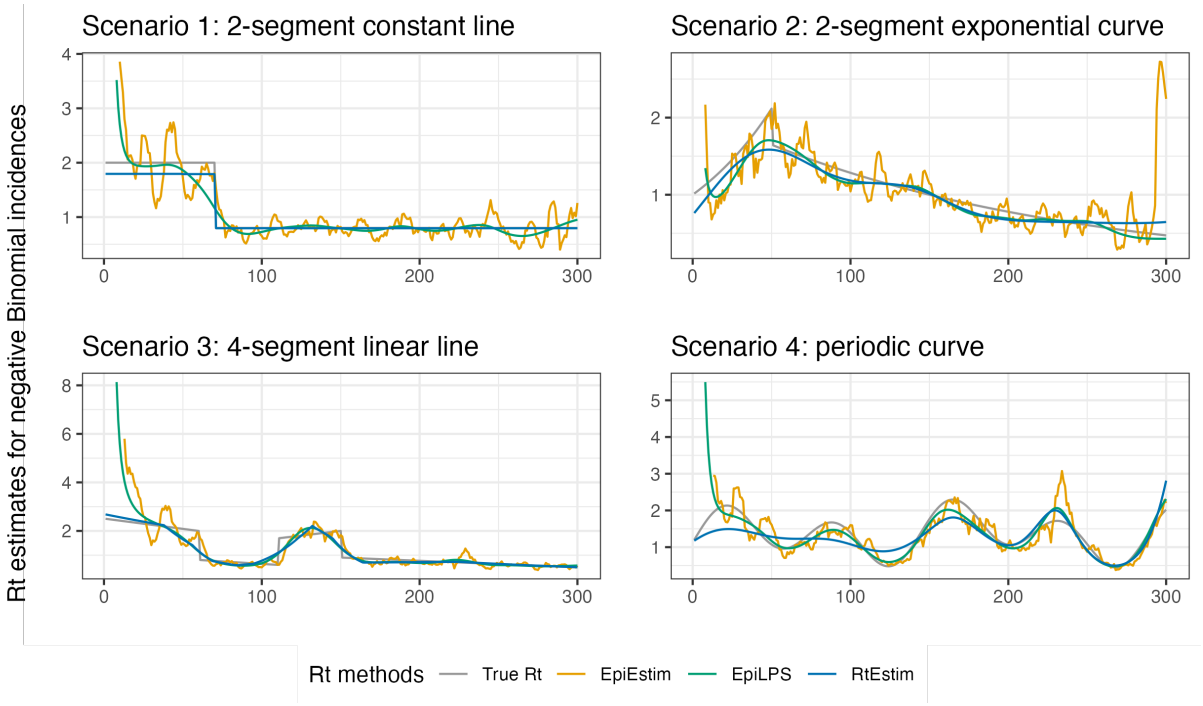

Figure 3: Effective reproduction number estimation for negative Binomial incidences.

gamma distribution with shape 2.5 and scale 2.5 to approximate the serial interval function, which is empirically found to be a reasonable choice.

Considering the temporal evolutions of effective reproduction numbers across $3, 4, 5$ days, the estimated reproduction numbers of Covid-19 in British Columbia (illustrated in Figure 6) are less than 3 during most of the time, which means that one distinct infected individuals can on average infect less than three other individuals in the population. The three degrees of the temporal evolution (across all regularization levels λ) all yield similar results that $\hat{\mathcal{R}}_t$ comes to the highest peak around the end of 2021 and then drops down to the lowest trough shortly thereafter. Throughout the estimated curves, the peaks and troughs of the reproduction numbers roughly come prior to the following growths and decays of confirmed cases respectively. We also visualize the 95% empirical confidence bands of the point estimates for the "best" tuning parameter (in terms of MSEs).

The reproduction numbers are relatively unstable before April 1st, 2022. The highest peak coincides with the emergence and globally spread of the Omicron variant. The estimated reproduction numbers are apparently below the threshold 1 during two time periods – roughly from April 1st, 2021 to July 1st, 2021 and from January 1st, 2022 to April 1st, 2022. The first trough coincides with the first authorization for use of Covid-19 vaccines in British Columbia. The second trough shortly after the greatest peak may credit to many aspects, including self-isolation of the infected individuals and application of the second shot of Covid-19 vaccines. Since around April 1st, 2022, the reproduction numbers stay stable (fluctuating around 1) and the infected cases stay low.

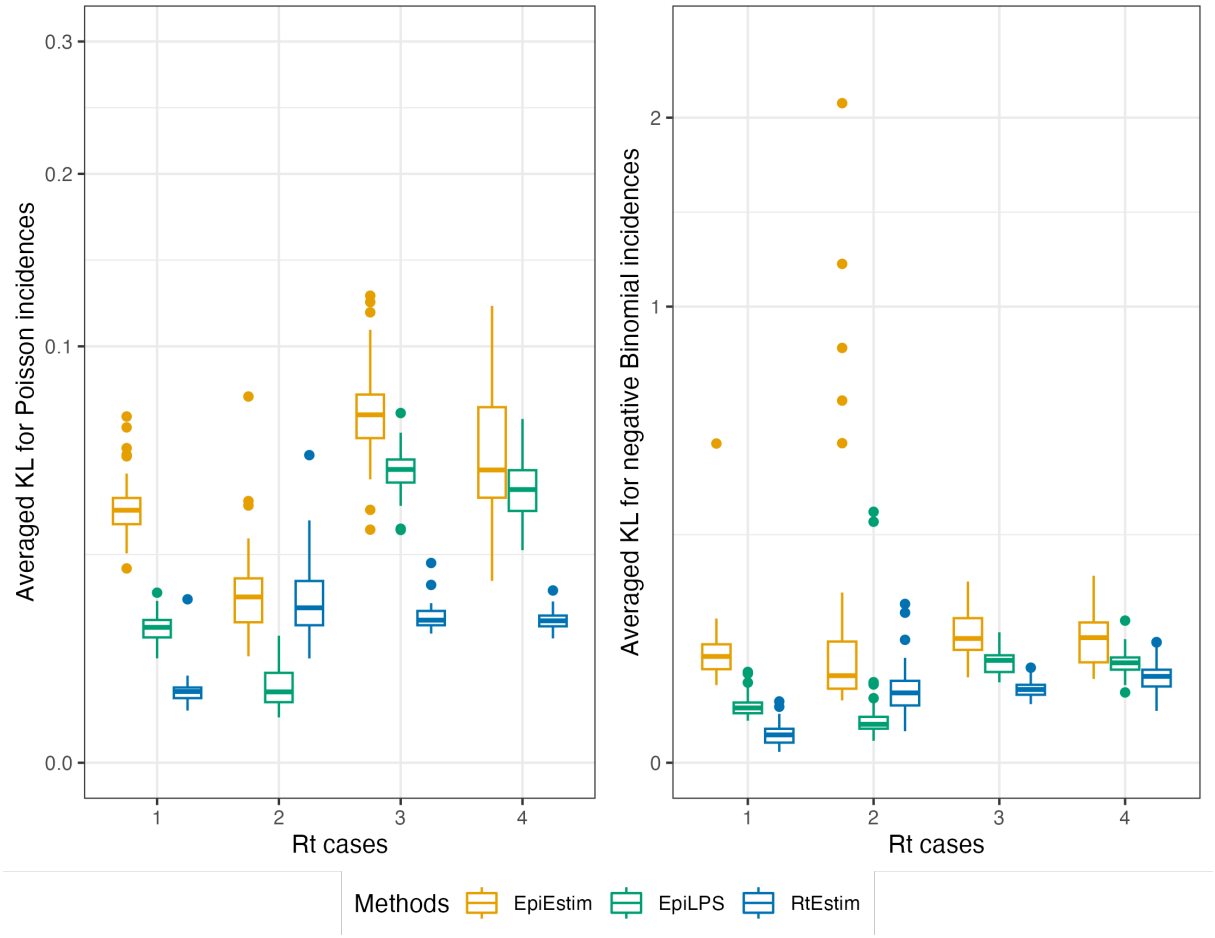

Figure 4: Boxplot of Kullback-Leibler divergence between the estimated effective reproduction numbers and the true ones across all methods given Poisson incidences and negative Binomial incidences across 50 samples. Left panel visualizes the KL divergences for the Poisson incidence cases. Right panel displays the KL divergences for the negative Binomial incidence cases. Outliers (larger than 3) of EpiEstim is excluded for a better visualization.

Greater regularization levels (by using larger λs) result in smoother estimated curves. Smoother curves suggest that the estimated reproduction numbers are around 1 during most time periods; however, it may miss to capture some outbreaks of the pandemic. More wiggly curves better reflect the fluctuation of $\mathcal{R}_t$, but sometimes fail to highlight the significant peaks or troughs. The tuning parameter λ needs to be chosen corresponding to the information in practice for a better interpretation. Here, we provide the CV-chosen $\mathcal{R}_t$ estimates with confidence bands.

3.4 Pandemic influenza in Baltimore, Maryland, 1918

We then apply `RtEstim` on the pandemic influenza in Baltimore, Maryland, 1918. Dataset in Figure 7 is obtained from the `R` package `EpiEstim`. The 1918 influenza, caused by H1N1 influenza A virus, was an unprecedentedly deadly influenza that infected almost one-third of the population

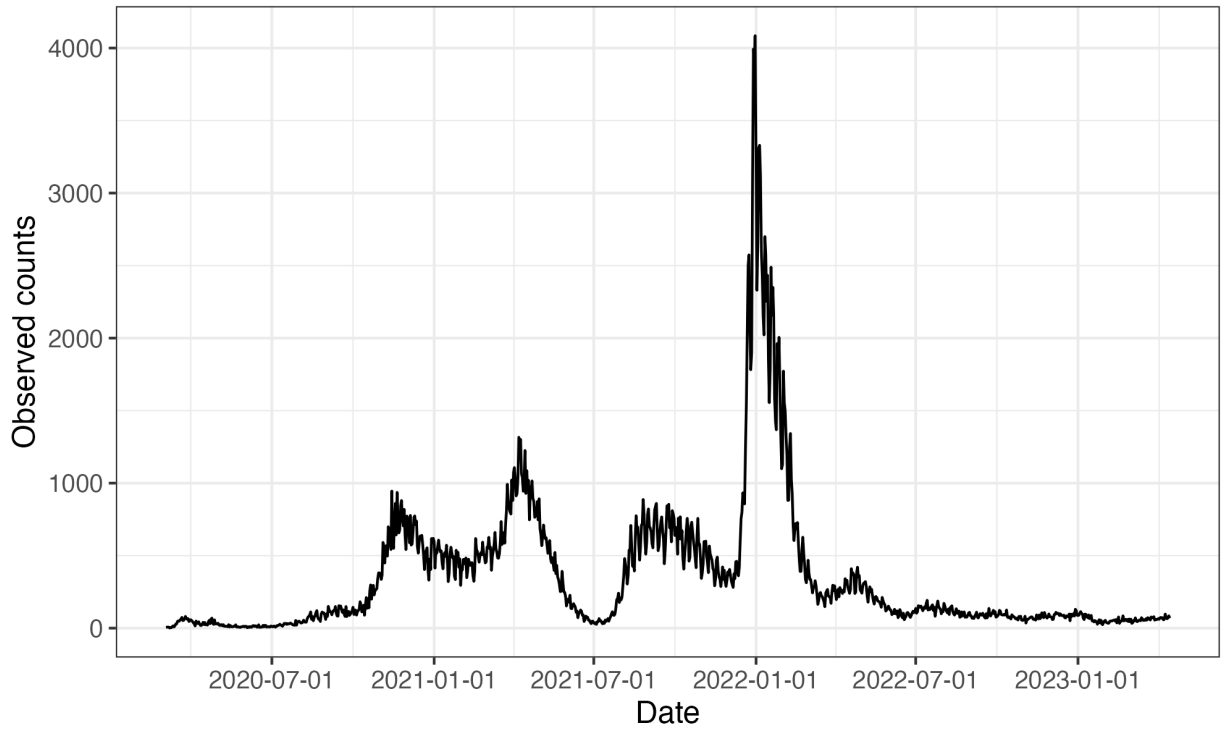

Figure 5: Covid19 daily confirmed incidence counts between March 1st, 2020 and April 15th, 2023 in British Columbia, Canada.

across the world (Taubenberger and Morens, 2006). In the estimation displayed in Figure 8, the CV-tuned piecewise cubic estimates better capture the growing tendency at the beginning of the pandemic. It suggests that the pandemic has yielded a decrease after around 20 days and reached 1 when the pandemic has lasted for nearly 50 days. However, it also suggests an increase at the end of the period, while a steady decline (as in CV-tuned piecewise constant and linear estimates) is more reasonable. The smoothness of $\mathcal{R}_t$ curves should be chosen based on the purpose of the study in practice, e.g., epidemic forecasting may require a more wiggly curve that contains more fluctuation information, while retrospective studies that solely target on understanding of the pandemic may prefer a smoother curve with less important information smoothed out.

4 Discussion

`RtEstim` provides a locally adaptive estimator using Poisson trend filtering on univariate data. It captures the heterogeneous smoothness of effective reproduction numbers given the observed incidence time series in a certain region. This is a nonparametric regression model which can be written as a convex optimization(minimization) problem. Minimizing the distance (averaged KL divergence per coordinate) between the estimators and (functions of) observations guarantees the data fidelity; minimizing a certain order of divided differences between each pair of neighboring

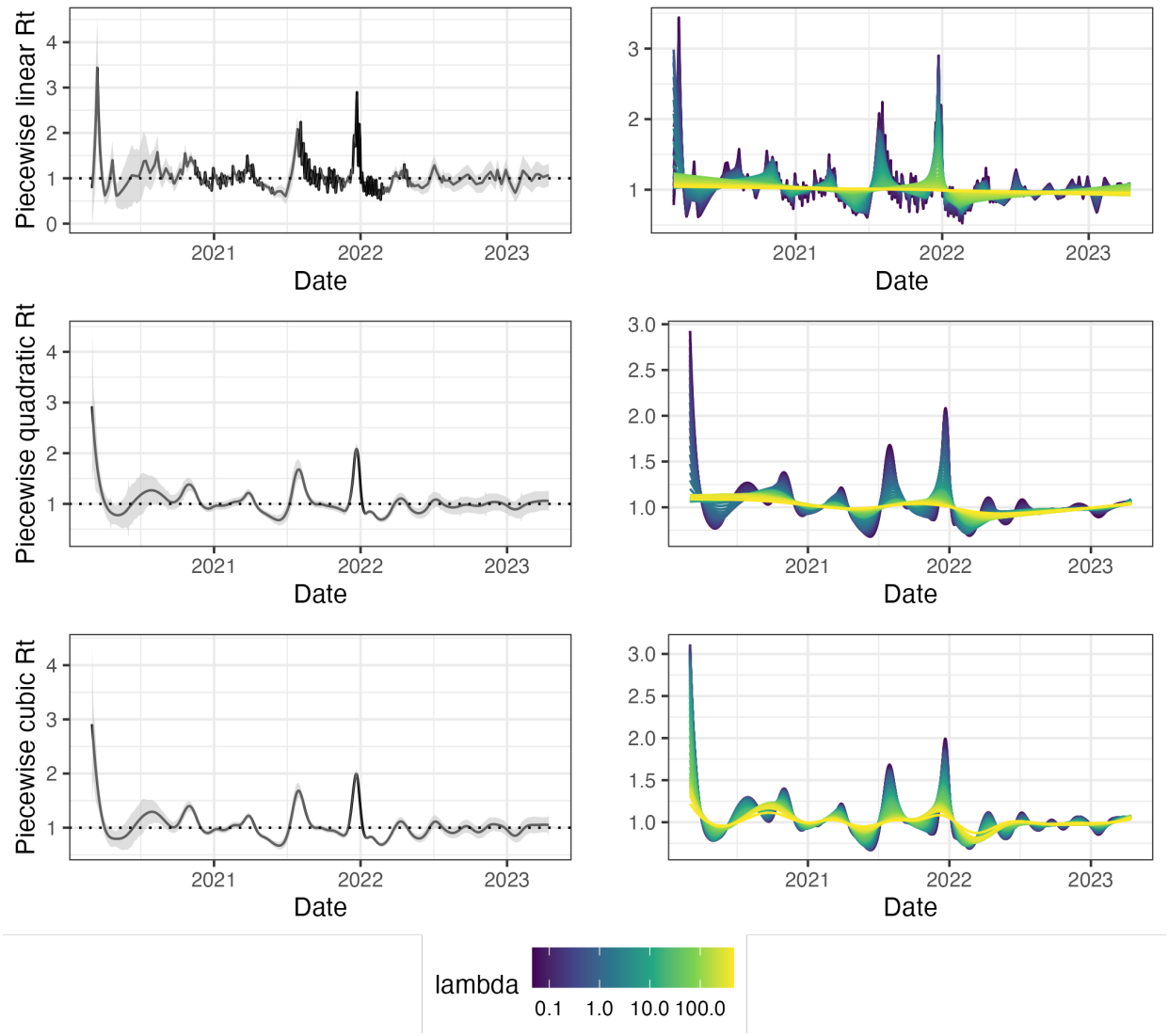

Figure 6: Estimated effective reproduction numbers for Covid19 daily confirmed counts between March 1st, 2020 and April 15th, 2023 in British Columbia, Canada. The left panels display the CV-tuned estimates with 95% confidence intervals. The right panels demonstrate estimates corresponding to 50 tuning parameters. The top, medium and bottom panels illustrate the estimated reproduction numbers ($\mathcal{R}_t$) using the Poisson trend filtering (in Equation (3)) with degrees $k = 1, 2, 3$ respectively.

parameters regularizes the smoothness. The ℓ_1 regularization introduces sparsity to the divided differences, which leads to heterogeneous smoothness within certain periods of time. The homogeneous smoothness within a time period can be either performed by a constant reproduction number, or a constant rate of changes, or a constant graphical curvature depending on the prescribed degree ($k = 0, 1, 2$ respectively).

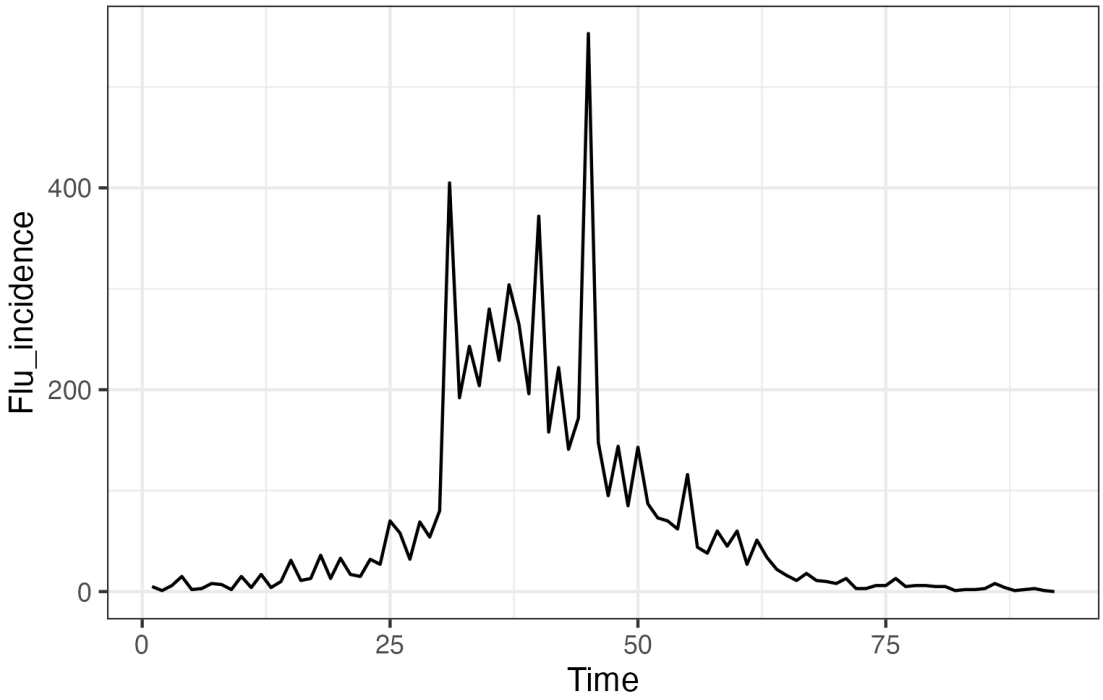

Figure 7: Pandemic influenza incidence counts in Baltimore, Maryland in 1918.

The property of local adaptivity is useful to distinguish, for example, the seasonal outbreaks from the un-seasonal outbreak periods. Given a properly chosen degree of polynomials, for example, the growth rate of un-seasonal outbreak periods can suggest a potential upcoming outbreak, which alerts epidemiologists to propose sanitary policies to prevent the progressing outbreak ahead of the infection surge. The effective reproduction numbers can be estimated afterwards to check the efficiency of the sanitary policies referring to whether they are below the threshold, their tendencies of reduction, or their graphical curvatures.

Our method `RtEstim` provides a natural way to deal with missing data, e.g., on weekends and holidays. We linearly impute the missing data in the computation of total primary infectiousness by assuming these values are missing at random. While solving the convex optimization problem, the edge lengths of the line graphs can be adjusted, so we can manually increase the length between two observations while penalizing the distance between them. It is remarkable that our focus is to provide a mathematical model for epidemiologists to use, rather than to focus on a specific disease. In addition, more specialized methodologies are needed for the diseases with relatively long incubation periods (e.g., HIV and HBV).

A group of epidemiological models are compartmental models. They establish the epidemic transmission process by creating compartments with labels and connecting them by directed edges. A simple compartmental model – for example, *Susceptible-Infectious-Susceptible* (SIS) model –

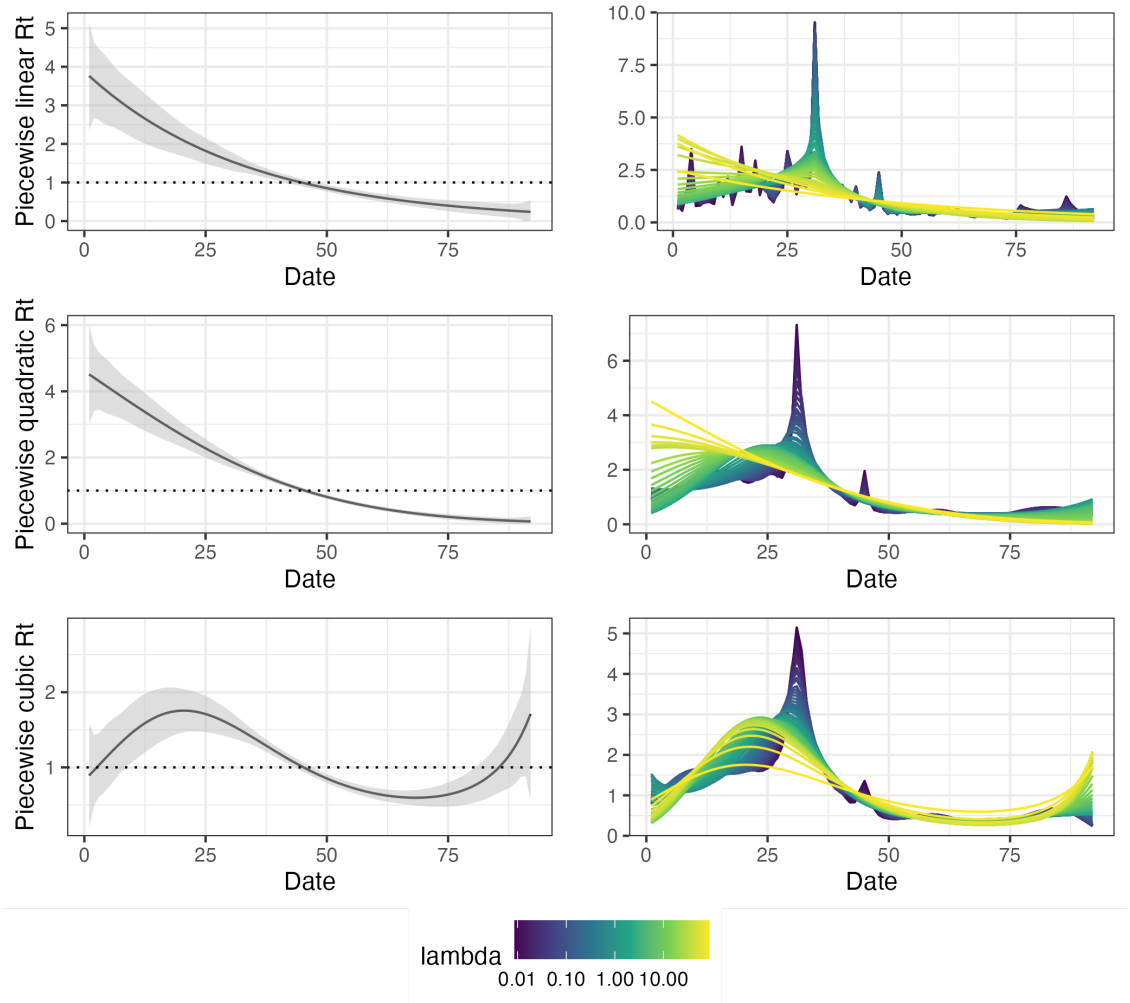

Figure 8: Estimated effective reproduction numbers for pandemic influenza incidence counts in Baltimore, Maryland in 1918. The left panels display the CV-tuned estimates with 95% confidence intervals. The right panels demonstrate estimates corresponding to 50 tuning parameters. The top, medium and bottom panels illustrate the estimated reproduction numbers ($\mathcal{R}_t$) using the Poisson trend filtering (in Equation (3)) with degrees $k = 1, 2, 3$ respectively.

divides the population (N) into two compartments for susceptible cases (S) and infectious cases (I) respectively and connects them in serial as $S \to I \to S$. It only focuses on susceptible individuals. Each directed edge corresponds to a ratio of transmission (say, α, β respectively). In such models, reproduction numbers are defined as functions of the estimated transmission parameters and the numbers of compartments or population, e.g., $\hat{\mathcal{R}}_0 = \hat{\beta}N/\hat{\alpha}$ in the SIS models Brauer et al. (2019), as by-products. Compartmental models usually solve ordinary differential equations (ODE) systems for transmission numbers (e.g., α, β in the SIS model). A disadvantage of such parametric models is that they are less flexible than nonparametric models and the number of parameters to be estimated grows along with the increase of compartments in practice, which results in a growing computational complexity. Since the epidemic mechanism depends highly on the contexts, e.g., if a latency period

exists or not, such models are lack of generalizability. Moreover, data of high quality are not always available for all compartments especially when there is a pandemic outbreak that results in a sudden shortage of resources in collecting daily new infections.

There are more practical considerations that may influence the quality of $\mathcal{R}_t$ estimation to be considered late. In our approach, we consider a homogeneous population without distinguishing the imported cases from the local cases. Poisson distribution is frequently used to model non-negative count data with heteroskedasticity. Another common alternative is negative Binomial distribution with or without a specified level of overdispersion. We consider a fixed serial interval throughout the transmission dynamics, but as the factors such as population immunity vary, the serial interval may vary as well due to the change of population factors such as herd immunity. Another common statement is that the distribution density of serial intervals is generally wider than the correspondence of generation intervals as serial interval includes both generation time and incubation time. If we assume generation time and incubation time both follow gamma distributions, the serial interval is likely to perform as a bimodal density.

References

Sam Abbott, Joel Hellewell, Robin N Thompson, Katharine Sherratt, Hamish P Gibbs, Nikos I Bosse, James D Munday, Sophie Meakin, Emma L Doughty, June Young Chun, et al. Estimating the time-varying reproduction number of sars-cov-2 using national and subnational case counts. *Wellcome Open Research*, 5(112):112, 2020.

Patrice Abry, Nelly Pustelnik, Stéphane Roux, Pablo Jensen, Patrick Flandrin, Rémi Gribonval, Charles-Gérard Lucas, Éric Guichard, Pierre Borgnat, and Nicolas Garnier. Spatial and temporal regularization to estimate covid-19 reproduction number r (t): Promoting piecewise smoothness via convex optimization. *Plos one*, 15(8):e0237901, 2020.

Amin Azmon, Christel Faes, and Niel Hens. On the estimation of the reproduction number based on misreported epidemic data. *Statistics in medicine*, 33(7):1176–1192, 2014.

Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical models in epidemiology*, volume 32. Springer, 2019.

Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology*, 178(9):1505–1512, 2013.

Anne Cori, Simon Cauchemez, Neil M Ferguson, Christophe Fraser, Elisabeth Dahlqwist, P Alex Demarsh, Thibaut Jombart, Zhian N Kamvar, Justin Lessler, Shikun Li, et al. Package 'epiestim'. *CRAN: Vienna Austria*, 2020.

Oswaldo Gressani, Christel Faes, and Niel Hens. An approximate bayesian approach for estimation of the reproduction number under misreported epidemic data. *MedRxiv*, pages 2021–05, 2021.

Oswaldo Gressani, Jacco Wallinga, Christian L Althaus, Niel Hens, and Christel Faes. Epilps: A fast and flexible bayesian tool for estimation of the time-varying reproduction number. *PLoS computational biology*, 18(10):e1010618, 2022.

Gary Hettinger, David Rubin, and Jing Huang. Estimating the instantaneous reproduction number with imperfect data: A method to account for case-reporting variation and serial interval uncertainty. *arXiv preprint arXiv:2302.12078*, 2023.

Faith Ho, Kris V Parag, Dillon C Adam, Eric HY Lau, Benjamin J Cowling, and Tim K Tsang. Accounting for the potential of overdispersion in estimation of the time-varying reproduction number. *Epidemiology*, 34(2):201–205, 2023.

Shihui Jin, Borame Lee Dickens, Jue Tao Lim, and Alex R Cook. Epimix: A novel method to estimate effective reproduction number. *Infectious Disease Modelling*, 2023.

Nicholas A Johnson. A dynamic programming algorithm for the fused lasso and ℓ_0-segmentation. *Journal of Computational and Graphical Statistics*, 22(2):246–260, 2013.

Seung-Jean Kim, Kwangmoo Koh, Stephen Boyd, and Dimitry Gorinevsky. ℓ_1 trend filtering. *SIAM review*, 51(2):339–360, 2009.

Adrian Lison, Sam Abbott, Jana Huisman, and Tanja Stadler. Generative bayesian modeling to nowcast the effective reproduction number from line list data with missing symptom onset dates. *arXiv preprint arXiv:2308.13262*, 2023.

Rebecca K Nash, Samir Bhatt, Anne Cori, and Pierre Nouvellet. Estimating the epidemic reproduction number from temporally aggregated incidence data: A statistical modelling approach and software tool. *PLOS Computational Biology*, 19(8):e1011439, 2023.

Kris V Parag. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Computational Biology*, 17(9):e1009347, 2021.

Barbara Pascal, Patrice Abry, Nelly Pustelnik, Stéphane Roux, Rémi Gribonval, and Patrick Flandrin. Nonsmooth convex optimization to estimate the covid-19 reproduction number space-time evolution with robustness against low quality data. *IEEE Transactions on Signal Processing*, 70:2859–2868, 2022.

Eugen Pircalabelu. A spline-based time-varying reproduction number for modelling epidemiological outbreaks. *Journal of the Royal Statistical Society Series C: Applied Statistics*, page qlad027, 2023a.

Eugen Pircalabelu. A spline-based time-varying reproduction number for modelling epidemiological outbreaks. *Journal of the Royal Statistical Society Series C: Applied Statistics*, page qlad027, 2023b.

Aaditya Ramdas and Ryan J Tibshirani. Fast and flexible admm algorithms for trend filtering. *Journal of Computational and Graphical Statistics*, 25(3):839–858, 2016.

Veeranjaneyulu Sadhanala, Robert Bassett, James Sharpnack, and Daniel J McDonald. Exponential family trend filtering on lattices. *arXiv preprint arXiv:2209.09175*, 2022.

Sam Abbott, Joel Hellewell, Katharine Sherratt, Katelyn Gostic, Joe Hickson, Hamada S. Badr, Michael DeWitt, Robin Thompson, EpiForecasts, and Sebastian Funk. *EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters*, 2020.

Jeffery K Taubenberger and David M Morens. 1918 influenza: the mother of all pandemics. *Revista Biomedica*, 17(1):69–79, 2006.

Robin N Thompson, Jake E Stockwin, Rolina D van Gaalen, Jonny A Polonsky, Zhian N Kamvar, P Alex Demarsh, Elisabeth Dahlqwist, Siyang Li, Eve Miguel, Thibaut Jombart, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29:100356, 2019.

Ryan J Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1):285–323, 2014.

Ryan J Tibshirani et al. Divided differences, falling factorials, and discrete splines: Another look at trend filtering and related problems. *Foundations and Trends® in Machine Learning*, 15(6): 694–846, 2022.

Cristiano Trevisin, Enrico Bertuzzo, Damiano Pasetto, Lorenzo Mari, Stefano Miccoli, Renato Casagrandi, Marino Gatto, and Andrea Rinaldo. Spatially explicit effective reproduction numbers from incidence and mobility data. *Proceedings of the National Academy of Sciences*, 120(20): e2219816120, 2023.