

EpiLPS: a fast and flexible Bayesian tool for near real-time estimation of the time-varying reproduction number

Non Peer-reviewed author version

GRESSANI, Oswaldo; Wallinga, Jacco; Althaus, Christian; HENS, Niel & FAES, Christel (2021) EpiLPS: a fast and flexible Bayesian tool for near real-time estimation of the time-varying reproduction number.

DOI: 10.1101/2021.12.02.21267189

Handle: <http://hdl.handle.net/1942/36417>

# EpiLPS: a fast and flexible Bayesian tool for near real-time estimation of the time-varying reproduction number

Oswaldo Gressani <sup>a,\*</sup>, Jacco Wallinga<sup>b,c</sup>, Christian Althaus<sup>d</sup>, Niel Hens<sup>a,e</sup>, Christel Faes<sup>a</sup>

<sup>a</sup> Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat),  
Data Science Institute, Hasselt University, Belgium

<sup>b</sup> Centre for Infectious Disease Control, National Institute for Public Health  
and the Environment, Bilthoven, The Netherlands

<sup>c</sup> Department of Biomedical Data Sciences, Leiden University Medical Centre,  
Leiden, The Netherlands

<sup>d</sup> Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

<sup>e</sup> Centre for Health Economics Research and Modelling Infectious Diseases,  
Vaxinfectio, University of Antwerp, Belgium

## Author summary

The instantaneous reproduction number  $R(t)$  is a key metric that provides important insights into an epidemic outbreak. We present a flexible Bayesian approach called EpiLPS (Epidemiological modeling with Laplacian-P-splines) for smooth estimation of the epidemic curve and  $R(t)$ . Computational speed and absence of arbitrary assumptions on smoothing makes EpiLPS an interesting tool for near real-time estimation of the reproduction number. An R software package is available (<https://github.com/oswaldogressani>).

---

\*Corresponding author. E-mail address: [oswaldo.gressani@uhasselt.be](mailto:oswaldo.gressani@uhasselt.be)

## Abstract

In infectious disease epidemiology, the instantaneous reproduction number  $R(t)$  is a time-varying metric defined as the average number of secondary infections generated by individuals who are infectious at time  $t$ . It is therefore a crucial epidemiological parameter that assists public health decision makers in the management of an epidemic. We present a new Bayesian tool for robust estimation of the time-varying reproduction number. The proposed methodology smooths the epidemic curve and allows to obtain (approximate) point estimates and credible envelopes of  $R(t)$  by employing the renewal equation, using Bayesian P-splines coupled with Laplace approximations of the conditional posterior of the spline vector. Two alternative approaches for inference are presented: (1) an approach based on a maximum a posteriori argument for the model hyperparameters, delivering estimates of  $R(t)$  in only a few seconds; and (2) an approach based on a MCMC scheme with underlying Langevin dynamics for efficient sampling of the posterior target distribution. Case counts per unit of time are assumed to follow a Negative Binomial distribution to account for potential excess variability in the data that would not be captured by a classic Poisson model. Furthermore, after smoothing the epidemic curve, a “plug-in” estimate of the reproduction number can be obtained from the renewal equation yielding a closed form expression of  $R(t)$  as a function of the spline parameters. The approach is extremely fast and free of arbitrary smoothing assumptions. EpiLPS is applied on data of SARS-CoV-1 in Hong-Kong (2003), influenza A H1N1 (2009) in the USA and current SARS-CoV-2 pandemic (2020-2021) for Belgium, Portugal, Denmark and France.

## 1 Motivation

The instantaneous reproduction number  $R(t)$  is a time-varying metric defined as the average number of secondary cases generated by infectious individuals at time  $t$ . During epidemic outbreaks,  $R(t)$  provides a snapshot (often on a daily basis) that quantifies the extent to which a given infectious disease is transmissible in a population and is therefore an important tool that assists governmental organizations in the management of a public

health crisis. The reproduction number is also a good proxy for measuring the real-time growth phase of an epidemic and as such, constitutes a key signal about the severity of the outbreak and the required control effort. For this reason, having a robust, accurate and timely estimator of  $R(t)$  is a crucial matter that has attracted considerable interest in developing new statistical approaches during the last two decades as summarized in [White et al. \(2021\)](#). The paper of [Gostic et al. \(2020\)](#) compares several methods for estimating  $R(t)$  and gives clear insights about the main challenges and obstacles that have to be faced. They recommend the method of [Cori et al. \(2013\)](#) and its associated EpiEstim package ([Cori, 2021](#)) as an appropriate and accurate tool for near real-time estimation of the instantaneous reproduction number. Another recent approach is proposed in [Parag \(2021\)](#), where a recursive Bayesian smoother based on Kalman filtering is used to derive a robust estimate of  $R(t)$  in periods of low incidence. The EpiNow2 package ([Abbott et al., 2020](#)) also provides interesting extensions and implementations of current best practices for precise estimation and forecast of the reproduction number using a Bayesian latent variable framework. Spline based approaches have shown to be an interesting tool for flexible modeling of the reproduction number. [Azmon et al. \(2014\)](#) use penalized radial splines for estimating  $R(t)$  under a Bayesian setting with misreported data and [Gressani et al. \(2021\)](#) accelerated the computational implementation by replacing the MCMC scheme with Laplace approximations. From a frequentist perspective, [Pircalabelu \(2021\)](#) uses truncated polynomials and radial basis splines to model the series of new infections and a derivative thereof as a candidate estimator for the reproduction number.

In this article, we propose a new Bayesian approach termed “EpiLPS” for estimating  $R(t)$  based on case incidence data and the serial interval distribution (the time elapsed between the onset of symptoms in an infector and the onset of symptoms in an infected case). Our estimator of  $R(t)$  is based on epidemic renewal-equations ([Fraser, 2007](#); [Wallinga and Lipsitch, 2007](#)) and Laplacian-P-splines (LPS) smoothing of the mean number of new cases by day of reporting. Time series of new cases by day of reporting are assumed to follow a Negative Binomial distribution to account for potential excess variability in the

data that would not be captured by a classic Poisson model. Algorithms related to Laplace approximations and evaluations of B-spline bases are coded in C++ and embedded in the R language through the Rcpp package (Eddelbuettel et al., 2011), making computational speed another key strength of EpiLPS as  $R(t)$  can be estimated in seconds. In addition, EpiLPS can also be used to obtain a smoothed estimate of the epidemic curve that can be of potential interest to further visualize an epidemic outbreak.

The proposed Bayesian methodology is based on a latent Gaussian model for the B-spline amplitudes and opens up two possible paths for inference. The first is called LPSMAP, a fully “sampling free” approach based on Laplace approximations to the conditional posterior of B-spline coefficients. The hyperparameter vector is fixed at its *maximum a posteriori* and credible envelopes of  $R(t)$  are computed via the “delta” method. The second path is called LPSMALA and is an MCMC-based approach with Langevin diffusions for efficient exploration of the posterior distribution of latent variables. The latter approach is computationally heavier than LPSMAP but has the merit of taking into account the uncertainty surrounding the hyperparameters. The underlying Metropolis-within-Gibbs structure keeps the practical implementation to a fairly simple level and the computational cost is reasonable even for long chain lengths.

Compared to existing methods, EpiLPS resembles EpiEstim from a methodological point of view in the sense that  $R(t)$  is estimated from incidence time series and a serial interval distribution, yet the two approaches fundamentally differ in many aspects. First, the methodology of Cori et al. (2013) assumes that incidence at time  $t$  is Poisson distributed, while EpiLPS assumes a Negative Binomial model. Second, as our approach uses penalized spline based approximations, the prior specifications are imposed on the roughness penalty parameter and not directly on  $R(t)$  as in EpiEstim. Third and most importantly, EpiLPS is free of any sliding window specification. An R package for EpiLPS has been developed and is available at <https://github.com/oswaldogressani>. The software also allows to compute the Cori et al. (2013) estimate of  $R(t)$  for the sake of comparison.

The manuscript is organized as follows. Section 2 aims at presenting the Laplacian-P-

splines model for smoothing count data. We show how the Laplace approximation applies to the conditional posterior of the B-spline amplitudes and also derive the (approximate) posterior of the hyperparameter vector to be optimized. This yields the *maximum a posteriori* estimate of the spline vector via Laplacian-P-splines (LPSMAP). Section 3 uses LPSMAP and proposes a “plug-in” estimate of  $R(t)$  based on renewal equations. Approximate credible intervals for  $R(t)$  will also be shown in this section. Section 4 shows an alternative path for estimation of  $R(t)$  based on Markov chain Monte Carlo (MCMC). The latter approach uses Langevin dynamics for efficient sampling of the target posterior distribution. This approach is termed LPSMALA for “Laplacian-P-splines with a Metropolis-adjusted Langevin algorithm”. Section 5 is devoted to assess the performance of EpiLPS in various simulation scenarios and make comparisons with EpiEstim. In Section 6, we apply EpiLPS to real world epidemic outbreaks. Finally, Section 7 concludes with a discussion.

## 2 Methodology behind EpiLPS

### 2.1 Negative Binomial model for case incidence data

Let  $\{y_t, t = 1, \dots, T\}$  be a time series of counts during an epidemic of  $T$  days with  $y_t \in \mathbb{N}$  (set of non-negative integers) denoting the total number of new contaminations on day  $t$ . We assume that the number of cases on day  $t$  follows a Negative Binomial distribution  $y_t \sim \text{NegBin}(\mu(t), \rho)$ , with  $\mu(t), \rho \in \mathbb{R}_+^* := \{x \in \mathbb{R} | x > 0\}$  and probability mass function (see e.g. [Anscombe, 1950](#); [Piegorisch, 1990](#)):

$$p(y_t | \mu(t), \rho) = \frac{\Gamma(y_t + \rho)}{\Gamma(y_t + 1)\Gamma(\rho)} \left( \frac{\mu(t)}{\mu(t) + \rho} \right)^{y_t} \left( \frac{\rho}{\rho + \mu(t)} \right)^\rho, \quad (1)$$

where  $\Gamma(\cdot)$  is the gamma function. Under the above parameterization, we have  $\mathbb{E}(y_t) = \mu(t)$  and  $\mathbb{V}(y_t) = \mu(t) + \mu(t)^2/\rho$ , so that  $1/\rho$  is the parameter responsible for overdispersion (variance larger than the mean) that is absent in a Poisson setting. In the limiting case  $\lim_{\rho \rightarrow +\infty} \mathbb{V}(y_t) = \mu(t) = \mathbb{E}(y_t)$  and we recover the Poisson model. We assume that  $\mu(t)$

evolves smoothly over the time course of the epidemic and model it with cubic B-splines (Eilers and Marx, 1996):

$$\log(\mu(t)) = \sum_{k=1}^K \theta_k b_k(t) = \boldsymbol{\theta}^\top b(t), \quad (2)$$

where  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)^\top$  is the vector of B-spline amplitudes to be estimated and  $b(\cdot) = (b_1(\cdot), \dots, b_K(\cdot))^\top$  is a cubic B-spline basis defined on the domain  $\mathcal{T} = [r_l, T]$ , where  $r_l$  is a lower bound on the time axis, typically the first day of the epidemic (i.e.  $r_l = 1$ ). The philosophy behind P-splines consists in specifying a “large” number  $K$  of basis functions together with a discrete roughness penalty  $\lambda \boldsymbol{\theta}^\top P \boldsymbol{\theta}$  as a counterforce to the induced flexibility of the fit. The parameter  $\lambda > 0$  acts as a tuning parameter calibrating the “degree” of smoothness and  $P = D_r^\top D_r + \varepsilon I_K$  is a penalty matrix built from  $r$ th order difference matrices  $D_r$  of dimension  $(K - r) \times K$  perturbed by an  $\varepsilon$ -multiple (here  $\varepsilon = 10^{-6}$ ) of the  $K$ -dimensional identity matrix  $I_K$  to ensure full rankedness. The reader is redirected to Eilers and Marx (2021) for a complete textbook treatment of P-splines. Following Lang and Brezger (2004), we impose a Gaussian prior on the spline vector  $\boldsymbol{\theta} | \lambda \sim \mathcal{N}_{\dim(\boldsymbol{\theta})}(0, Q_\lambda^{-1})$ , with precision matrix  $Q_\lambda = \lambda P$ . For full Bayesian inference, the following priors are imposed on the model hyperparameters. Following Jullion and Lambert (2007), a robust Gamma prior is specified for the roughness penalty parameter  $\lambda | \delta \sim \mathcal{G}(\phi/2, (\phi\delta)/2)$ , where  $\mathcal{G}(a, b)$  is a Gamma distribution with mean  $a/b$  and variance  $a/b^2$ ,  $\phi = 2$  and  $\delta$  is an additional dispersion parameter with hyperprior  $\delta \sim \mathcal{G}(a_\delta = 10, b_\delta = 10)$ . Finally, the following uninformative prior is imposed on the overdispersion parameter  $\rho \sim \mathcal{G}(a_\rho = 0.0001, b_\rho = 0.0001)$ . Let  $\boldsymbol{\eta} := (\lambda, \rho)^\top$  denote the vector of hyperparameters. The full Bayesian model is thus:

$$\begin{aligned} y_t | \mu(t), \rho &\sim \text{NegBin}(\mu(t), \rho), \\ \log(\mu(t)) &= \boldsymbol{\theta}^\top b(t), \\ \boldsymbol{\theta} | \lambda &\sim \mathcal{N}_{\dim(\boldsymbol{\theta})}(0, Q_\lambda^{-1}), \end{aligned}$$

$$\begin{aligned}\lambda|\delta &\sim \mathcal{G}(\phi/2, (\phi\delta)/2), \\ \delta &\sim \mathcal{G}(a_\delta, b_\delta), \\ \rho &\sim \mathcal{G}(a_\rho, b_\rho).\end{aligned}$$

## 2.2 Laplace approximation to the conditional posterior of $\theta$

The log-likelihood for the Negative Binomial model is given by:

$$\ell(\theta, \rho; \mathcal{D}) \doteq \sum_{t=1}^T \{g(y_t, \rho) + y_t \theta^\top b(t) + \rho \log(\rho) - (y_t + \rho) \log(\exp(\theta^\top b(t)) + \rho)\}, \quad (3)$$

with  $g(y_t, \rho) = \log \Gamma(y_t + \rho) - \log \Gamma(\rho)$  and  $\doteq$  denoting equality up to an additive constant.

The gradient of the log-likelihood with respect to the spline coefficients is:

$$\nabla \ell(\theta, \rho; \mathcal{D}) = \left( \frac{\partial \ell(\theta, \rho; \mathcal{D})}{\partial \theta_1}, \dots, \frac{\partial \ell(\theta, \rho; \mathcal{D})}{\partial \theta_K} \right)^\top,$$

where:

$$\frac{\partial \ell(\theta, \rho; \mathcal{D})}{\partial \theta_k} = \sum_{t=1}^T y_t b_k(t) - \sum_{t=1}^T \frac{(y_t + \rho) \exp(\theta^\top b(t))}{(\exp(\theta^\top b(t)) + \rho)} b_k(t), \quad k = 1, \dots, K.$$

The Hessian of the log-likelihood with respect to the B-spline amplitudes is:

$$\nabla^2 \ell(\theta, \rho; \mathcal{D}) = \begin{pmatrix} \frac{\partial^2 \ell(\theta, \rho; \mathcal{D})}{\partial \theta_1^2} & \cdots & \frac{\partial^2 \ell(\theta, \rho; \mathcal{D})}{\partial \theta_1 \partial \theta_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell(\theta, \rho; \mathcal{D})}{\partial \theta_K \partial \theta_1} & \cdots & \frac{\partial^2 \ell(\theta, \rho; \mathcal{D})}{\partial \theta_K^2} \end{pmatrix},$$

with entries:



$$\frac{\partial^2 \ell(\boldsymbol{\theta}, \rho; \mathcal{D})}{\partial \theta_k \partial \theta_l} = - \sum_{t=1}^T \rho(y_t + \rho) \frac{\exp(\boldsymbol{\theta}^\top b(t))}{(\exp(\boldsymbol{\theta}^\top b(t)) + \rho)^2} b_k(t) b_l(t), \quad k, l = 1, \dots, K.$$

Using Bayes' rule, the conditional posterior of  $\boldsymbol{\theta}$  for a given  $\boldsymbol{\eta}$  is:

$$\begin{aligned} p(\boldsymbol{\theta} | \boldsymbol{\eta}, \mathcal{D}) &\propto \mathcal{L}(\boldsymbol{\theta}, \rho; \mathcal{D}) p(\boldsymbol{\theta} | \lambda) \\ &\propto \exp\left(\ell(\boldsymbol{\theta}, \rho; \mathcal{D}) - \frac{\lambda}{2} \boldsymbol{\theta}^\top P \boldsymbol{\theta}\right). \end{aligned} \quad (4)$$

The gradient and Hessian of the log-likelihood (3) can be used to compute the gradient and Hessian of the (log-)conditional posterior (4), namely:

$$\begin{aligned} \nabla \log p(\boldsymbol{\theta} | \boldsymbol{\eta}, \mathcal{D}) &= \nabla \ell(\boldsymbol{\theta}, \rho; \mathcal{D}) - \lambda P \boldsymbol{\theta}, \\ \nabla^2 \log p(\boldsymbol{\theta} | \boldsymbol{\eta}, \mathcal{D}) &= \nabla^2 \ell(\boldsymbol{\theta}, \rho; \mathcal{D}) - \lambda P. \end{aligned}$$

The above two equations will be used in a Newton-Raphson algorithm to obtain the Laplace approximation to the conditional posterior of  $\boldsymbol{\theta}$ :

$$\tilde{p}_G(\boldsymbol{\theta} | \boldsymbol{\eta}, \mathcal{D}) = \mathcal{N}_{\dim(\boldsymbol{\theta})}(\boldsymbol{\theta}^*(\boldsymbol{\eta}), \Sigma^*(\boldsymbol{\eta})), \quad (5)$$

where  $\boldsymbol{\theta}^*(\boldsymbol{\eta})$  and  $\Sigma^*(\boldsymbol{\eta})$  is the mode and variance-covariance respectively after convergence of the Newton-Raphson algorithm. The latter two quantities are functions of the hyperparameter vector  $\boldsymbol{\eta}$ . An intuitive choice for  $\boldsymbol{\eta}$  is to fix it at its *maximum a posteriori* (MAP). This is the option retained here, although it is also possible to work with a grid-based approach (Rue et al., 2009; Gressani and Lambert, 2021).

### 2.3 Hyperparameter optimization

The hyperparameter vector  $\boldsymbol{\eta} = (\lambda, \rho)^\top$  will be calibrated by posterior optimization. Following Tierney and Kadane (1986) and Rue et al. (2009), the hyperparameter vector can be approximated as follows:

$$\tilde{p}(\boldsymbol{\eta}, \delta | \mathcal{D}) \propto \frac{\mathcal{L}(\boldsymbol{\theta}, \rho; \mathcal{D}) p(\boldsymbol{\theta} | \lambda) p(\lambda | \delta) p(\delta) p(\rho)}{\tilde{p}_G(\boldsymbol{\theta} | \boldsymbol{\eta}, \mathcal{D})} \Bigg|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*(\boldsymbol{\eta})}, \quad (6)$$

Approximation (6) can be written extensively as follows:

$$\begin{aligned} \tilde{p}(\boldsymbol{\eta}, \delta | \mathcal{D}) &\propto \lambda^{\frac{K+\phi}{2}-1} \delta^{\frac{\phi}{2}+a_\delta-1} \exp\left(-\delta \left(\frac{\phi\lambda}{2} + b_\delta\right)\right) \rho^{a_\rho-1} \\ &\times |\Sigma^*(\boldsymbol{\eta})|^{\frac{1}{2}} \exp\left(\ell(\boldsymbol{\theta}^*(\boldsymbol{\eta}), \rho; \mathcal{D}) - \frac{\lambda}{2} \boldsymbol{\theta}^{*\top}(\boldsymbol{\eta}) P \boldsymbol{\theta}^*(\boldsymbol{\eta}) - b_\rho \rho\right), \end{aligned}$$

where the  $K/2$  power of  $\lambda$  comes from the determinant  $|Q_\lambda^{-1}|^{-1/2} = |\lambda P|^{1/2} \propto \lambda^{K/2}$ . As  $\delta^{\frac{\phi}{2}+a_\delta-1} \exp(-\delta(\frac{\phi\lambda}{2} + b_\delta))$  is the kernel of a Gamma distribution for the dispersion parameter  $\delta$ , the following integral can be analytically solved:

$$\begin{aligned} \int_0^{+\infty} \tilde{p}(\boldsymbol{\eta}, \delta | \mathcal{D}) d\delta &= \tilde{p}(\boldsymbol{\eta} | \mathcal{D}) \\ &\propto \lambda^{\frac{K+\phi}{2}-1} \left(\frac{\phi\lambda}{2} + b_\delta\right)^{-\left(\frac{\phi}{2}+a_\delta\right)} \rho^{a_\rho-1} \\ &\times |\Sigma^*(\boldsymbol{\eta})|^{\frac{1}{2}} \exp\left(\ell(\boldsymbol{\theta}^*(\boldsymbol{\eta}), \rho; \mathcal{D}) - \frac{\lambda}{2} \boldsymbol{\theta}^{*\top}(\boldsymbol{\eta}) P \boldsymbol{\theta}^*(\boldsymbol{\eta}) - b_\rho \rho\right). \end{aligned}$$

Using the transformation of variables (ensuring numerical stability during optimization)  $w = \log(\rho)$ ,  $v = \log(\lambda)$ , one can show that  $\tilde{p}(\boldsymbol{\eta} | \mathcal{D})$  can be written as follows after using the multivariate transformation method:

$$\begin{aligned} \tilde{p}(\tilde{\boldsymbol{\eta}} | \mathcal{D}) &\propto \exp(v)^{\frac{K+\phi}{2}} \left(\frac{\phi \exp(v)}{2} + b_\delta\right)^{-\left(\frac{\phi}{2}+a_\delta\right)} \exp(w)^{a_\rho} \\ &\times |\Sigma^*(\tilde{\boldsymbol{\eta}})|^{\frac{1}{2}} \exp\left(\ell(\boldsymbol{\theta}^*(\tilde{\boldsymbol{\eta}}), \exp(w); \mathcal{D}) - \frac{\exp(v)}{2} \boldsymbol{\theta}^{*\top}(\tilde{\boldsymbol{\eta}}) P \boldsymbol{\theta}^*(\tilde{\boldsymbol{\eta}}) - b_\rho \exp(w)\right), \end{aligned}$$

where  $\tilde{\boldsymbol{\eta}} = (w, v)^\top$ . The approximated log-posterior becomes:

$$\begin{aligned} \log \tilde{p}(\tilde{\boldsymbol{\eta}} | \mathcal{D}) &\doteq 0.5 \log |\Sigma^*(\tilde{\boldsymbol{\eta}})| + 0.5(K + \phi)v + a_\rho w - (0.5\phi + a_\delta) \log(0.5\phi \exp(v) + b_\delta) \\ &+ \ell(\boldsymbol{\theta}^*(\tilde{\boldsymbol{\eta}}), \exp(w); \mathcal{D}) - 0.5 \exp(v) \boldsymbol{\theta}^{*\top}(\tilde{\boldsymbol{\eta}}) P \boldsymbol{\theta}^*(\tilde{\boldsymbol{\eta}}) - b_\rho \exp(w). \end{aligned} \quad (7)$$

Equation (7) is numerically optimized and yields  $\tilde{\boldsymbol{\eta}}^* = \operatorname{argmax}_{\tilde{\boldsymbol{\eta}}} \log \tilde{p}(\tilde{\boldsymbol{\eta}}|\mathcal{D})$ . Plugging the latter vector into the Laplace approximation (5), one obtains the estimate  $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^*(\tilde{\boldsymbol{\eta}}^*)$  of the spline vector. The latter can be seen as a *maximum a posteriori* (MAP) estimate of  $\boldsymbol{\theta}$ . Thus, the approximated posterior of the spline vector is:

$$\tilde{p}(\boldsymbol{\theta}|\mathcal{D}) = \mathcal{N}_{\dim(\boldsymbol{\theta})}(\boldsymbol{\theta}^*(\tilde{\boldsymbol{\eta}}^*), \Sigma^*(\tilde{\boldsymbol{\eta}}^*)), \quad (8)$$

and can be used to construct credible intervals for functions that indirectly depend on  $\boldsymbol{\theta}$ , such as  $R(t)$  as shown in the following section.

### 3 Estimation of $R(t)$ with LPSMAP

#### 3.1 The renewal equation “plug-in” estimate

In this section, we show how the Negative Binomial model for smoothing incidence counts can be used to estimate  $R(t)$  through the renewal equation. Let  $\boldsymbol{\varphi} = \{\varphi_1, \dots, \varphi_k\}$  be a known  $k$ -dimensional vector representing the serial interval distribution. The renewal equation describes the link between the number of new cases at day  $t$  and past infected cases (up to time point  $t - 1$ ) weighted by the serial interval distribution. Mathematically:

$$y_t = \sum_s R(t) \varphi_s y_{t-s}. \quad (9)$$

Rearranging (9) and taking the length  $k$  of the serial interval into account, we obtain an equation with the instantaneous reproduction number on the left-hand side:

$$R(t) = \begin{cases} y_t & \text{for } t = 1, \\ y_t \left( \sum_{s=1}^{t-1} \varphi_s y_{t-s} \right)^{-1} & \text{for } 2 \leq t \leq k, \\ y_t \left( \sum_{s=1}^k \varphi_s y_{t-s} \right)^{-1} & \text{for } k < t \leq T. \end{cases} \quad (10)$$

A “plug-in” estimator of  $R(t)$  at any  $t \in \mathcal{T}$  is obtained by replacing the number of new cases  $y_t$  by the estimated mean number of cases  $\hat{\mu}(t) = \exp(\hat{\boldsymbol{\theta}}^\top b(t))$ , yielding:

$$\hat{R}(t) = \begin{cases} \exp(\hat{\boldsymbol{\theta}}^\top b(t)) & \text{for } t = 1, \\ \exp(\hat{\boldsymbol{\theta}}^\top b(t)) \left( \sum_{s=1}^{t-1} \varphi_s \exp(\hat{\boldsymbol{\theta}}^\top b(t-s)) \right)^{-1} & \text{for } 2 \leq t \leq k, \\ \exp(\hat{\boldsymbol{\theta}}^\top b(t)) \left( \sum_{s=1}^k \varphi_s \exp(\hat{\boldsymbol{\theta}}^\top b(t-s)) \right)^{-1} & \text{for } k < t \leq T. \end{cases} \quad (11)$$

Using the indicator function  $\mathbb{I}(\cdot)$ , i.e.  $\mathbb{I}(A) = 1$  if condition  $A$  is true and  $\mathbb{I}(A) = 0$  otherwise, the above “plug-in” estimate of  $R(t)$  can be written in a single line:

$$\begin{aligned} \hat{R}(t) = & \exp(\hat{\boldsymbol{\theta}}^\top b(t)) \left\{ \mathbb{I}(t = 1) + \left( \sum_{s=1}^{t-1} \varphi_s \exp(\hat{\boldsymbol{\theta}}^\top b(t-s)) \right)^{-1} \mathbb{I}(2 \leq t \leq k) \right. \\ & \left. + \left( \sum_{s=1}^k \varphi_s \exp(\hat{\boldsymbol{\theta}}^\top b(t-s)) \right)^{-1} \mathbb{I}(k < t \leq T) \right\}. \end{aligned} \quad (12)$$

### 3.2 Credible intervals for $R(t)$

Replacing the number of new cases  $y_t$  by the (theoretical) mean number of cases  $\mu(t) = \exp(\boldsymbol{\theta}^\top b(t))$  in (10) and using the compact notation, one has:

$$\begin{aligned} R(t) = & \exp(\boldsymbol{\theta}^\top b(t)) \left\{ \mathbb{I}(t = 1) + \left( \sum_{s=1}^{t-1} \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) \right)^{-1} \mathbb{I}(2 \leq t \leq k) \right. \\ & \left. + \left( \sum_{s=1}^k \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) \right)^{-1} \mathbb{I}(k < t \leq T) \right\}. \end{aligned} \quad (13)$$

Let us denote the log of the instantaneous reproduction number in (13) as:

$$\begin{aligned} \log R(t) & := h(\boldsymbol{\theta}|t) \\ & = \boldsymbol{\theta}^\top b(t) + \log \zeta(\boldsymbol{\theta}), \end{aligned}$$

with

$$\begin{aligned}\zeta(\boldsymbol{\theta}) &= \mathbb{I}(t = 1) + \left( \sum_{s=1}^{t-1} \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) \right)^{-1} \mathbb{I}(2 \leq t \leq k) \\ &\quad + \left( \sum_{s=1}^k \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) \right)^{-1} \mathbb{I}(k < t \leq T).\end{aligned}$$

Note that  $h(\boldsymbol{\theta}|t)$  is seen here as a function of the spline vector  $\boldsymbol{\theta}$  for a given time point  $t \in \mathcal{T}$ . A  $(1 - \alpha) \times 100\%$  approximate credible interval for  $R(t)$  is obtained via a “delta” method. Consider a first-order Taylor expansion of  $h(\boldsymbol{\theta}|t)$  around  $\boldsymbol{\theta}^*(\tilde{\boldsymbol{\eta}}^*)$  (henceforth  $\boldsymbol{\theta}^*$  for the sake of a light notation), the mean of the Laplace approximated posterior of the spline vector in (8):

$$h(\boldsymbol{\theta}|t) \approx h(\boldsymbol{\theta}^*|t) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \nabla h(\boldsymbol{\theta}|t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}, \quad (14)$$

where the  $k$ th entry of the gradient vector  $\nabla h(\boldsymbol{\theta}|t) = (\partial h(\boldsymbol{\theta}|t)/\partial \theta_1, \dots, \partial h(\boldsymbol{\theta}|t)/\partial \theta_K)^\top$  is:

$$\begin{aligned}\frac{\partial h(\boldsymbol{\theta}|t)}{\partial \theta_k} &= b_k(t) + \zeta^{-1}(\boldsymbol{\theta}) \frac{\partial \zeta(\boldsymbol{\theta})}{\partial \theta_k}. \\ \frac{\partial \zeta(\boldsymbol{\theta})}{\partial \theta_k} &= - \left( \sum_{s=1}^{t-1} \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) \right)^{-2} \sum_{s=1}^{t-1} \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) b_k(t-s) \mathbb{I}(2 \leq t \leq k) \\ &\quad - \left( \sum_{s=1}^k \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) \right)^{-2} \sum_{s=1}^k \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) b_k(t-s) \mathbb{I}(k < t \leq T).\end{aligned}$$

It follows that for  $k = 1, \dots, K$  we have:

$$\begin{aligned}\frac{\partial h(\boldsymbol{\theta}|t)}{\partial \theta_k} &= b_k(t) + \left\{ 0 \mathbb{I}(t = 1) \right. \\ &\quad - \left( \sum_{s=1}^{t-1} \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) \right)^{-1} \sum_{s=1}^{t-1} \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) b_k(t-s) \mathbb{I}(2 \leq t \leq k) \\ &\quad \left. - \left( \sum_{s=1}^k \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) \right)^{-1} \sum_{s=1}^k \varphi_s \exp(\boldsymbol{\theta}^\top b(t-s)) b_k(t-s) \mathbb{I}(k < t \leq T) \right\}.\end{aligned}$$

The Taylor expansion in (14) is a linear combination of the vector  $\boldsymbol{\theta}$  that is *a posteriori* (approximately) Gaussian due to the Laplace approximation. As the family of Gaussian distributions is closed under linear combinations, it follows that  $h(\boldsymbol{\theta}|t)$  (and hence  $\log R(t)$ ) is *a posteriori* also (approximately) Gaussian with mean  $\mathbb{E}(h(\boldsymbol{\theta}|t)) \approx h(\boldsymbol{\theta}^*|t)$  and variance-covariance matrix  $\mathbb{V}(h(\boldsymbol{\theta}|t)) \approx \nabla^\top h(\boldsymbol{\theta}|t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \Sigma^* \nabla h(\boldsymbol{\theta}|t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ , where  $\Sigma^* := \Sigma^*(\tilde{\boldsymbol{\eta}}^*)$  is the covariance matrix of the Laplace approximation (8). This suggests to write:

$$\left( \log R(t) | \mathcal{D} \right) \approx \mathcal{N}_1 \left( h(\boldsymbol{\theta}^*|t), \nabla^\top h(\boldsymbol{\theta}|t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \Sigma^* \nabla h(\boldsymbol{\theta}|t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right). \quad (15)$$

Thus, from (15) a quantile-based  $(1 - \alpha) \times 100\%$  approximate credible interval for  $R(t)$  is:

$$CI_{R(t)}^{1-\alpha} = \exp \left( h(\boldsymbol{\theta}^*|t) \pm z_{\alpha/2} \sqrt{\nabla^\top h(\boldsymbol{\theta}|t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \Sigma^* \nabla h(\boldsymbol{\theta}|t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}} \right), \quad (16)$$

where  $z_{\alpha/2}$  is the  $\alpha/2$ -upper quantile of a standard normal variate.

## 4 Estimation of $R(t)$ with LPSMALA

In Bayesian statistics, posterior distributions obtained with Bayes' theorem often entail a high degree of analytical complexity and to a certain extent a flavor of incompleteness in the sense that the posterior is typically only known up to a normalizing constant. Markov chain Monte Carlo (MCMC) methods are simulation techniques for generating samples from a (possibly unnormalized) target distribution (Chib and Greenberg, 1996). One of the most popular MCMC methods together with the Gibbs sampler (Geman and Geman, 1984) is the Metropolis-Hastings (MH) algorithm originally proposed by Metropolis et al. (1953) and later generalized by Hastings (1970). In this section, we propose to implement a modified version of the Metropolis-adjusted Langevin algorithm (MALA) (Roberts and Tweedie, 1996) within the EpiLPS framework. The major advantage of MALA as compared to MH algorithms is that the proposal distribution is based upon a discretized approximation of the Langevin diffusion that uses the gradient of the target posterior distribution. These

“smarter” proposals make use of additional information about the target density so that algorithms based on Langevin dynamics can converge at sub-geometric rates and tend to be more efficient than naive random-walk Metropolis algorithms ([Roberts and Rosenthal, 1998, 2001](#)).

This motivates our choice for embedding a MALA algorithm in EpiLPS as an efficient way of obtaining MCMC samples for inference on the instantaneous reproduction number  $R(t)$  via the renewal equation. The end-user will thus have a fully flexible choice regarding the underlying approach for estimating  $R(t)$  either via Laplacian-P-splines where the uncertainty surrounding the parameter  $\lambda$  responsible for smoothing is ignored and fixed at its “maximum a posteriori” (LPSMAP); or via a modified MALA algorithm where the uncertainty surrounding the penalty (and overdispersion) parameter is fully taken into account (LPSMALA). The approach permits to obtain samples from the joint posterior of the spline vector and the penalty and overdispersion parameters. The latter can then be injected in functionals of the spline vector and hence obtain smooth estimates of the epidemic curve as well as the instantaneous reproduction number. Another advantage is that highest posterior density intervals can be easily calculated from LPSMALA.

## 4.1 Conditional posteriors for a “Metropolis-within-Gibbs”

### 4.1.1 Joint posterior of $(\zeta, \lambda)$

Let  $\zeta = (\boldsymbol{\theta}^\top, \rho)^\top$  be the  $(K + 1)$ -dimensional vector gathering the B-spline coefficients  $\boldsymbol{\theta}$  and the overdispersion parameter  $\rho$ . Using Bayes’ theorem, the joint posterior distribution for  $\zeta$ ,  $\lambda$  and  $\delta$  is:

$$\begin{aligned} p(\zeta, \lambda, \delta | \mathcal{D}) &= \frac{p(\mathcal{D} | \zeta, \lambda, \delta) p(\zeta, \lambda, \delta)}{p(\mathcal{D})} \\ &\propto \mathcal{L}(\zeta; \mathcal{D}) p(\boldsymbol{\theta} | \lambda) p(\lambda | \delta) p(\delta) p(\rho) \\ &\propto \exp(\ell(\zeta; \mathcal{D})) p(\boldsymbol{\theta} | \lambda) p(\lambda | \delta) p(\delta) p(\rho), \end{aligned} \tag{17}$$

The analytical formulas of the chosen priors are:

$$\begin{aligned}
 p(\boldsymbol{\theta}|\lambda) &\propto \lambda^{\frac{K}{2}} \exp(-0.5\lambda\boldsymbol{\theta}^\top P\boldsymbol{\theta}), \\
 p(\lambda|\delta) &\propto \delta^{\frac{\phi}{2}} \lambda^{\frac{\phi}{2}-1} \exp(-0.5\phi\delta\lambda), \\
 p(\delta) &\propto \delta^{a_\delta-1} \exp(-b_\delta\delta), \\
 p(\rho) &\propto \rho^{a_\rho-1} \exp(-b_\rho\rho).
 \end{aligned}$$

Injecting the above priors into (17) yields:

$$p(\boldsymbol{\zeta}, \lambda, \delta|\mathcal{D}) \propto \exp(\ell(\boldsymbol{\zeta}; \mathcal{D}) - b_\rho\rho - 0.5\lambda\boldsymbol{\theta}^\top P\boldsymbol{\theta}) \rho^{a_\rho-1} \lambda^{\frac{K+\phi}{2}-1} \delta^{\frac{\phi}{2}+a_\delta-1} \exp(-\delta(0.5\phi\lambda + b_\delta)). \quad (18)$$

#### 4.1.2 Conditional posteriors of $\boldsymbol{\zeta}$ , $\lambda$ and $\delta$

The following conditional posterior distributions can be directly obtained from (18):

$$p(\boldsymbol{\zeta}|\lambda, \delta, \mathcal{D}) \propto \rho^{a_\rho-1} \exp(\ell(\boldsymbol{\zeta}; \mathcal{D}) - b_\rho\rho - 0.5\lambda\boldsymbol{\theta}^\top P\boldsymbol{\theta}), \quad (19)$$

$$(\lambda|\boldsymbol{\zeta}, \delta, \mathcal{D}) \sim \mathcal{G}(0.5(K + \phi), 0.5(\boldsymbol{\theta}^\top P\boldsymbol{\theta} + \delta\phi)), \quad (20)$$

$$(\delta|\boldsymbol{\zeta}, \lambda, \mathcal{D}) \sim \mathcal{G}(0.5\phi + a_\delta, 0.5\phi\lambda + b_\delta). \quad (21)$$

## 4.2 Sampling from the joint posterior $p(\boldsymbol{\zeta}, \lambda, \delta|\mathcal{D})$

As the full conditionals  $p(\boldsymbol{\zeta}|\lambda, \delta, \mathcal{D})$ ,  $p(\lambda|\boldsymbol{\zeta}, \delta, \mathcal{D})$  and  $p(\delta|\boldsymbol{\zeta}, \lambda, \mathcal{D})$  are available, we follow a ‘‘Metropolis-within-Gibbs’’ strategy to sample the joint posterior  $p(\boldsymbol{\zeta}, \lambda, \delta|\mathcal{D})$ . In particular, the hyperparameters  $\lambda$  and  $\delta$  will be sampled in a Gibbs step, while  $\boldsymbol{\zeta}$  will be sampled using a modified Langevin-Hastings algorithm. This approach is presented in [Lambert and Eilers \(2009\)](#) in the context of Bayesian density estimation (see also [Lambert and Eilers \(2005\)](#) for the use of MALA in a proportional hazards model). We adapt the algorithm of the latter reference to our EpiLPS methodology by incorporating the added value of our Laplacian-P-splines method. In particular, the variance-covariance matrix in the Langevin diffusion will be replaced by the variance-covariance matrix of LPSMAP. The correlation structure borrowed from LPSMAP improves convergence and chain mixing.



### 4.3 The modified Metropolis-adjusted Langevin algorithm

In what follows, we prefer to work under the  $\log(\cdot)$  parameterization for  $\rho$ , i.e.  $w = \log(\rho)$  and denote by  $\tilde{\zeta} = (\boldsymbol{\theta}^\top, w)^\top$ , the  $(K + 1)$ -dimensional vector of B-spline amplitudes and overdispersion  $w$ . Under this parameterization, the conditional posterior of  $\tilde{\zeta}$  given  $\lambda$  and  $\delta$  can be obtained from (19) by using the transformation method of random variables:

$$p(\tilde{\zeta}|\lambda, \delta, \mathcal{D}) \propto \exp(w)^{a_\rho} \exp\left(\ell(\tilde{\zeta}; \mathcal{D}) - b_\rho \exp(w) - 0.5\lambda\boldsymbol{\theta}^\top P\boldsymbol{\theta}\right), \quad (22)$$

with the following log-likelihood under the reparameterization:

$$\begin{aligned} \ell(\tilde{\zeta}; \mathcal{D}) \doteq & \sum_{t=1}^T \left\{ \log \Gamma(y_t + \exp(w)) - \log \Gamma(\exp(w)) + y_t \boldsymbol{\theta}^\top b(t) + \exp(w)w \right. \\ & \left. - (y_t + \exp(w)) \log \left( \exp(\boldsymbol{\theta}^\top b(t)) + \exp(w) \right) \right\}. \end{aligned} \quad (23)$$

Let us denote by  $\tilde{\zeta}^{(m-1)} \in \mathbb{R}^{(K+1)}$  the state of the chain at iteration  $(m - 1)$ . In the Langevin-Hastings algorithms, the proposal for the vector  $\tilde{\zeta}$  at iteration  $m$  is a draw from the following multivariate Gaussian distribution:

$$\tilde{\zeta}^{(\text{prop})} \sim \mathcal{N}_{(K+1)}\left(\tilde{\zeta}^{(m-1)} + 0.5\varrho\Sigma_{LH}\nabla_{\tilde{\zeta}}\log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D})\Big|_{\tilde{\zeta}=\tilde{\zeta}^{(m-1)}}, \varrho\Sigma_{LH}\right), \quad (24)$$

where  $\varrho > 0$  is a tuning parameter that has to be carefully chosen in order to reach a desired acceptance rate and  $\Sigma_{LH}$  is the following block-diagonal variance-covariance matrix:

$$\Sigma_{LH} = \begin{pmatrix} \Sigma^* & 0 \\ 0 & 1 \end{pmatrix}, \quad (25)$$

where  $\Sigma^*$  is the  $K$ -dimensional covariance matrix obtained with LPSMAP. The gradient of  $\log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D}) = \ell(\tilde{\zeta}; \mathcal{D}) - 0.5\lambda\boldsymbol{\theta}^\top P\boldsymbol{\theta} - b_\rho \exp(w) + a_\rho w$  can be decomposed as follows:

$$\nabla_{\tilde{\zeta}} \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D}) = \left( \nabla_{\boldsymbol{\theta}} \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D}), \frac{\partial \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D})}{\partial w} \right)^\top, \quad (26)$$

and is analytically available (see [Appendix A](#) for more details). All the quantities related to the Langevin-Hastings proposal have been analytically derived, so that the draw in (24) can be obtained (for a given value of  $\lambda$  and  $\delta$ ). As in a classic MH algorithm, the next step consists in computing the acceptance probability:

$$\pi\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right) = \min\left\{1, \frac{p(\tilde{\zeta}^{(\text{prop})}|\lambda, \delta, \mathcal{D}) q\left(\tilde{\zeta}^{(\text{prop})}, \tilde{\zeta}^{(m-1)}\right)}{p(\tilde{\zeta}^{(m-1)}|\lambda, \delta, \mathcal{D}) q\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right)}\right\}, \quad (27)$$

where  $q(\cdot, \cdot)$  denotes the (Gaussian) proposal distribution and  $p(\cdot|\lambda, \delta, \mathcal{D})$  the target (conditional) posterior distribution. Finally, we generate a uniform random variable  $u \sim \mathcal{U}(0, 1)$  and accept the proposed vector  $\tilde{\zeta}^{(\text{prop})}$  if  $u \leq \pi\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right)$  and reject it otherwise. While iterating through the Metropolis-within-Gibbs algorithm, the tuning parameter  $\varrho$  is automatically adapted to reach the optimal acceptance rate of 0.57 ([Haario et al., 2001](#); [Atchadé and Rosenthal, 2005](#); [Roberts and Rosenthal, 1998](#)). The pseudo-code below summarizes the LPSMALA algorithm.

---

**LPSMALA algorithm to sample the posterior  $p(\boldsymbol{\theta}, \rho, \lambda, \delta|\mathcal{D})$ .**

---

1: Fix initial values  $m = 0$ ,  $\lambda^{(0)}$ ,  $\delta^{(0)}$ ,  $\varrho^{(0)}$  and  $\tilde{\zeta}^{(0)} = (\boldsymbol{\theta}^{(0)\top}, w^{(0)\top})^\top$ .

2: **for**  $m = 1, \dots, M$  **do**

3:     (*Langevin-Hastings*)

4:     Compute Langevin diffusion:  $\mathcal{E}\left(\tilde{\zeta}^{(m-1)}\right) = \tilde{\zeta}^{(m-1)} + 0.5\varrho^{(m-1)}\Sigma_{LH}\nabla_{\tilde{\zeta}}\log p(\tilde{\zeta}|\lambda^{(m-1)}, \delta^{(m-1)}, \mathcal{D})|_{\tilde{\zeta}=\tilde{\zeta}^{(m-1)}}$ .

5:     Generate a proposal:  $\tilde{\zeta}^{(\text{prop})} \sim \mathcal{N}_{(K+1)}\left(\mathcal{E}\left(\tilde{\zeta}^{(m-1)}\right), \varrho^{(m-1)}\Sigma_{LH}\right)$ .

6:     Compute acceptance probability  $\pi\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right) = \min\left\{1, \frac{p(\tilde{\zeta}^{(\text{prop})}|\lambda^{(m-1)}, \delta^{(m-1)}, \mathcal{D}) q\left(\tilde{\zeta}^{(\text{prop})}, \tilde{\zeta}^{(m-1)}\right)}{p(\tilde{\zeta}^{(m-1)}|\lambda^{(m-1)}, \delta^{(m-1)}, \mathcal{D}) q\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right)}\right\}$ .

7:     Draw  $u \sim \mathcal{U}(0, 1)$ .

8:     **if**  $u \leq \pi$  **set**  $\tilde{\zeta}^{(m)} = \tilde{\zeta}^{(\text{prop})}$  (accept), **else**  $\tilde{\zeta}^{(m)} = \tilde{\zeta}^{(m-1)}$  (reject).

9:     (*Gibbs sampler*)

10:     Draw  $\delta^{(m)} \sim \mathcal{G}\left(0.5\phi + a_\delta, 0.5\phi\lambda^{(m-1)} + b_\delta\right)$ ,

11:     Draw  $\lambda^{(m)} \sim \mathcal{G}\left(0.5(K + \phi), 0.5(\boldsymbol{\theta}^{(m)\top} P \boldsymbol{\theta}^{(m)} + \delta^{(m)}\phi)\right)$ .

12:     (*Adaptive tuning*)

13:     Update  $\varrho^{(m)} = \mathcal{H}^2\left(\sqrt{\varrho^{(m-1)}} + m^{-1}\left(\pi\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right) - 0.57\right)\right)$ .

14: **end for**

---

The adaptive tuning part (on line 13) involves the step function  $\mathcal{H}(z) = \epsilon \mathbb{I}(z < \epsilon) + z \mathbb{I}(\epsilon \leq z \leq \mathcal{A}) + \mathcal{A} \mathbb{I}(z > \mathcal{A})$ , with  $\epsilon = 10^{-4}$  and  $\mathcal{A} = 10^4$ , see [Lambert and Eilers \(2009\)](#) for details. Finally, the ratio  $q\left(\tilde{\zeta}^{(\text{prop})}, \tilde{\zeta}^{(m-1)}\right) q^{-1}\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right)$  entering the computation of the acceptance probability (line 6) is derived in [Appendix B](#).

## 4.4 Posterior inference

Provided the LPSMALA algorithm (cf. Section 4.3) is iterated long enough, say after  $\widetilde{M}$  iterations, MCMC theory certifies that  $\mathcal{S} = \left\{ \left( \boldsymbol{\theta}^{(m)\top}, \rho^{(m)}, \lambda^{(m)}, \delta^{(m)} \right) \right\}_{m=\widetilde{M}+1}^M$  can be viewed as random draws from the target posterior distribution  $p(\boldsymbol{\theta}, \rho, \lambda, \delta | \mathcal{D})$ . Note that a convenient starting point for the initial values of the parameters might be to fix them at their LPSMAP estimate. Given the sample  $\mathcal{S}$ , inference on quantities that are functions of  $\boldsymbol{\theta}$  becomes straightforward in the sense that point estimates and credibility envelopes can be easily obtained. A point estimate for the mean number of incidence counts at time  $t$  is taken to be the posterior mean (after discarding the burn-in phase):

$$\widehat{\mu}(t) = \frac{1}{M - \widetilde{M}} \sum_{m=\widetilde{M}+1}^M \exp\left(\boldsymbol{\theta}^{(m)\top} b(t)\right). \quad (28)$$

Note also that  $\mathcal{S}$  can be used to compute highest posterior density intervals of  $\mu(t)$  at any point  $t$ . Using the renewal equation and the MCMC sample, one can apply the ‘‘plug-in’’ estimate of Section 3.1 and recover the following estimate of the instantaneous reproduction number at any given point  $t$ :

$$\begin{aligned} \widehat{R}(t) &= \frac{1}{M - \widetilde{M}} \sum_{m=\widetilde{M}+1}^M \exp\left(\boldsymbol{\theta}^{(m)\top} b(t)\right) \left\{ \mathbb{I}(t = 1) + \left( \sum_{s=1}^{t-1} \varphi_s \exp\left(\boldsymbol{\theta}^{(m)\top} b(t-s)\right) \right)^{-1} \mathbb{I}(2 \leq t \leq k) \right. \\ &\quad \left. + \left( \sum_{s=1}^k \varphi_s \exp\left(\boldsymbol{\theta}^{(m)\top} b(t-s)\right) \right)^{-1} \mathbb{I}(k < t \leq T) \right\}. \end{aligned} \quad (29)$$

Also, using  $\mathcal{S}$ , one can compute a highest posterior density interval of  $R(t)$  at any point  $t$ .

## 5 Numerical study

In this section, we implement a numerical study with four underlying epidemic scenarios to assess the accuracy with which EpiLPS is able to “track” the target reproduction number over time. EpiLPS results are compared with the `estimate.R()` routine of the EpiEstim package using two sliding window options (the default weekly window of 7 days and a shorter window of 3 days). In each scenario, we simulate incidence data for an epidemic lasting 50 days according to a Poisson data generating process with mean counts governed by the renewal equation (see Eq.10). In total,  $S = 200$  epidemics are simulated for each scenario and a cubic B-spline basis of size  $K = 20$  with a second order penalty is used for inference in EpiLPS. In Scenario 1, a constant instantaneous reproduction number at  $R(t) = 2.5$  is considered. In Scenario 2, an intervention strategy is replicated, so that  $R(t) = 2.5$  and a sudden drop to  $R(t) = 0.7$  occurs at day  $t = 20$ . The latter scenario allows to check whether EpiLPS is able to quickly react to such a brutal change in  $R(t)$ . Scenario 3 implements a more wiggly structure for  $R(t)$  and finally, Scenario 4 considers the case of a vanishing epidemic with a monotonic decreasing reproduction number. Table 1 summarizes the target  $R(t)$  functions and the serial interval distribution for the considered scenarios. Table 2 reports for each scenario the average Bias, MSE, coverage of 90% and 95% credible envelopes and credible interval width for the  $R(t)$  estimator with EpiLPS (LPSMAP and LSPMALA) and EpiEstim. The average is computed over days  $t = 8, \dots, 50$ . For LSPMALA, a chain of length 5,000 and a warm up period of 2,000 is specified.

Table 1: Different scenarios for the simulation of epidemic data based on a Poisson data generating process for the incidence time series over  $T = 50$  days.

Scenario	Target $R(t)$	Serial interval distribution
1	$R(t) = 2.5$	$\varphi = (0.05, 0.05, 0.1, 0.2, 0.2, 0.1, 0.1, 0.1, 0.1)$
2	$R(t) = 2.5 \mathbb{I}(t < 20) + 0.7 \mathbb{I}(t \geq 20)$	$\varphi = (0.15, 0.25, 0.35, 0.13, 0.12)$
3	$R(t) = 0.4 + \exp(\cos(t/5))$	$\varphi = (0.13, 0.22, 0.18, 0.15, 0.07, 0.13, 0.12)$
4	$R(t) = \exp(\cos(t/15))$	$\varphi = (0.32, 0.24, 0.15, 0.09, 0.2)$

Table 2: Numerical results for EpiLPS and EpiEstim under different simulation scenarios. The Bias, MSE, 90% and 95% coverage of credible intervals and 95% CI width are averaged over days  $t = 8, \dots, 50$ .

Scenario	Method	Bias	MSE	CP <sub>90%</sub>	CP <sub>95%</sub>	CI <sub>95%</sub> width
Scenario 1 “Constant $R(t)$ ”	LPSMAP	-0.026	0.018	98.756	99.407	0.946
	LPSMALA	-0.026	0.020	93.535	95.977	0.351
	EpiEstim (7d window)	-0.003	0.016	90.291	95.221	0.335
	EpiEstim (3d window)	0.003	0.036	90.444	95.567	0.499
Scenario 2 “Step $R(t)$ ”	LPSMAP	-0.008	0.038	89.337	90.651	0.395
	LPSMALA	-0.005	0.030	90.384	90.814	0.328
	EpiEstim (7d window)	0.085	0.071	76.733	81.302	0.123
	EpiEstim (3d window)	0.030	0.036	86.000	90.691	0.253
Scenario 3 “Wave $R(t)$ ”	LPSMAP	-0.007	0.009	97.372	98.895	0.530
	LPSMALA	0.000	0.009	92.058	95.919	0.329
	EpiEstim (7d window)	0.051	0.145	19.895	24.395	0.321
	EpiEstim (3d window)	0.039	0.050	60.957	67.638	0.532
Scenario 4 “Decaying $R(t)$ ”	LPSMAP	0.000	0.001	98.477	99.465	0.257
	LPSMALA	-0.001	0.001	90.767	95.419	0.112
	EpiEstim (7d window)	0.125	0.021	17.988	20.616	0.097
	EpiEstim (3d window)	0.046	0.010	46.436	51.819	0.214

For the first scenario, EpiEstim exhibits slightly lower bias than EpiLPS and the MSE is more or less of the same magnitude for both approaches. A general pattern that is apparent from Table 2 is that EpiLPS tends to overcover with LPSMAP (except for Scenario 2), while LPSMALA is closer to the nominal value in all scenarios. This phenomenon can be attributed to the following two main sources of approximations inherent to LPSMAP: (1) the Laplace approximation to the conditional posterior of the B-spline amplitudes and (2) the *maximum a posteriori* estimate of the hyperparameter vector that ignores the uncertainty related to the estimation of the roughness penalty parameter. In Scenario 2, EpiLPS has smaller bias than EpiEstim and the latter approach also suffers from mild undercoverage with a weekly sliding window. In Scenarios 3 and 4, where the target  $R(t)$  curve is less linear than in the previous scenarios, a more pronounced difference between EpiEstim and EpiLPS appears. EpiLPS shows better performance, especially in terms of coverage of credible intervals where EpiEstim exhibits severe undercoverage. Even when decreasing the sliding window to 3 days, the coverage probability is far from the nominal value.

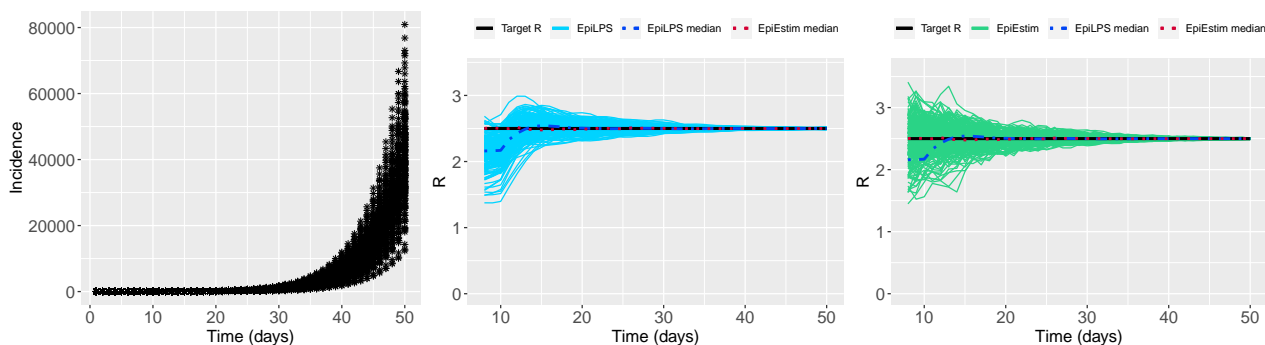


Figure 1: (Left) Simulated incidence data for Scenario 1. (Center) Estimated trajectories of  $R(t)$  for each generated dataset with EpiLPS (LPSMAP). (Right) Estimated trajectories of  $R(t)$  for each generated dataset with EpiEstim and a weekly sliding window. The pointwise median estimate of  $R(t)$  for EpiLPS (dashed) and EpiEstim (dotted) is also shown.

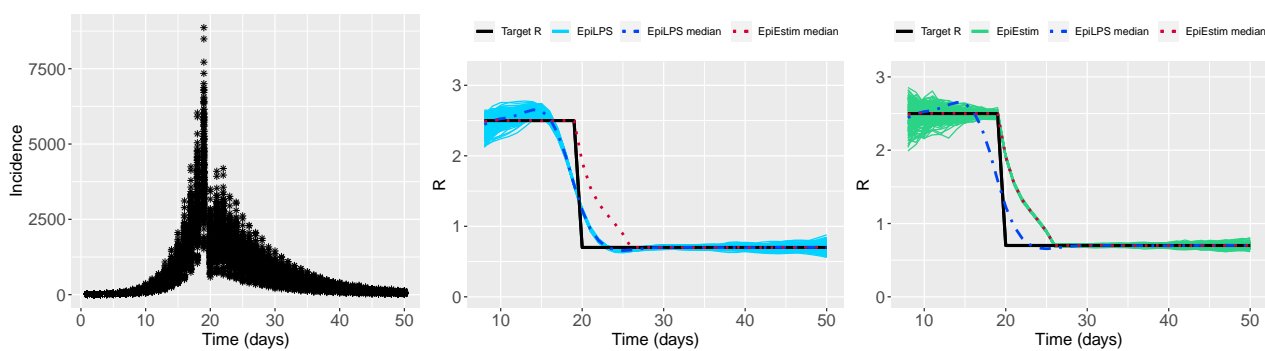


Figure 2: (Left) Simulated incidence data for Scenario 2. (Center) Estimated trajectories of  $R(t)$  for each generated dataset with EpiLPS (LPSMAP). (Right) Estimated trajectories of  $R(t)$  for each generated dataset with EpiEstim and a weekly sliding window. The pointwise median estimate of  $R(t)$  for EpiLPS (dashed) and EpiEstim (dotted) is also shown.

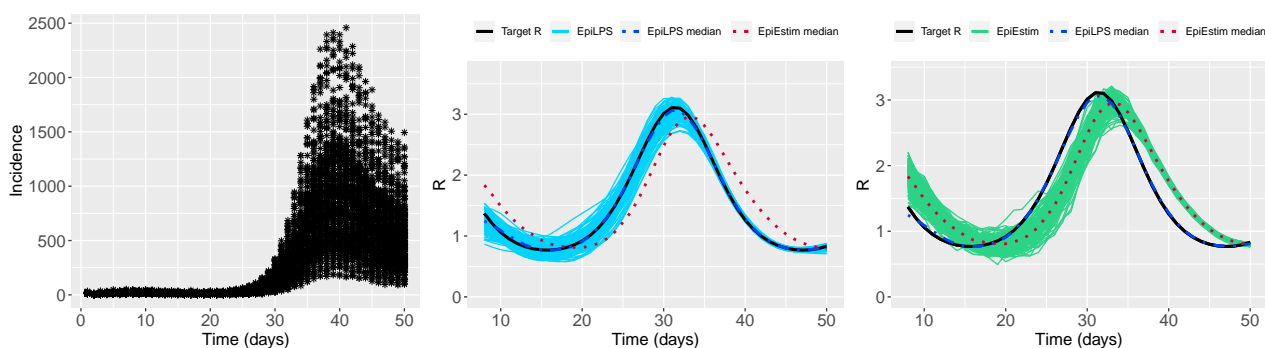


Figure 3: (Left) Simulated incidence data for Scenario 3. (Center) Estimated trajectories of  $R(t)$  for each generated dataset with EpiLPS (LPSMAP). (Right) Estimated trajectories of  $R(t)$  for each generated dataset with EpiEstim and a weekly sliding window. The pointwise median estimate of  $R(t)$  for EpiLPS (dashed) and EpiEstim (dotted) is also shown.

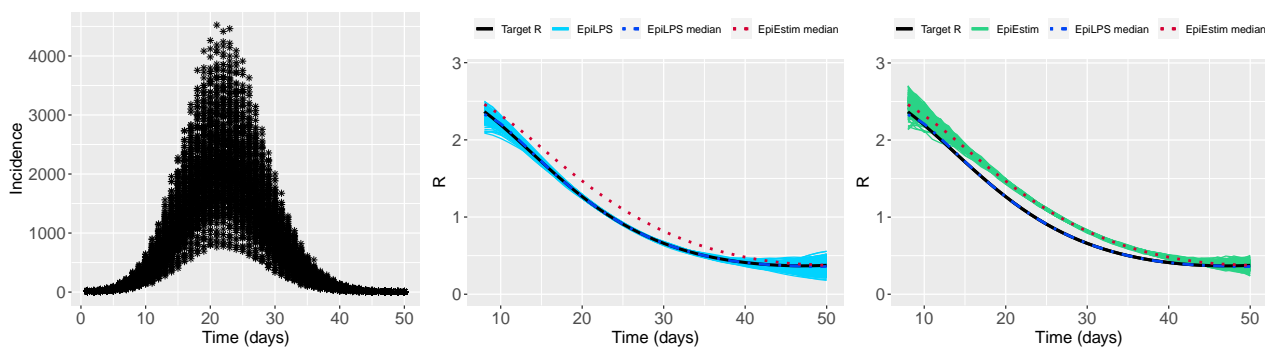


Figure 4: (Left) Simulated incidence data for Scenario 4. (Center) Estimated trajectories of  $R(t)$  for each generated dataset with EpiLPS (LPSMAP). (Right) Estimated trajectories of  $R(t)$  for each generated dataset with EpiEstim and a weekly sliding window. The pointwise median estimate of  $R(t)$  for EpiLPS (dashed) and EpiEstim (dotted) is also shown.

Figures 1-4 show the simulated incidence data together with the estimated trajectories of  $R(t)$  with EpiLPS (using LPSMAP) and EpiEstim respectively. Under linear configurations of  $R(t)$ , the two competing approaches have similar behavior, while for curved  $R(t)$  functions, EpiEstim estimates (with a weekly sliding window) are shifted as compared to the target. With EpiEstim, the magnitude of this shift can be controlled by decreasing the sliding window, yet the optimal *a priori* choice of a smoothing window is far from being an easy task in practice and thus alleviating the window size assumption (as in EpiLPS) can be of interest.

## 6 Application to observed case counts in infectious disease epidemics

### 6.1 Epidemics of SARS-CoV-1 and influenza A H1N1

In this section, the LPSMALA algorithm is applied on two historical outbreak datasets presented in [Cori et al. \(2013\)](#). In particular, we consider the 2003 SARS outbreak in Hong Kong and the 2009 pandemic influenza in a school in Pennsylvania (USA). We use  $K = 40$  B-splines with a second-order penalty and the serial interval distributions provided in the EpiEstim package [Cori \(2021\)](#). The LPSMALA algorithm is implemented with a

chain of length 25,000. Acceptance rates for the generated chains are close to the optimal value of 57% and the posterior samples have converged according to the Geweke (1992) diagnostic test (at the 1% level of significance). Figure 5 shows the smoothed epidemic curves and the estimated  $R(t)$  for the two outbreaks. Results for the SARS data show that the reproduction number reaches a first peak during the third week where  $\hat{R}(t) = 9.66$  (95% CI : 5.12 – 16.82) and a second more moderate peak around week 6 with  $\hat{R}(t) = 2.79$  (95% CI : 1.97 – 3.81). After day  $t = 43$ , the epidemic is under control and  $R(t)$  smoothly decays below 1. For the pandemic influenza in Pennsylvania, in the end of the second week  $R(t)$  is around 2.04 (95% CI : 1.24 – 3.07). During the middle of the third week, the situation is less severe and  $R(t)$  points below 1. As noted in Cori et al. (2013), a few cases appeared in the last days of the epidemic generating an upward trend in  $R(t)$  estimates.

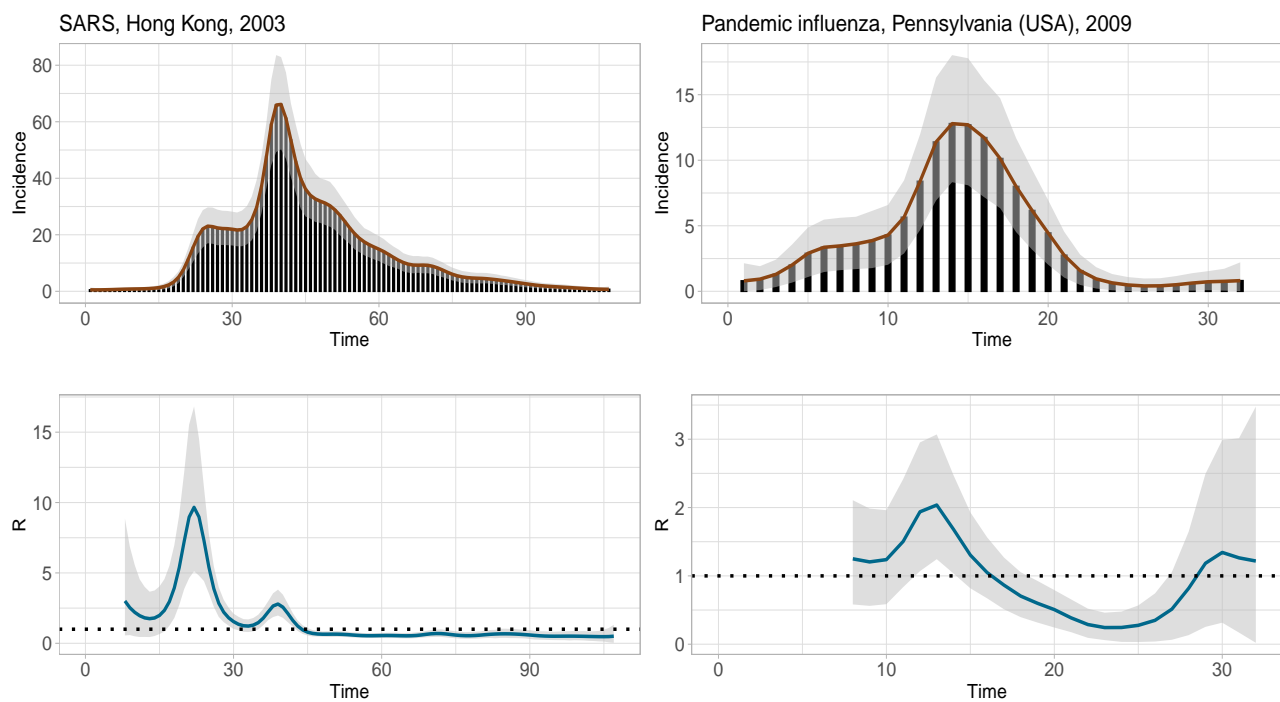


Figure 5: (Left column) EpiLPS fit for the epidemic curve (top) and the instantaneous reproduction number  $R(t)$  (bottom) of the SARS outbreak in Hong Kong, 2003. (Right column) EpiLPS fit for the epidemic curve (top) and the instantaneous reproduction number  $R(t)$  (bottom) of the pandemic influenza in Pennsylvania, 2009.



## 6.2 Application on the SARS-CoV-2 pandemic

The EpiLPS methodology is illustrated on the SARS-CoV-2 pandemic using publicly available data from the Covid-19 Data Hub ([Guidotti and Ardia, 2020](https://covid19datahub.com/)) and its associated COVID19 package on CRAN (<https://cran.r-project.org/package=COVID19>). Country-level data on hospitalizations for Belgium, Denmark, Portugal and France from April, 5th 2020 to October 31st, 2021 is used with a uniform serial interval distribution over five days, i.e.  $\varphi = (0.2, 0.2, 0.2, 0.2, 0.2)$ . In Figure 6, the estimated reproduction number over the considered period is shown for the four countries. Results are obtained with the LPSMAP algorithm using  $K = 30$  B-splines and a second order penalty. From a computational perspective, it takes less than 3 seconds to fit the EpiLPS model for the four countries with LPSMAP. The fitted reproduction numbers reflect the different waves of the COVID-19 pandemic and the recent rise in infections in the beginning of September 2021.

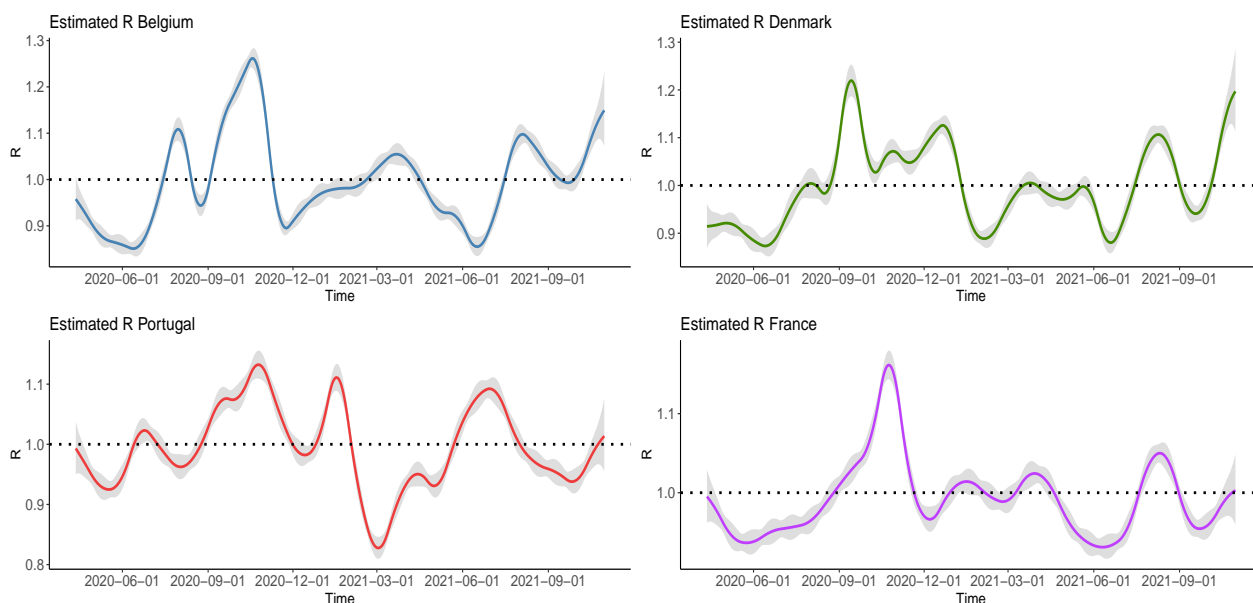


Figure 6: Estimated reproduction number from 2020-04-05 to 2021-10-31 for Belgium, Denmark, Portugal and France using a five days serial interval and LPSMAP with  $K = 30$  B-splines and a second order penalty. The gray shaded surface corresponds to the 95% (approximate) credible interval.

## 7 Conclusion

EpiLPS (an acronym for **E**pidemiological modelling with **L**aplacian-**P**-**S**plines) is a fast and flexible tool for near real-time estimation of the instantaneous reproduction number  $R(t)$  during epidemic outbreaks. The tool is flexible in the sense that (penalized) spline based approximations provide smoothed estimates of  $R(t)$  with little computational effort and without the constraint of imposing any sliding window assumption that could potentially affect the timing and accuracy of the estimator. Moreover, the end user has the choice between a fully sampling-free approach (LPSMAP) or an efficient MCMC-based approach with Langevin diffusions (LPSMALA) for inference. The available EpiLPS package (<https://github.com/oswaldogressani>) allows public health policy makers to analyze incoming data faster than existing methods relying on classic MCMC samplers, thus allowing them to be better informed when taking decisions on control measures for the ongoing SARS-CoV-2 epidemic. Simulation studies in this manuscript provide encouraging results and support EpiLPS as being a robust tool capable of a precise tracking of  $R(t)$  over time. The available EpiLPS software package is also straightforward to use with simple function calls that are specifically interesting for routine usage.

The EpiLPS project opens up several future research directions. A possible extension would be to formulate the EpiLPS model within a zero-inflated Poisson framework to cope with incidence time series characterized by an excess of zero counts. Another interesting extension would be to adapt the model to account for regional variation and imported cases.

## Software

Simulation results and real data applications in this manuscript can be fully reproduced with the EpiLPS package available here <https://github.com/oswaldogressani>.

## Conflict of interest

The authors declare no conflicts of interest.

## Acknowledgments

This project is funded by the European Union's Research and Innovation Action under the H2020 work programme, EpiPose (grant number 101003688).

## Appendix A

In this appendix, we derive the analytical expressions of the gradient:

$$\nabla_{\tilde{\zeta}} \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D}) = \left( \nabla_{\theta} \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D}), \frac{\partial \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D})}{\partial w} \right)^{\top},$$

where the target function is  $\log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D}) = \ell(\tilde{\zeta}; \mathcal{D}) - 0.5\lambda\theta^{\top}P\theta - b_{\rho}\exp(w) + a_{\rho}w$ . Let us first concentrate on the partial derivatives with respect to the spline components:

$$\nabla_{\theta} \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D}) = \nabla_{\theta} \ell(\tilde{\zeta}; \mathcal{D}) - \lambda P\theta. \quad (30)$$

As already shown in Section 2.2, the gradient for the log-likelihood  $\nabla_{\theta} \ell(\tilde{\zeta}; \mathcal{D})$  is:

$$\frac{\partial \ell(\tilde{\zeta}; \mathcal{D})}{\partial \theta_k} = \sum_{t=1}^T y_t b_k(t) - \sum_{t=1}^T \frac{(y_t + \exp(w)) \exp(\theta^{\top} b(t))}{(\exp(\theta^{\top} b(t)) + \exp(w))} b_k(t), \quad k = 1, \dots, K. \quad (31)$$

The last term to be computed to recover the full gradient is:

$$\frac{\partial \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D})}{\partial w} = \frac{\partial \ell(\tilde{\zeta}; \mathcal{D})}{\partial w} - b_{\rho} \exp(w) + a_{\rho}, \quad (32)$$

where

$$\begin{aligned} \frac{\partial \ell(\tilde{\zeta}; \mathcal{D})}{\partial w} &= \sum_{t=1}^T \left\{ \underbrace{\frac{\partial \log \Gamma(y_t + \exp(w))}{\partial w}}_{\text{Term I}} - \underbrace{\frac{\partial \log \Gamma(\exp(w))}{\partial w}}_{\text{Term II}} + \underbrace{\frac{\partial \exp(w)w}{\partial w}}_{\text{Term III}} \right. \\ &\quad \left. - \underbrace{\frac{\partial}{\partial w} (y_t + \exp(w)) \log \left( \exp(\theta^{\top} b(t)) + \exp(w) \right)}_{\text{Term IV}} \right\}. \end{aligned} \quad (33)$$

For Term I, using the chain rule, one recovers:

$$\begin{aligned} \frac{\partial \log \Gamma(y_t + \exp(w))}{\partial w} &= \frac{\partial \log \Gamma(y_t + \exp(w))}{\partial \Gamma(y_t + \exp(w))} \frac{\partial \Gamma(y_t + \exp(w))}{\partial (y_t + \exp(w))} \frac{\partial (y_t + \exp(w))}{\partial w} \\ &= \frac{\Gamma'(y_t + \exp(w))}{\Gamma(y_t + \exp(w))} \exp(w) \\ &= \psi(y_t + \exp(w)) \exp(w), \end{aligned} \quad (34)$$

where  $\psi(\cdot)$  is the digamma function. Using the same chain rule argument, one can easily show that for Term II:

$$\frac{\partial \log \Gamma(\exp(w))}{\partial w} = \psi(\exp(w)) \exp(w). \quad (35)$$

Term III is also trivial:

$$\frac{\partial \exp(w)w}{\partial w} = \exp(w)(1 + w). \quad (36)$$

Term IV is as follows:

$$\begin{aligned} \frac{\partial}{\partial w} (y_t + \exp(w)) \log \left( \exp(\boldsymbol{\theta}^\top b(t)) + \exp(w) \right) &= \exp(w) \log \left( \exp(\boldsymbol{\theta}^\top b(t)) + \exp(w) \right) \\ &\quad + (y_t + \exp(w)) \exp(w) \left( \exp(\boldsymbol{\theta}^\top b(t)) + \exp(w) \right)^{-1} \\ &= \exp(w) \left\{ \log \left( \exp(\boldsymbol{\theta}^\top b(t)) + \exp(w) \right) \right. \\ &\quad \left. + (y_t + \exp(w)) \left( \exp(\boldsymbol{\theta}^\top b(t)) + \exp(w) \right)^{-1} \right\}. \quad (37) \end{aligned}$$

Gathering all the above intermediate results, we obtain the following derivative for (32):

$$\begin{aligned} \frac{\partial \log p(\tilde{\zeta} | \lambda, \delta, \mathcal{D})}{\partial w} &= \sum_{t=1}^T \left\{ \exp(w) \left[ \psi(y_t + \exp(w)) - \psi(\exp(w)) + (1 + w) - \log \left( \exp(\boldsymbol{\theta}^\top b(t)) + \exp(w) \right) \right. \right. \\ &\quad \left. \left. - (y_t + \exp(w)) \left( \exp(\boldsymbol{\theta}^\top b(t)) + \exp(w) \right)^{-1} \right] \right\} - b_\rho \exp(w) + a_\rho. \quad (38) \end{aligned}$$

## Appendix B

To determine the analytical form of the ratio of proposal distributions in the Metropolis-within-Gibbs algorithm, let us use the compact notation  $\mathbf{Y}_{(m-1)} = \nabla_{\tilde{\zeta}} \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D})|_{\tilde{\zeta}=\tilde{\zeta}^{(m-1)}}$ ,  $\mathbf{Y}_{(\text{prop})} = \nabla_{\tilde{\zeta}} \log p(\tilde{\zeta}|\lambda, \delta, \mathcal{D})|_{\tilde{\zeta}=\tilde{\zeta}^{(\text{prop})}}$  and define the difference  $\Delta\tilde{\zeta} = \tilde{\zeta}^{(\text{prop})} - \tilde{\zeta}^{(m-1)}$ , so that:

$$q\left(\tilde{\zeta}^{(\text{prop})}, \tilde{\zeta}^{(m-1)}\right) = (2\pi)^{-\frac{(K+1)}{2}} |\varrho \Sigma_{LH}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\varrho} \left(\tilde{\zeta}^{(m-1)} - \tilde{\zeta}^{(\text{prop})} - 0.5\varrho \Sigma_{LH} \mathbf{Y}_{(\text{prop})}\right)^\top \Sigma_{LH}^{-1} \left(\tilde{\zeta}^{(m-1)} - \tilde{\zeta}^{(\text{prop})} - 0.5\varrho \Sigma_{LH} \mathbf{Y}_{(\text{prop})}\right)\right\}. \quad (39)$$

The kernel of (39) is thus:

$$\begin{aligned} \ker\left(q\left(\tilde{\zeta}^{(\text{prop})}, \tilde{\zeta}^{(m-1)}\right)\right) &= \exp\left\{-\frac{1}{2\varrho} \left(-(\Delta\tilde{\zeta})^\top - 0.5\varrho \Sigma_{LH} \mathbf{Y}_{(\text{prop})}\right)^\top \Sigma_{LH}^{-1} \left(-(\Delta\tilde{\zeta})^\top - 0.5\varrho \Sigma_{LH} \mathbf{Y}_{(\text{prop})}\right)\right\} \\ &= \exp\left\{-\frac{1}{2\varrho} \left[ \left(-(\Delta\tilde{\zeta})^\top \Sigma_{LH}^{-1} - 0.5\varrho \mathbf{Y}_{(\text{prop})}^\top \Sigma_{LH} \Sigma_{LH}^{-1}\right) \left(-(\Delta\tilde{\zeta})^\top - 0.5\varrho \Sigma_{LH} \mathbf{Y}_{(\text{prop})}\right) \right]\right\} \\ &= \exp\left\{-\frac{1}{2\varrho} \left[ (\Delta\tilde{\zeta})^\top \Sigma_{LH}^{-1} (\Delta\tilde{\zeta})^\top + \varrho \mathbf{Y}_{(\text{prop})}^\top (\Delta\tilde{\zeta}) + \frac{\varrho^2}{4} \mathbf{Y}_{(\text{prop})}^\top \Sigma_{LH} \mathbf{Y}_{(\text{prop})} \right]\right\} \\ &= \exp\left\{-\frac{1}{2\varrho} (\Delta\tilde{\zeta})^\top \Sigma_{LH}^{-1} (\Delta\tilde{\zeta})^\top - \frac{1}{2} \mathbf{Y}_{(\text{prop})}^\top (\Delta\tilde{\zeta}) - \frac{\varrho}{8} \mathbf{Y}_{(\text{prop})}^\top \Sigma_{LH} \mathbf{Y}_{(\text{prop})}\right\}. \end{aligned}$$

Using a similar argument, one can show that the kernel of  $q\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right)$  is:

$$\ker\left(q\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right)\right) = \exp\left\{-\left[\frac{1}{2\varrho} (\Delta\tilde{\zeta})^\top \Sigma_{LH}^{-1} (\Delta\tilde{\zeta})^\top - \frac{1}{2} \mathbf{Y}_{(m-1)}^\top (\Delta\tilde{\zeta}) + \frac{\varrho}{8} \mathbf{Y}_{(m-1)}^\top \Sigma_{LH} \mathbf{Y}_{(m-1)}\right]\right\}.$$

This can be used to compute the ratio:

$$\begin{aligned} \frac{q\left(\tilde{\zeta}^{(\text{prop})}, \tilde{\zeta}^{(m-1)}\right)}{q\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right)} &= \ker\left(q\left(\tilde{\zeta}^{(\text{prop})}, \tilde{\zeta}^{(m-1)}\right)\right) \ker^{-1}\left(q\left(\tilde{\zeta}^{(m-1)}, \tilde{\zeta}^{(\text{prop})}\right)\right) \\ &= \exp\left\{-\frac{1}{2} \mathbf{Y}_{(\text{prop})}^\top (\Delta\tilde{\zeta}) - \frac{\varrho}{8} \mathbf{Y}_{(\text{prop})}^\top \Sigma_{LH} \mathbf{Y}_{(\text{prop})} - \frac{1}{2} \mathbf{Y}_{(m-1)}^\top (\Delta\tilde{\zeta}) + \frac{\varrho}{8} \mathbf{Y}_{(m-1)}^\top \Sigma_{LH} \mathbf{Y}_{(m-1)}\right\} \\ &= \exp\left\{-\frac{1}{2} (\mathbf{Y}_{(\text{prop})} + \mathbf{Y}_{(m-1)})^\top (\Delta\tilde{\zeta}) - \frac{\varrho}{8} \left[ (\mathbf{Y}_{(\text{prop})} + \mathbf{Y}_{(m-1)})^\top \Sigma_{LH} (\mathbf{Y}_{(\text{prop})} - \mathbf{Y}_{(m-1)}) \right]\right\} \\ &= \exp\left\{-\frac{1}{2} (\mathbf{Y}_{(\text{prop})} + \mathbf{Y}_{(m-1)})^\top \left(\Delta\tilde{\zeta} + \frac{\varrho \Sigma_{LH}}{4} (\mathbf{Y}_{(\text{prop})} - \mathbf{Y}_{(m-1)})\right)\right\}. \quad (40) \end{aligned}$$

## References

- Abbott, S., Hellewell, J., Sherratt, K., Gostic, K., Hickson, J., Badr, H. S., DeWitt, M., Thompson, R., EpiForecasts, and Funk, S. (2020). *EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters*.
- Anscombe, F. J. (1950). Sampling theory of the Negative Binomial and logarithmic series distributions. *Biometrika*, 37(3/4):358–382. <http://www.jstor.org/stable/2332388>.
- Atchadé, Y. F. and Rosenthal, J. S. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, 11(5):815–828. <https://doi.org/10.3150/bj/1130077595>.
- Azmon, A., Faes, C., and Hens, N. (2014). On the estimation of the reproduction number based on misreported epidemic data. *Statistics in medicine*, 33(7):1176–1192.
- Chib, S. and Greenberg, E. (1996). Markov chain Monte Carlo simulation methods in Econometrics. *Econometric Theory*, 12(3):409–431. <http://www.jstor.org/stable/3532527>.
- Cori, A. (2021). EpiEstim: estimate time varying reproduction numbers from epidemic curves. <https://cran.r-project.org/web/packages/EpiEstim/index.html>.
- Cori, A., Ferguson, N. M., Fraser, C., and Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9):1505–1512. <https://doi.org/10.1093/aje/kwt133>.
- Eddelbuettel, D., François, R., Allaire, J., Ushey, K., Kou, Q., Russel, N., Chambers, J., and Bates, D. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science*, 11(2):89–121. <https://doi.org/10.1214/ss/1038425655>.
- Eilers, P. H. C. and Marx, B. D. (2021). *Practical Smoothing: The Joys of P-splines*. Cambridge University Press.

- Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PloS one*, 2(8):e758. <https://doi.org/10.1371/journal.pone.0000758>.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, PAMI-6(6):721–741. <https://doi.org/10.1109/TPAMI.1984.4767596>.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, 4:169–193.
- Gostic, K. M., McGough, L., Baskerville, E. B., Abbott, S., Joshi, K., Tedijanto, C., Kahn, R., Niehus, R., Hay, J. A., De Salazar, P. M., Hellewell, J., Meakin, S., Munday, J. D., Bosse, N. I., Sherratt, K., Thompson, R. N., White, L. F., Huisman, J. S., Scire, J., Bonhoeffer, S., Stadler, T., Wallinga, J., Funk, S., Lipsitch, M., and Cobey, S. (2020). Practical considerations for measuring the effective reproductive number,  $R_t$ . *PLOS Computational Biology*, 16(12):1–21.
- Gressani, O., Faes, C., and Hens, N. (2021). An approximate Bayesian approach for estimation of the reproduction number under misreported epidemic data. *MedRxiv*. <https://doi.org/10.1101/2021.05.19.21257438>.
- Gressani, O. and Lambert, P. (2021). Laplace approximations for fast bayesian inference in generalized additive models based on P-splines. *Computational Statistics & Data Analysis*, 154:107088. <https://doi.org/10.1016/j.csda.2020.107088>.
- Guidotti, E. and Ardia, D. (2020). Covid-19 data hub. *Journal of Open Source Software*, 5(51):2376.
- Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242. <https://www.jstor.org/stable/3318737>.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109. <https://doi.org/10.1093/biomet/57.1.97>.



- Jullion, A. and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Computational Statistics & Data Analysis*, 51(5):2542–2558.
- Lambert, P. and Eilers, P. H. C. (2005). Bayesian proportional hazards model with time-varying regression coefficients: A penalized Poisson regression approach. *Statistics in Medicine*, 24(24):3977–3989. <https://doi.org/10.1002/sim.2396>.
- Lambert, P. and Eilers, P. H. C. (2009). Bayesian density estimation from grouped continuous data. *Computational Statistics & Data Analysis*, 53(4):1388–1399. <https://doi.org/10.1016/j.csda.2008.11.022>.
- Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212. <https://doi.org/10.1198/1061860043010>.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092. <https://doi.org/10.1063/1.1699114>.
- Parag, K. V. (2021). Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLOS Computational Biology*, 17(9):1–23.
- Piegorsch, W. W. (1990). Maximum likelihood estimation for the Negative Binomial dispersion parameter. *Biometrics*, 46(3):863–867. <https://doi.org/10.2307/2532104>.
- Pircalabelu, E. (2021). A spline-based time-varying reproduction number for modelling epidemiological outbreaks. *LIDAM Discussion Paper ISBA*. <http://hdl.handle.net/2078.1/244926>.
- Roberts, G. O. and Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268. <https://doi.org/10.1111/1467-9868.00123>.

- Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367. <https://www.jstor.org/stable/3182776>.
- Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363. <https://www.jstor.org/stable/3318418>.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using Integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86. <https://doi.org/10.1080/01621459.1986.10478240>.
- Wallinga, J. and Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604.
- White, L. F., Moser, C. B., Thompson, R. N., and Pagano, M. (2021). Statistical estimation of the reproductive number from case notification data. *American Journal of Epidemiology*, 190(4):611–620. <https://doi.org/10.1093/aje/kwaa211>.