Bayesian nowcasting with Laplacian-P-splines

Bryan Sumalinab*[1,2], Oswaldo Gressani[1], Niel Hens[1,3], Christel Faes[1]

[1]Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Data Science Institute (DSI), Hasselt University, Hasselt, Belgium

[2]Department of Mathematics and Statistics, College of Science and Mathematics, Mindanao State University - Iligan Institute of Technology, Iligan City, Philippines

[3]Centre for Health Economic Research and Modelling Infectious Diseases (CHERMID), Vaccine & Infectious Disease Institute, Antwerp University, Antwerp, Belgium

*Corresponding author. *E-mail address*: bryan.sumalinab@uhasselt.be

Abstract

During an epidemic, the daily number of reported infected cases, deaths or hospitalizations is often lower than the actual number due to reporting delays. Nowcasting aims to estimate the cases that have not yet been reported and combine it with the already reported cases to obtain an estimate of the daily cases. In this paper, we present a fast and flexible Bayesian approach to do nowcasting by combining P-splines and Laplace approximations. The main benefit of Laplacian-P-splines (LPS) is the flexibility and faster computation time compared to Markov chain Monte Carlo (MCMC) algorithms that are often used for Bayesian inference. In addition, it is natural to quantify the prediction uncertainty with LPS in the Bayesian framework, and hence prediction intervals are easily obtained. Model performance is assessed through simulations and the method is applied to COVID-19 mortality and incidence cases in Belgium.

Keywords: Bivariate P-splines, COVID-19, Epidemic, Reporting delay.

1 Introduction

Nowcasting is a term used for estimating the occurred-but-not-yet-reported-events (Donker et al., 2011; Van de Kassteele et al., 2019). In epidemiology, real-time updates of new symptomatic/infected individuals are helpful to assess the present situation and provide recommendations for rapid planning and for implementing essential measures to contain an epidemic outbreak. The exact number of new daily cases is frequently subject to reporting delays, resulting in underreporting of the real number of infected individuals for that day. Failing to account for the reporting delays will lead to possibly biased predictions that might have an effect on policy making (Gutierrez et al., 2020). The main goal of nowcasting is to estimate the actual number of new cases by combining the (predicted) not-yet-reported-cases with the already reported cases.

Several early references that establish the statistical framework for this type of problem can be found in the paper of Lawless (1994); De Angelis and Gilks (1994); and Lindsey (1996). In disease surveillance, Höhle and an der Heiden (2014) applied nowcasting to the outbreak of Shiga toxin-producing Escherichia coli in Germany and also to the SARS-CoV-2 outbreak (Glöckner et al., 2020; Günther et al., 2021). Their approach is formulated within a hierarchical Bayesian framework that consists of estimating the epidemic curve by using a quadratic spline based on a truncated power basis function and the time-varying reporting delay distribution which is approximated by a discrete time survival model. Van de Kassteele et al. (2019) pointed out that a potential drawback of such a method is the long computation time required by the Markov chain Monte Carlo (MCMC) algorithm and therefore proposed an alternative fast and flexible modelling strategy based on bivariate P-splines (Penalized B-splines). P-splines (Eilers and

Marx, 1996) provide a flexible smoothing tool used to describe trends in the data. It introduces a penalty parameter that controls the roughness of the fit and counterbalances the flexibility of a rich B-splines basis (Eilers and Marx, 1996; Eilers et al., 2015). Other attractive features of the P-splines smoother are the relatively simple structure of the penalty matrix that is effortlessly computed and the natural extension to a Bayesian framework (Lang and Brezger, 2004). Based on the approach of Van de Kassteele et al. (2019), the number of cases are structured in a two-dimensional table (with calendar time as the first dimension and delay time as the second dimension), yielding the data matrix used as an input in the model. The reporting intensity is assumed to be a smooth surface and is modelled using two-dimensional P-splines.

In this paper, we build upon the work of Van de Kassteele et al. (2019) by proposing a new nowcasting methodology based on Laplacian-P-splines (LPS) in a fully Bayesian framework. A key advantage of working with the Bayesian approach is the ease to obtain the predictive distribution and quantify the uncertainty associated with the predictions. In addition, the posterior distribution of the penalty parameter can be explored, and hence its uncertainty can be accounted for. The Laplace approximation uses a second-order Taylor expansion to approximate the posterior distribution of the regression parameters by a Gaussian density. It is a sampling-free method with the major advantage of faster computational time as opposed to MCMC approaches that are commonly used in Bayesian inference. Therefore, given the flexibility of previously mentioned P-splines smoothers and the computational benefit of Laplace approximations, it can be a helpful tool in the daily monitoring of new cases during an epidemic period. Laplacian-P-splines already proved to be useful in survival models (Gressani and Lambert, 2018; Gressani et al., 2022a), generalized additive models (Gressani and Lambert, 2021) and also in epidemic models for estimating the effective reproduction number (Gressani et al., 2022b). We build on the work of Gressani and Lambert (2021) to extend the Laplacian-P-splines methodol-

ogy to nowcasting, thereby providing a fast and flexible (fully) Bayesian alternative to Van de Kassteele et al. (2019). To evaluate the (predictive) performance of our method, a simulation study is implemented and several performance measures are reported such as the (relative) bias, prediction interval coverage, and prediction interval width. Finally, we apply our method to the COVID-19 mortality and incidence data in Belgium for the year 2021 and 2022, respectively. The R codes used for the simulation study are available on GitHub through the link https://github.com/bryansumalinab/Laplacian-P-spline-nowcasting.git.

2 Methodology

2.1 Bayesian model formulation

In this section, we follow the work of Van de Kassteele et al. (2019) to formulate a fully Bayesian model based on P-splines. Let $y_{t,d}$ denote the number of cases that occurred at time $t = 1, 2, ..., T$ (corresponding to the calendar day) and are reported with a delay of $d = 0, 1, 2..., D$ days. The information on cases can be summarized in matrix form:

$$Y = \begin{bmatrix} y_{1,0} & y_{1,1} & y_{1,2} & \cdots & y_{1,D} \\ y_{2,0} & y_{2,1} & y_{2,2} & \cdots & y_{2,D} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ y_{T-(D-1),0} & y_{T-(D-1),1} & y_{T-(D-1),2} & \cdots & \boldsymbol{y_{T-(D-1),D}} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ y_{T-1,0} & y_{T-1,1} & \boldsymbol{y_{T-1,2}} & \cdots & \boldsymbol{y_{T-1,D}} \\ y_{T,0} & \boldsymbol{y_{T,1}} & \boldsymbol{y_{T,2}} & \cdots & \boldsymbol{y_{T,D}} \end{bmatrix},$$

with cases that have not yet been reported (at time T) highlighted in bold. The not-yet-reported cases correspond to (t, d) combinations satisfying $t > T - d$. The main objective is to predict the total number of cases, $y_t = \sum_{d=0}^{D} y_{t,d}$, for $t = T - (D - 1), \ldots, T$ for

which the nowcasted and already reported cases can be combined.

Let $\mathcal{D} := \boldsymbol{y} = (y_1, y_2 \ldots, y_n)^\top$ denote the vector of the observed number of cases by stacking the columns of matrix Y for the reported cases, where each entry corresponds to each (t, d) combination of reported cases $y_{t,d}$. The model assumes that the number of cases either follows a Poisson or a negative binomial (NB) distribution, i.e., $y_{t,d} \sim \text{Poisson}(\mu_{t,d})$ or $y_{t,d} \sim \text{NB}(\mu_{t,d}, \phi)$ with mean $\mu_{t,d} > 0$. For the negative binomial, $\phi > 0$ is an overdispersion parameter and the variance is $\mathbb{V}(y_{t,d}) = \mu_{t,d} + \frac{\mu_{t,d}^2}{\phi}$. Following Van de Kassteele et al. (2019), the (log) mean number of cases is modeled with two dimensional B-splines:

$$\log(\mu_{t,d}) = \beta_0 + \sum_{j=1}^{K_T} \sum_{k=1}^{K_D} \theta_{j,k} b_j(t) b_k(d) + \sum_{l=1}^{p} \beta_l z_l(t, d), \tag{1}$$

where β_0 is the intercept; $b_j(\cdot)$ and $b_k(\cdot)$ are univariate B-splines basis functions specified in the time and delay dimensions, respectively; and $z_l(t, d)$ represents additional covariates with regression coefficients β_l. In matrix notation:

$$\log(\boldsymbol{\mu}) = B\boldsymbol{\theta} + Z\boldsymbol{\beta}, \tag{2}$$

where the matrices B and Z correspond to the basis functions and covariates, respectively, and vectors $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are the associated parameters to be estimated (Details in supplementary material S1).

In the philosophy of P-splines (Eilers and Marx, 1996), we use a rich (cubic) B-splines basis and counterbalance the associated flexibility by imposing a discrete roughness penalty on contiguous B-spline coefficients. For the two dimensional P-splines, the penalty can be obtained based on row-wise (direction of calendar time) and column-wise (direction of reporting delay) differences for matrix $\Theta = (\theta_{j,k})$ with $j = 1, \ldots, K_T$ and $k = 1, \ldots, K_D$ (supplementary material S1) (see Durbán et al. (2002) and Fahrmeir et al. (2013) (pp.

507-508)). Let D_t^m and D_d^m denote the mth order row-wise and column-wise difference matrix with dimensions $(K_T - m) \times K_T$ and $(K_D - m) \times K_D$, respectively. In this paper, we use a second order ($m = 2$) difference penalty. For ease of notation, let $D_t = D_t^m$ and $D_d = D_d^m$. The difference matrix for vector $\boldsymbol{\theta}$ can be obtained by expanding the difference matrix into $I_{K_D} \otimes D_t$ and $D_d \otimes I_{K_T}$, where $\otimes$ denotes the Kronecker product. Using this notation, the row-wise and column-wise difference penalty can be written, respectively, as:

$$\|(I_{K_D} \otimes D_t)\boldsymbol{\theta}\|^2 = \boldsymbol{\theta}^\top (I_{K_D} \otimes D_t)^\top (I_{K_D} \otimes D_t)\boldsymbol{\theta} = \boldsymbol{\theta}^\top (I_{K_D} \otimes D_t^\top D_t)\boldsymbol{\theta} \quad \text{and}$$
$$\|(D_d \otimes I_{K_T})\boldsymbol{\theta}\|^2 = \boldsymbol{\theta}^\top (D_d \otimes I_{K_T})^\top (D_d \otimes I_{K_T})\boldsymbol{\theta} = \boldsymbol{\theta}^\top (D_d^\top D_d \otimes I_{K_T})\boldsymbol{\theta}.$$

Let $\lambda_t > 0$ and $\lambda_d > 0$ denote the row-wise and column-wise penalty parameter, respectively. The penalty for the two dimensional B-splines is then given by:

$$\begin{aligned}
&\lambda_t \boldsymbol{\theta}^\top (I_{K_D} \otimes D_t^\top D_t)\boldsymbol{\theta} + \lambda_d \boldsymbol{\theta}^\top (D_d^\top D_d \otimes I_{K_T})\boldsymbol{\theta} \\
&= \boldsymbol{\theta}^\top \Big(\lambda_t (I_{K_D} \otimes D_t^\top D_t) + \lambda_d (D_d^\top D_d \otimes I_{K_T}) \Big) \boldsymbol{\theta}.
\end{aligned}$$

Let us define the penalty matrices $P_t = D_t^\top D_t + \delta I_{K_T}$ and $P_d = D_d^\top D_d + \delta I_{K_D}$, where δ is a small number (say $\delta = 10^{-12}$), to ensure that the penalty matrices are full rank and thus invertible. Following Lang and Brezger (2004), the penalty can be translated in the Bayesian framework by specifying a Gaussian prior $(\boldsymbol{\theta}|\boldsymbol{\lambda}) \sim \mathcal{N}_{dim(\boldsymbol{\theta})}(\mathbf{0}, \mathcal{P}^{-1}(\boldsymbol{\lambda}))$, where $\boldsymbol{\lambda} = (\lambda_t, \lambda_d)^\top$ is the penalty vector and $\mathcal{P}(\boldsymbol{\lambda}) = \lambda_t (I_{K_D} \otimes P_t) + \lambda_d (P_d \otimes I_{K_T})$ is the penalty matrix. A Gaussian prior is also assumed for $\boldsymbol{\beta}$, namely $\boldsymbol{\beta} \sim \mathcal{N}_{dim(\boldsymbol{\beta})}(\mathbf{0}, V_{\boldsymbol{\beta}}^{-1})$ with $V_{\boldsymbol{\beta}} = \zeta I_{p+1}$ (small ζ, e.g. $\zeta = 10^{-5}$). Denote by $X = (B, Z)$ the global design matrix, $\boldsymbol{\xi} = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ the latent parameter vector and $Q_{\boldsymbol{\xi}}^{\boldsymbol{\lambda}} = \text{blkdiag}(V_{\boldsymbol{\beta}}, \mathcal{P}(\boldsymbol{\lambda}))$ the precision matrix for $\boldsymbol{\xi}$ where blkdiag$(\cdot)$ refers to a block diagonal matrix. From here, we focus on the negative binomial model for the number of cases. The Poisson model is described

in detail in supplementary material S4. The full Bayesian (negative binomial) model is summarized as follows:

$$
\begin{aligned}
&(y_i|\boldsymbol{\xi}) \sim \mathrm{NB}(\mu_i, \phi) \text{ with } \log(\boldsymbol{\mu}) = X\boldsymbol{\xi}, \\
&(\boldsymbol{\xi}|\boldsymbol{\lambda}) \sim \mathcal{N}_{dim(\boldsymbol{\xi})}(\mathbf{0}, (Q_{\boldsymbol{\xi}}^{\boldsymbol{\lambda}})^{-1}), \\
&(\lambda_t|\delta_t) \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu\delta_t}{2}\right), \\
&(\lambda_d|\delta_d) \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{\nu\delta_d}{2}\right), \\
&\delta_t \sim \mathcal{G}(a_\delta, b_\delta), \\
&\delta_d \sim \mathcal{G}(a_\delta, b_\delta), \\
&\phi \sim \mathcal{G}(a_\phi, b_\phi),
\end{aligned}
$$

where $\mathcal{G}(a, b)$ denotes a Gamma distribution with mean a/b and variance a/b^2. This robust prior specification on the penalty parameters follows from Jullion and Lambert (2007). They have shown that when the hyperparameters a_δ, b_δ are chosen to be equal and small enough (say 10^{-5}), then the resulting fit is robust to the value of ν (e.g. $\nu = 3$ in this paper). In addition, we also set $a_\phi = b_\phi = 10^{-5}$.

2.2 Laplace approximation to the conditional posterior of $\boldsymbol{\xi}$

The conditional posterior of the latent vector $\boldsymbol{\xi}$ is approximated by a Gaussian distribution via the Laplace approximation. The gradient and Hessian of the (log) conditional posterior are analytically derived and used in a Newton-Raphson algorithm to obtain the Gaussian approximation to the conditional posterior distribution of $\boldsymbol{\xi}$.

Note that for a negative binomial distributed y_i with mean $E(y_i) = \mu_i$, the probability distribution can be written as an exponential dispersion family given by $p(y_i; \gamma, \phi) = \exp\{[y_i\gamma_i - b(\gamma_i)]/a(\phi) + c(y_i, \phi)\}$, where $\gamma_i = \phi \log\left(\frac{\mu_i}{\mu_i+\phi}\right)$, $b(\eta_i) = -\phi^2 \log\left(\frac{\phi}{\mu_i+\phi}\right)$,

$c(y_i, \phi) = \log\left(\frac{\Gamma(y_i+\phi)}{\Gamma(y_i+1)\Gamma(\phi)}\right)$, $a(\phi) = \phi$ and $\Gamma(\cdot)$ is the gamma function. Thus, for fixed ϕ, a negative binomial regression model is a generalized linear model (GLM) where μ_i is linked on the linear predictor through the link function $g(\mu_i)$ such that $g(\mu_i) = \sum_{i=1}^{q} \xi_j x_{ij}$ (see Agresti (2013) for a detailed account about GLMs). Here, we use the log-link function $g(\mu_i) = \log(\mu_i)$. The log-likelihood is given by $\log \mathcal{L}(\boldsymbol{\xi}, \phi; \mathcal{D}) = \sum_{i=1}^{n} \{[y_i\gamma_i - b(\gamma_i)]/a(\phi) + c(y_i, \phi)\}$ and it can be shown that for a negative binomial model, the gradient and Hessian are:

$$\nabla_{\boldsymbol{\xi}} \log \mathcal{L}(\boldsymbol{\xi}, \phi; \mathcal{D}) = X^\top W D(\boldsymbol{y} - \boldsymbol{\mu}) \quad \text{and} \quad \nabla^2_{\boldsymbol{\xi}} \log \mathcal{L}(\boldsymbol{\xi}, \phi; \mathcal{D}) = X^\top M V X - X^\top W X,$$

where $W = \text{diag}(w_1, \ldots, w_n)$, $w_i = [\mathbb{V}(y_i)(g'(\mu_i))^2]^{-1}$, $D = \text{diag}(g'(\mu_1), \ldots, g'(\mu_n))$, $V = \text{diag}(v_1, v_2, \ldots, v_n)$, $v_i = \frac{1}{g'(\mu_i)}\left(\frac{1}{\mathbb{V}(y_i)}\frac{\partial(g'(\mu_i)^{-1})}{\partial \mu_i} + \frac{1}{g'(\mu_i)}\frac{\partial(\mathbb{V}(y_i)^{-1})}{\partial \mu_i}\right)$ and $M = \text{diag}(y_1 - \mu_1, \ldots, y_n - \mu_n)$ (see details in supplementary material S2).

Using Bayes' rule, the posterior of $\boldsymbol{\xi}$ conditional on the penalty vector $\boldsymbol{\lambda}$ and overdispersion parameter ϕ is:

$$\begin{aligned} p(\boldsymbol{\xi}|\boldsymbol{\lambda}, \phi, \mathcal{D}) &\propto \mathcal{L}(\boldsymbol{\xi}, \phi; \mathcal{D}) p(\boldsymbol{\xi}|\boldsymbol{\lambda}) \\ &\propto \exp\left\{\frac{1}{\phi}\sum_{i=1}^{n}[y_i\gamma_i - b(\gamma_i)] - \frac{1}{2}(\boldsymbol{\xi}^\top Q^{\boldsymbol{\lambda}}_{\boldsymbol{\xi}} \boldsymbol{\xi})\right\}. \end{aligned}$$

The gradient and Hessian for the log-conditional posterior of $\boldsymbol{\xi}$ are:

$$\begin{aligned} \nabla_{\boldsymbol{\xi}} \log p(\boldsymbol{\xi}|\boldsymbol{\lambda}, \phi, \mathcal{D}) &= X^\top W D(\boldsymbol{y} - \boldsymbol{\mu}) - Q^{\boldsymbol{\lambda}}_{\boldsymbol{\xi}} \boldsymbol{\xi}, \\ \nabla^2_{\boldsymbol{\xi}} \log p(\boldsymbol{\xi}|\boldsymbol{\lambda}, \phi, \mathcal{D}) &= X^\top M V X - X^\top W X - Q^{\boldsymbol{\lambda}}_{\boldsymbol{\xi}}. \end{aligned}$$

The above gradient and Hessian can then be used in a Newton-Raphson algorithm to obtain the mode of the conditional posterior of $\boldsymbol{\xi}$. After convergence, the Laplace approximation of the conditional posterior of $\boldsymbol{\xi}$ is a multivariate Gaussian density denoted

by $\widetilde{p}_G(\boldsymbol{\xi}|\boldsymbol{\lambda},\phi,\mathcal{D}) = \mathcal{N}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}}, \widehat{\Sigma}_{\boldsymbol{\lambda}})$ where $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}}$ is the mean/mode and $\widehat{\Sigma}_{\boldsymbol{\lambda}}$ is the covariance matrix. There are instances where this approximation may be insufficient, for example, when dealing with sparse likelihoods or low counts. In such scenarios, the resulting posterior distribution could exhibit skewness. Lambert and Gressani (2023) proposed an approach to address this asymmetry issue but due to complexity, this is beyond the scope of this paper.

2.3 Optimization of hyperparameters and overdispersion parameter

In this section, we derive the (approximate) posterior distribution of the hyperparameters $\boldsymbol{\lambda} = (\lambda_t, \lambda_d)^\top$ and $\boldsymbol{\delta} = (\delta_t, \delta_d)^\top$, and overdispersion parameter ϕ. Let $\boldsymbol{\eta} = (\lambda_t, \lambda_d, \delta_t, \delta_d)^\top$ denote the vector of hyperparameters. Using Bayes' theorem, the joint marginal posterior of $\boldsymbol{\eta}$ and ϕ is $p(\boldsymbol{\eta},\phi|\mathcal{D}) \propto \frac{\mathcal{L}(\boldsymbol{\xi},\phi;\mathcal{D})p(\boldsymbol{\xi}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\boldsymbol{\delta})p(\boldsymbol{\delta})p(\phi)}{p(\boldsymbol{\xi}|\boldsymbol{\lambda},\boldsymbol{\phi},\mathcal{D})}$. Following Rue et al. (2009), the above posterior can be approximated by replacing $p(\boldsymbol{\xi}|\boldsymbol{\lambda},\phi,\mathcal{D})$ with $\widetilde{p}_G(\boldsymbol{\xi}|\boldsymbol{\lambda},\phi,\mathcal{D})$ obtained in Section 2.2 and by evaluating the latent vector at $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}}$. Note that $\gamma_i = \mathbf{x}_i^\top \boldsymbol{\xi}$, where $\mathbf{x}_i^\top$ corresponds to the ith row of the design matrix X. Also, the determinant $|Q_{\boldsymbol{\xi}}^{\boldsymbol{\lambda}}|$ in $p(\boldsymbol{\xi}|\boldsymbol{\lambda})$ is $|Q_{\boldsymbol{\xi}}^{\boldsymbol{\lambda}}| \propto |\mathcal{P}(\boldsymbol{\lambda})|$. Hence, the approximated joint marginal posterior of $\boldsymbol{\eta}$ and ϕ is:

$$
\begin{aligned}
\widetilde{p}(\boldsymbol{\eta},\phi|\mathcal{D}) &\propto \frac{\mathcal{L}(\boldsymbol{\xi},\phi;\mathcal{D})p(\boldsymbol{\xi}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\boldsymbol{\delta})p(\boldsymbol{\delta})p(\phi)}{p(\boldsymbol{\xi}|\boldsymbol{\lambda},\boldsymbol{\phi},\mathcal{D})}\Big|_{\boldsymbol{\xi}=\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}}} \\
&\propto \exp\left(\sum_{i=1}^{n}[y_i\mathbf{x}_i^\top\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}} - b(\mathbf{x}_i^\top\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}})]/\phi + \log(\Gamma(y_i+\phi)) - \Gamma(\phi)\right) \\
&\quad \times |\mathcal{P}(\boldsymbol{\lambda})|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}}^\top Q_{\boldsymbol{\xi}}^{\boldsymbol{\lambda}}\widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}}\right) \\
&\quad \times (\lambda_t\lambda_d)^{(\frac{\nu}{2}-1)}\,\delta_t^{(\frac{\nu}{2}+a_\delta-1)}\exp(-\delta_t(b_\delta + \frac{\nu}{2}\lambda_t)) \\
&\quad \times \delta_d^{(\frac{\nu}{2}+a_\delta-1)}\exp(-\delta_d(b_\delta + \frac{\nu}{2}\lambda_d)) \\
&\quad \times \phi^{(a_\phi-1)}\exp(-b_\phi\phi)|\widehat{\Sigma}_{\boldsymbol{\lambda}}|^{\frac{1}{2}}.
\end{aligned}
$$

Integrating out the hyperparameters δ_t and δ_d from $\widetilde{p}(\boldsymbol{\eta}, \phi|\mathcal{D})$, we obtain the joint marginal posterior of $\boldsymbol{\lambda}$ and ϕ as:

$$\begin{aligned}
\widetilde{p}(\boldsymbol{\lambda}, \phi|\mathcal{D}) &= \int_0^\infty \int_0^\infty \widetilde{p}(\boldsymbol{\eta}, \phi|\mathcal{D}) d\delta_t \, d\delta_d \\
&\propto \exp\left(\sum_{i=1}^n [y_i \mathbf{x}_i^\top \widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}} - b(\mathbf{x}_i^\top \widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}})]/\phi + \log(\Gamma(y_i + \phi)) - \Gamma(\phi) \right) \\
&\quad \times |\mathcal{P}(\boldsymbol{\lambda})|^{\frac{1}{2}} \exp\left(-\frac{1}{2} \widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}}^\top Q_{\boldsymbol{\xi}}^{\boldsymbol{\lambda}} \widehat{\boldsymbol{\xi}}_{\boldsymbol{\lambda}} \right) \phi^{(a_\phi - 1)} \exp(-b_\phi \phi) |\widehat{\Sigma}_{\boldsymbol{\lambda}}|^{\frac{1}{2}} \\
&\quad \times (\lambda_t \lambda_d)^{(\frac{\nu}{2} - 1)} \left((b_\delta + \frac{\nu}{2} \lambda_t)(b_\delta + \frac{\nu}{2} \lambda_d) \right)^{-(\frac{v}{2} + a_\delta)} .
\end{aligned}$$

To ensure numerical stability, we log-transform the penalty vector $\boldsymbol{v} = (v_t, v_d)^\top = (\log(\lambda_t), \log(\lambda_d))^\top$ and overdispersion parameter $v_\phi = \log(\phi)$. The joint log-posterior of $\boldsymbol{v}$ and v_ϕ is then given by:

$$\begin{aligned}
\log \widetilde{p}(\boldsymbol{v}, v_\phi; \mathcal{D}) \doteq & \sum_{i=1}^n [y_i \mathbf{x}_i^\top \widehat{\boldsymbol{\xi}}_{\boldsymbol{v}} - b(\mathbf{x}_i^\top \widehat{\boldsymbol{\xi}}_{\boldsymbol{v}})]/\exp(v_\phi) + \log(\Gamma(y_i + \exp(v_\phi))) - \Gamma(\exp(v_\phi)) \\
& + \frac{1}{2} \log|P(\boldsymbol{v})| - \frac{1}{2} \widehat{\boldsymbol{\xi}}_{\boldsymbol{v}}^\top Q_{\boldsymbol{\xi}}^{\boldsymbol{v}} \widehat{\boldsymbol{\xi}}_{\boldsymbol{v}} + a_\phi v_\phi - b_\phi e^{v_\phi} + \frac{1}{2} \log|\widehat{\Sigma}_{\boldsymbol{v}}| \\
& + \frac{\nu}{2} (v_d + v_t) - \left(\frac{\nu}{2} + a_\delta \right) \left(\log(b_\delta + \frac{\nu}{2} \exp(v_d)) + \log(b_\delta + \frac{\nu}{2} \exp(v_t)) \right) ,
\end{aligned}$$

where $\doteq$ denotes equality up to an additive constant. Finally, we maximize $\log \widetilde{p}(\boldsymbol{v}, v_\phi; \mathcal{D})$ using the *optim()* function in R to obtain the posterior mode as a point estimate for $\boldsymbol{v}$ and v_ϕ.

2.4 Nowcasting with prediction interval

To obtain the mean nowcast estimate with the prediction interval, note that $\log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\xi}$ or equivalently $\log(\mu_{t,d}) = \boldsymbol{x}_{t,d}^\top \boldsymbol{\xi}$, where $\widetilde{p}_G(\boldsymbol{\xi}|\boldsymbol{v}, \mathcal{D}) = \mathcal{N}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{v}}, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{v}})$. Thus, $\widetilde{p}(\log(\mu_{t,d})|\boldsymbol{v}, \mathcal{D}) = \mathcal{N}(\boldsymbol{x}_{t,d}^\top \widehat{\boldsymbol{\xi}}_{\boldsymbol{v}}, \boldsymbol{x}_{t,d}^\top \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{v}} \boldsymbol{x}_{t,d})$. The mean estimate for the not-yet-reported cases is calculated as $\widehat{\mu}_{t,d} = \exp(\boldsymbol{x}_{t,d}^\top \widehat{\boldsymbol{\xi}}_{\boldsymbol{v}})$ for all (t, d) combinations with $t > T - d$. Then, the estimate for the total number of cases for each t is obtained by summing

the already reported cases and the mean estimate for the not-yet-reported cases, i.e. $\widehat{\mu}_t = \sum_{\{d:t \leq T-d\}} y_{t,d} + \sum_{\{d:t > T-d\}} \widehat{\mu}_{t,d}$.

The prediction interval for the nowcast values is obtained by sampling from the posterior predictive distribution of the log-mean number of cases by following a five-step procedure:

1. For each (t, d) combinations with $t > T - d$ (corresponding to the not-yet-reported cases), generate 1000 random samples $(\widehat{\boldsymbol{y}}_{t,d})$ from a Gaussian distribution with mean $\boldsymbol{x}_{t,d}^\top \widehat{\boldsymbol{\xi}}_{\boldsymbol{v}}$ and variance $\boldsymbol{x}_{t,d}^\top \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{v}} \boldsymbol{x}_{t,d}$.

2. Exponentiate the sampled values from the previous step to obtain the average reporting intensities $\widehat{\boldsymbol{\mu}}_{t,d} = \exp(\widehat{\boldsymbol{y}}_{t,d})$.

3. Compute the average prediction for the not-yet-reported cases for each t. That is, compute $\widehat{\boldsymbol{\mu}}_t = \sum_d \widehat{\boldsymbol{\mu}}_{t,d}$ for $t > T - d$.

4. For each t, sample $\widehat{\boldsymbol{y}}_t$ containing 1000 values from $\mathrm{NB}(\widehat{\boldsymbol{\mu}}_t, \exp(\widehat{v}_\phi))$.

5. Finally, compute the quantiles of the sampled values $\widehat{\boldsymbol{y}}_t$ that correspond to the desired prediction interval.

2.5 Delay distribution

To obtain the smooth estimate of the delay distribution, only the first term on the right hand side of equation (2) is used (excluding the day of the week effects), as explained in the paper of Van de Kassteele et al. (2019). Specifically, the procedure to compute the delay distribution is as follows:

1. Compute the contribution of the smoothing term in equation (2) to the reporting intensity for all (t, d) combinations: $\boldsymbol{\mu}^{\text{smooth}} = \exp(B\boldsymbol{\theta})$.

2. Arrange $\boldsymbol{\mu}^{\text{smooth}}$ into a $T \times (D + 1)$ matrix with entries $\mu_{t,d}^{\text{smooth}}$.

3. For each $t = 1, \ldots, T$, compute the reporting delay distribution given by: $\hat{f}_t(d) = \frac{\mu_{t,d}^{\text{smooth}}}{\sum_{d=0}^{D} \mu_{t,d}^{\text{smooth}}}$.

3 Simulations

A simulation study is implemented in order to evaluate the predictive performance of the proposed method. The procedure to perform the simulations is as follows:

1. Consider a function $f(t)$ that represents the mean epidemic curve of all cases such that $\mu(t) = \exp(f(t))$ for $t = 1, ..., T$.

2. For each t, generate a random sample y_t from a negative binomial distribution with mean $\mu(t)$ and fixed overdispersion parameter (we choose the value of 10).

3. To account for possible delays $d = 0, 1, 2, \ldots, D$, generate samples from a multinomial distribution with probabilities $(p_0, p_1, p_2, \cdots, p_D$ such that $\sum_{d=0}^{D} p_d = 1)$, i.e.,

$$(y_{t,0}, y_{t,1}, y_{t,2}, \ldots, y_{t,D}) \sim Multinomial(y_t, p_0, p_1, p_2, \cdots, p_D).$$

   This sample represents the reported number of cases for each (t, d) combination.

4. Repeat steps 1 to 3 for 500 times to generate 500 possible realizations.

We consider two functions inspired from the paper of Noufaily et al. (2016) given by:

$$f(t) = \theta_1 + \theta_2 \sin\left(\frac{2\pi t}{150}\right),$$
$$h(t) = \theta_1 + \theta_2 \sin\left(\frac{2\pi t}{150}\right) + \theta_3 \sqrt{t},$$

for $t = 1, \ldots, 365$. That is, we assume a one year (365 days) time window in the simulation. In terms of delay probabilities, we consider a maximum delay of $D = 7$ days with probabilities $(p_0, p_1, p_2, \cdots, p_7) = (0.0, 0.1, 0.4, 0.2, 0.1, 0.1, 0.05, 0.05)$, as illustrated

in Figure 1. For example, 5% of the cases that occurred at time $T-6$ have a delay of 7 days (with probability $p_7 = 0.05$) and still need to be reported, or equivalently, 95% of the cases that occurred at time $T-6$ have already been reported. For cases that occurred at time $T-5$, only 90% of these cases have been reported, as cases with delays of 6 and 7 days are yet to be reported. For those cases that occurred on the nowcast day (time T), none of these cases is reported, that is, the delay is 100%.

Two scenarios are considered for $f(t)$: (i) the first has a small number of cases with values of $\theta_1 = 3$ and $\theta_2 = 1$; (ii) the second scenario has a relatively large number of cases with $\theta_1 = 3$ and $\theta_2 = 2$ and an additional factor of 50 is added to the mean function such that $\mu(t) = 50 + \exp(f(t))$. These scenarios are denoted by $f_1(t)$ and $f_2(t)$, respectively. Similarly, two scenarios are also considered for the second function $h(t)$: (i) the first scenario has values $\theta_1 = 0$, $\theta_2 = 0.4$ and $\theta_3 = 0.2$; (ii) the second scenario has values $\theta_1 = 1.5$, $\theta_2 = 0.4$ and $\theta_3 = 0.2$. We denote these functions by $h_1(t)$ and $h_2(t)$, respectively. The first function is symmetric with three peaks as shown in Figures 2a and 2b. On the other hand, the second function is not periodic as opposed to the first function (see Figures 2c and 2d). The plots for (one realization of) simulated cases based on these functions are shown in Figure 3 with (dashed) vertical lines corresponding to the different nowcast dates.

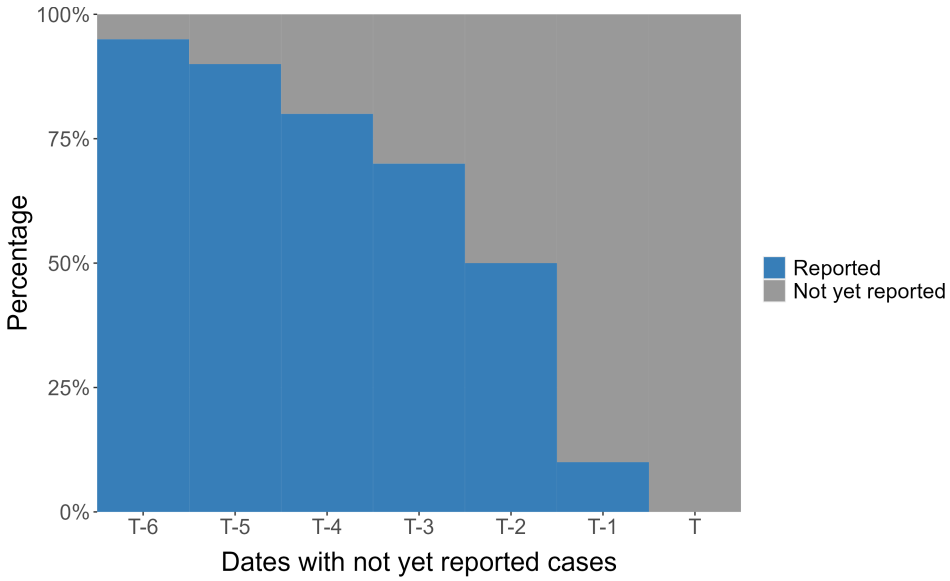
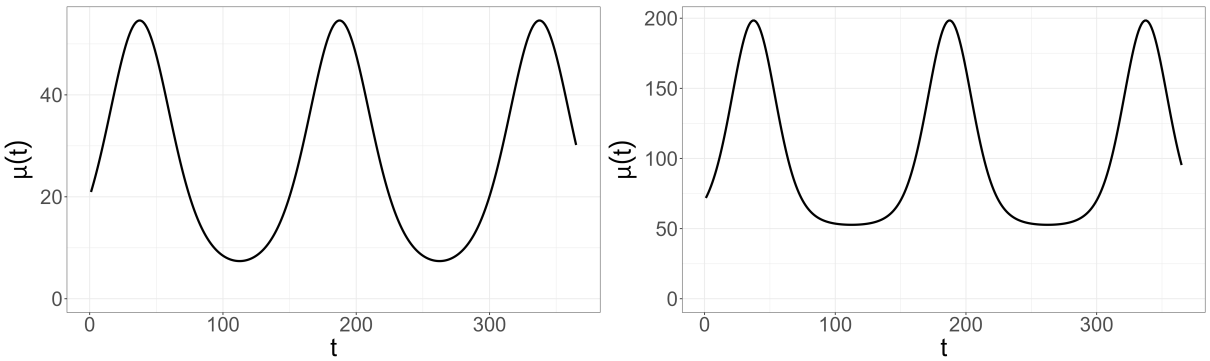
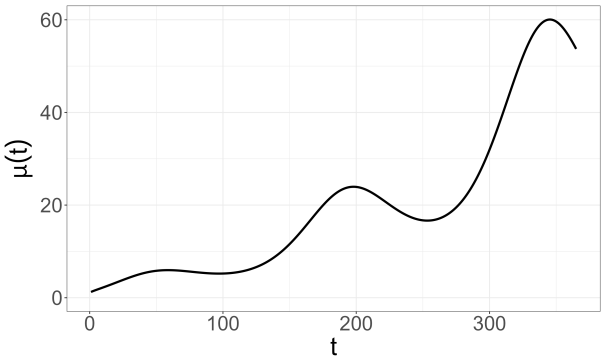


Figure 1: Illustration of the delay probabilities considered in the simulation.

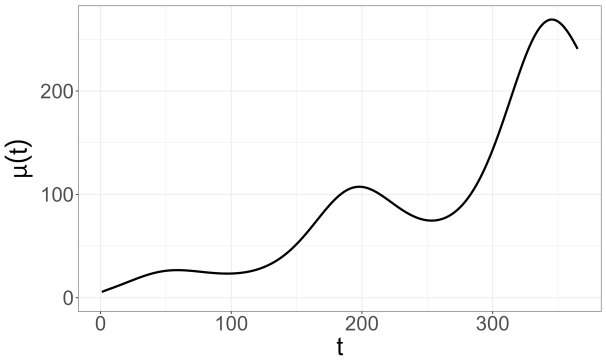


(a) $\mu(t) = \exp(f_1(t))$

(b) $\mu(t) = \exp(f_2(t))$

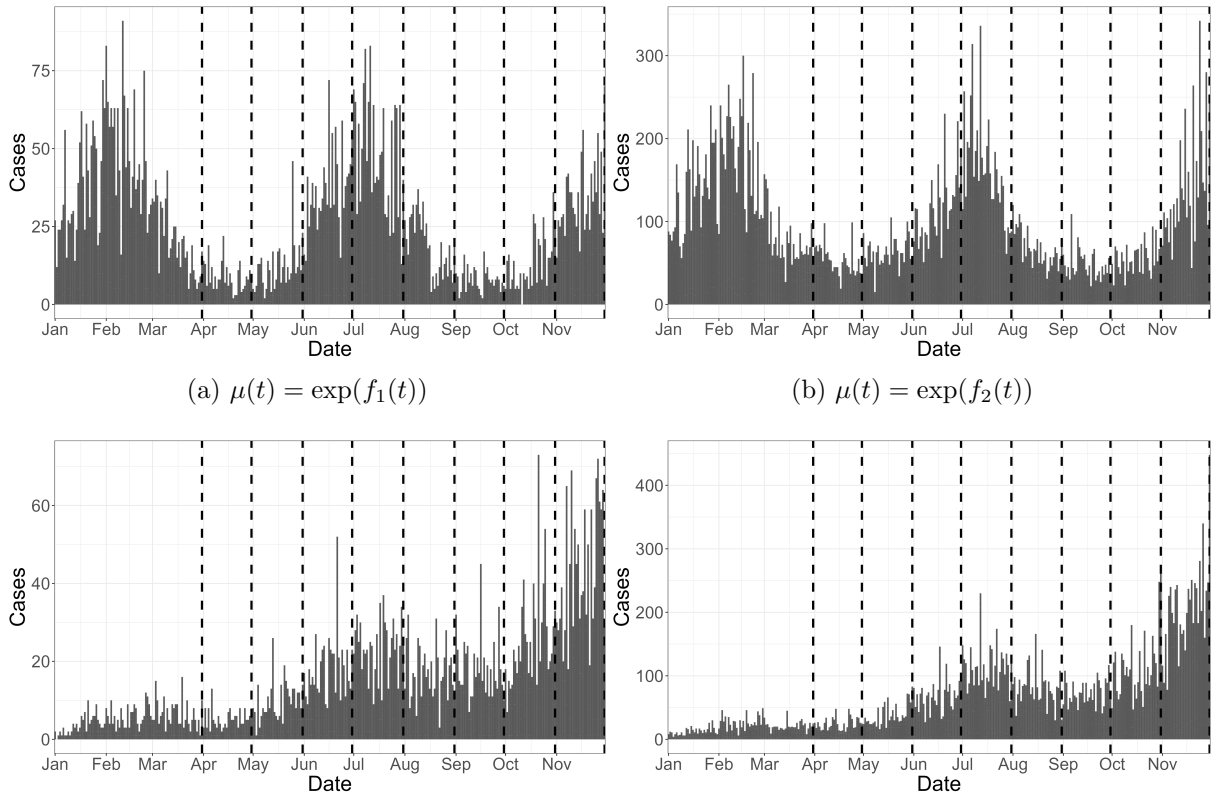


(c) $\mu(t) = \exp(h_1(t))$



(d) $\mu(t) = \exp(h_2(t))$

Figure 2: Mean epidemic curves considered in the simulations.



(a) $\mu(t) = \exp(f_1(t))$ (b) $\mu(t) = \exp(f_2(t))$

(c) $\mu(t) = \exp(h_1(t))$ (d) $\mu(t) = \exp(h_2(t))$

Figure 3: Simulated cases for different epidemic curves. The dashed vertical lines correspond to different nowcast dates.

For each generated data set, we fit the proposed models (LPS-NB and LPS-Poisson) as well as the method proposed by Van de Kassteele et al. (2019) (VDK) on the data, and compute the desired accuracy measures. The bias and relative bias (% bias) for the true mean epidemic curve ($\boldsymbol{\mu}$), 95% prediction interval (PI) coverage and prediction interval (PI) width are chosen to measure the predictive accuracy of our methodology. For a given calendar day t, the bias and relative bias are computed as $Bias_t = \frac{1}{S}\sum_{s=1}^{S}(\hat{\mu}_{st} - \mu_t)$ and $\%Bias_t = \frac{1}{S}\sum_{s=1}^{S}\left|\frac{\hat{\mu}_{st}-\mu_t}{\mu_{st}}\right| \times 100\%$, respectively, where $S = 500$ is the number of generated realizations, μ_t is the target value for the mean epidemic curve (step 1 of the simulation in Section 3) and $\hat{\mu}_{st}$ is the corresponding mean nowcast estimate at time t and simulation s. The prediction interval coverage is obtained by determining the percentage of (true) unreported cases that fall within the interval. Moreover, for each

simulation, the interval width is obtained, which is the difference between the upper and lower bound of the prediction interval.

In the simulation, we fix the nowcast date at the end of the month (from March to November). For each nowcast date, the prediction measures are computed for the dates having unreported cases, that is, for $t = T - (D - 1), \ldots, T$. As we have a maximum delay of 7 days, there will be seven dates (including the nowcast date) that involve the prediction of unreported cases. We only present here the simulation results on the nowcast day, that is, at time T. Tables 1-4 provide a summary of the results for the prediction measures on the nowcast date for the different functions being considered.

In terms of prediction interval, the Poisson model (LPS-P) generally has the narrowest prediction interval widths resulting in lower coverage rates for all functions. This is expected since we simulate the data from the negative binomial distribution with an overdispersion parameter which is not accounted for by the Poisson distribution. For function $f_1(t)$ (Table 1), the LPS-NB has more stable coverage rates and is closer to the nominal 95% prediction interval across all nowcast dates compared to VDK. The method of VDK on the other hand, tends to have high coverage rates resulting from wider prediction intervals, indicating higher uncertainty in its predictions. For functions $h_1(t)$ (Table 2) the PI coverage for LPS-NB and VDK are somewhat similar. The PI width of LPS-NB are wider from March to June and narrower from July to November compared to VDK.

The bias of the three methods varies across functions and months. LPS-NB and LPS-P often exhibit similar bias patterns, while VDK occasionally shows higher biases, particularly for function $f_1(t)$ during months like July and November. However, the biases are relatively small in magnitude. In terms of relative bias, VDK sometimes displays higher

values compared to LPS-NB and LPS-P, especially noticeable in function $f_1(t)$ during months like July and November. For functions $f_2(t)$ and $h_2(t)$ (Tables 3 and 4) with relatively high case numbers, we did not present the method of Van de Kassteele et al. (2019) since their results yield very wide prediction intervals and large relative bias. We believe that the reason for this is more of a computational issue rather than a methodological one. The LPS-NB model for these functions yield stable results with higher coverage rates, wider PI width, lower relative bias and generally lower bias compared to the LPS-P model.

Table 1: Performance measures on the nowcast date for function $f_1(t)$: **LPS-NB** - LPS model with a negative binomial distribution for the number of cases; **LPS-P** - LPS with a Poisson distribution for the number of cases; **VDK** - Methodology of Van de Kassteele et al. (2019).

	95% PI coverage			**PI width**			**Bias (μ)**			**% Bias (μ)**		
	LPS-NB	**LPS-P**	**VDK**	**LPS-NB**	**LPS-P**	**VDK**	**LPS-NB**	**LPS-P**	**VDK**	**LPS-NB**	**LPS-P**	**VDK**
March	95.8	78.6	99.8	48.9	32.7	74.9	1.7	3.1	4.0	32.1	47.7	70.2
Apr	97.6	86.0	99.2	20.7	15.6	28.2	-0.4	0.4	-1.6	23.9	30.9	32.3
May	96.6	73.0	100.0	41.5	25.2	73.7	-1.9	1.8	-3.0	19.3	26.4	27.9
June	96.0	54.0	99.6	98.6	42.2	226.5	-0.1	9.3	15.3	14.7	27.7	35.3
July	96.0	60.6	98.4	63.5	29.2	192.2	1.1	5.0	20.2	14.1	25.2	66.9
Aug	95.6	83.8	98.8	19.0	13.4	31.8	-0.7	0.2	-1.3	16.2	19.3	28.3
Sept	94.0	85.6	98.2	14.5	11.3	23.9	-1.3	-0.5	-1.9	19.2	19.1	31.9
Oct	95.6	72.6	99.2	38.6	21.3	69.2	-2.2	0.5	-3.0	14.8	16.8	23.7
Nov	97.2	48.8	99.0	96.6	34.0	232.5	4.7	4.9	21.7	14.5	17.8	43.0

Table 2: Performance measures on the nowcast date for function $h_1(t)$: **LPS-NB** - LPS model with a negative binomial distribution for the number of cases; **LPS-P** - LPS with a Poisson distribution for the number of cases; **VDK** - Methodology of Van de Kassteele et al. (2019).

	95% PI coverage			**PI width**			**Bias (μ)**			**% Bias (μ)**		
	LPS-NB	**LPS-P**	**VDK**	**LPS-NB**	**LPS-P**	**VDK**	**LPS-NB**	**LPS-P**	**VDK**	**LPS-NB**	**LPS-P**	**VDK**
March	96.6	81.2	95.6	26.5	19.1	13.2	1.3	2.0	-0.3	41.1	55.3	20.3
Apr	97.4	85.8	93.8	18.8	13.1	12.9	0.2	0.7	-0.9	25.5	33.1	23.2
May	97.2	78.4	95.2	26.4	16.4	20.9	-0.6	0.4	-1.7	20.4	25.4	19.2
June	95.8	75.2	95.2	40.5	21.7	39.6	-1.6	0.7	-0.5	16.6	21.4	15.3
July	96.4	72.0	96.4	40.0	21.7	49.2	-1.4	1.3	1.9	15.0	19.8	17.2
Aug	94.4	77.6	97.6	29.2	17.9	34.6	-1.6	0.4	-0.7	15.5	18.0	15.7
Sept	96.0	77.6	96.8	30.7	18.4	34.8	-2.3	-0.1	-2.6	16.5	17.5	18.5
Oct	96.2	66.0	97.8	54.7	25.8	69.2	-3.0	0.6	-1.1	13.9	16.4	14.7
Nov	97.6	49.4	96.4	94.4	34.3	157.1	0.2	3.8	16.7	11.3	16.6	30.9

Table 3: Performance measures on the nowcast date for function $f_2(t)$: **LPS-NB** - LPS model with a negative binomial distribution for the number of cases; **LPS-P** - LPS with a Poisson distribution for the number of cases.

	95% PI coverage		**PI width**		**Bias (μ)**		**% Bias (μ)**	
	LPS-NB	**LPS-P**	**LPS-NB**	**LPS-P**	**LPS-NB**	**LPS-P**	**LPS-NB**	**LPS-P**
March	96.8	49.4	296.2	147.5	12.5	32.9	30.0	71.6
Apr	98.0	54.8	147.8	72.6	2.0	17.4	19.5	47.1
May	98.4	50.8	142.2	63.7	-3.7	14.3	16.6	32.1
June	97.4	37.0	380.4	106.5	12.2	42.5	15.7	34.5
July	97.8	42.4	212.8	59.1	8.6	11.3	14.7	24.4
Aug	95.4	55.6	94.1	37.2	-3.5	3.6	14.0	19.8
Sept	96.6	55.8	90.8	35.2	-1.6	2.9	13.3	19.7
Oct	96.6	48.8	123.8	41.6	-5.6	0.7	13.0	17.4
Nov	97.8	36.6	373.4	71.2	28.7	14.4	17.7	19.8

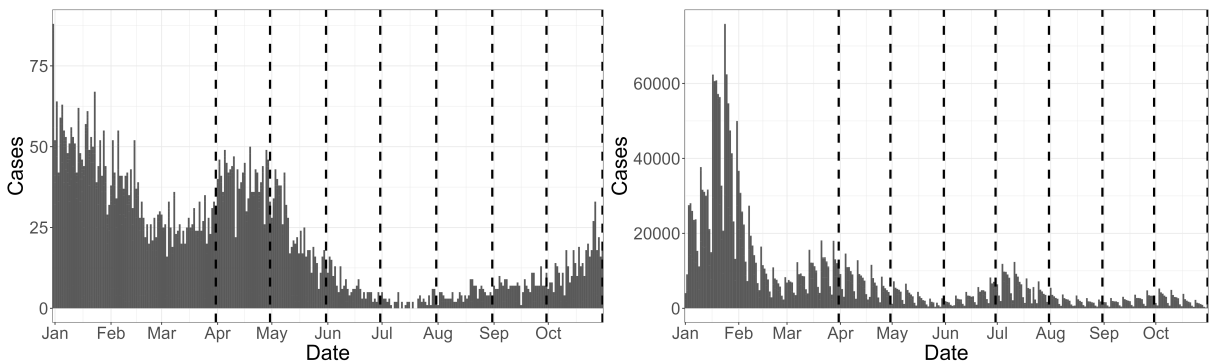
Table 4: Performance measures on the nowcast date for function $h_2(t)$: **LPS-NB** - LPS model with a negative binomial distribution for the number of cases; **LPS-P** - LPS with a Poisson distribution for the number of cases.

	95% PI coverage		**PI width**		**Bias (μ)**		**% Bias (μ)**	
	LPS-NB	**LPS-P**	**LPS-NB**	**LPS-P**	**LPS-NB**	**LPS-P**	**LPS-NB**	**LPS-P**
March	95.0	63.4	115.3	57.5	6.1	8.8	35.6	52.2
Apr	96.8	62.4	67.6	34.9	-0.3	4.1	20.0	32.9
May	97.6	53.8	103.6	45.6	-2.3	7.9	16.2	27.7
June	98.0	44.4	178.6	62.0	-1.1	16.8	14.0	28.3
July	97.0	43.8	184.6	56.6	2.2	12.6	13.5	24.0
Aug	97.2	49.0	129.2	45.0	-2.4	7.9	13.0	21.4
Sept	95.8	44.4	136.3	45.5	-4.6	6.2	12.5	20.4
Oct	97.6	38.2	257.7	63.0	-0.7	10.5	11.5	20.4
Nov	97.2	29.4	455.9	81.7	21.2	17.6	12.8	20.6

4 Real Data application

We apply our method to COVID-19 mortality data in Belgium for 2021 and incidence data for 2022. The raw data comes from the website of the Sciensano research institute (https://epistat.sciensano.be/covid/; Accessed December 20, 2022). The data contains the cumulative number of cases, reported up to the day of the file. The file is updated every day, and in this way, the number of cases and reporting delays are obtained. The data is structured in matrix format with the date of death/confirmed case as rows and number of days of reporting delay as columns. Figure 4 shows the total number of cases

with (dashed) vertical lines corresponding to different nowcast dates used for illustration. As no cases are reported immediately on the day of death/confirmed case in the data, all cases have a delay of at least one day. The proportion of cases reported for each delay is shown in Table 5. Majority of the cases (66% for mortality and 61% for incidence) were reported with a delay of 2 days. Followed by a delay of 1 day (14%) for mortality and 3 days (37%) for incidence. The rest of the delays account for very small percentages of the cases. The reporting delay is truncated to a maximum number of seven days.



(a) Mortality cases

(b) Incidence cases

Figure 4: Covid-19 death (a) and incidence (b) cases in Belgium. Dashed vertical lines correspond to different nowcast dates.

Table 5: Proportion of cases reported for each delay in the mortality data.

Delay (days)	0	1	2	3	4	5	6	7
Proportion	0.000	0.143	0.664	0.118	0.047	0.017	0.007	0.004

Table 6: Proportion of cases reported for each delay in the incidence data.

Delay (days)	0	1	2	3	4	5	6	7
Proportion	0.000	0.001	0.61	0.373	0.012	0.002	0.001	0.001

We choose $K_T = 40$ and $K_D = 10$ for the real data analysis. In addition, we also consider the day of the week effect as an additional covariate in the model. Specifically, we fit the following model $\log(\mu_{t,d}) = \beta_0 + \sum_{j=1}^{K_T} \sum_{k=1}^{K_D} \theta_{j,k} b_j(t) b_k(d) + \sum_{l=1}^{6} \beta_l z_l(t,d)$ as in (1), where $\sum_{l=1}^{6} \beta_l z_l(t,d)$ represents the day of the week with Monday taken as the reference

category. Algorithms to fit the model are available within the EpiLPS package (Gressani, 2021) through the *nowcasting()* routine.

Figure 5 presents the nowcasting results using the mortality data at different nowcast dates (namely, at every end of the month). The blue color represents the reported cases for the past 14 days on the nowcast date, the gray color represents cases that have not yet been reported, and the red points with error bars corresponds to the nowcast prediction with the 95% prediction interval. It can be seen that most of the nowcast predictions on the nowcast date are fairly close to the observed cases (gray). In addition, all of the observed cases fall within the prediction interval. For the incidence data, we used a maximum delay of 5 days. The nowcast results for the incidence data (Figure 6) tend to exhibit varying degrees of accuracy. This means that some nowcast values are close to the observed values, while others are further away, and some are moderately close. Notably, the nowcast estimate (on the nowcast day) for May and July are somewhat distant from the observed values.

The plot for the estimated delay density is shown in the supplementary material S5. The steps to obtain the time-varying delay distribution is outlined in Section 2.5. The plots show that the delay distribution is fairly constant through time for both incidence and mortality data. The density is highest for the delay of 2 days except in the beginning of the year for incidence data. This confirms the observed reporting intensity in our data that most cases are reported with a two-day delay.

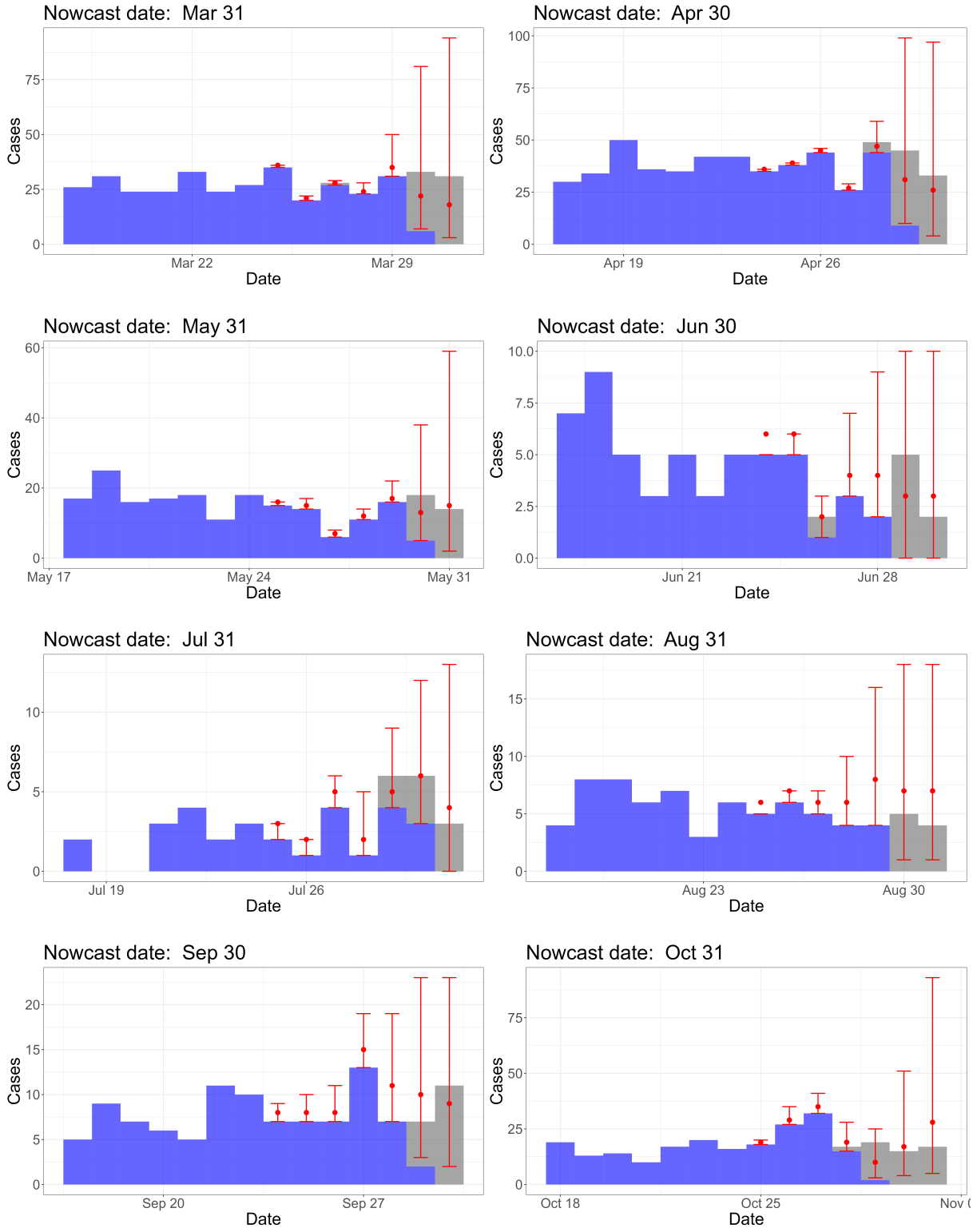


Figure 5: Nowcast for mortality data with different nowcast dates. Blue - reported cases ; Gray - not-yet-reported cases; Red points - nowcast estimates; Red error bar - 95% nowcast prediction interval.

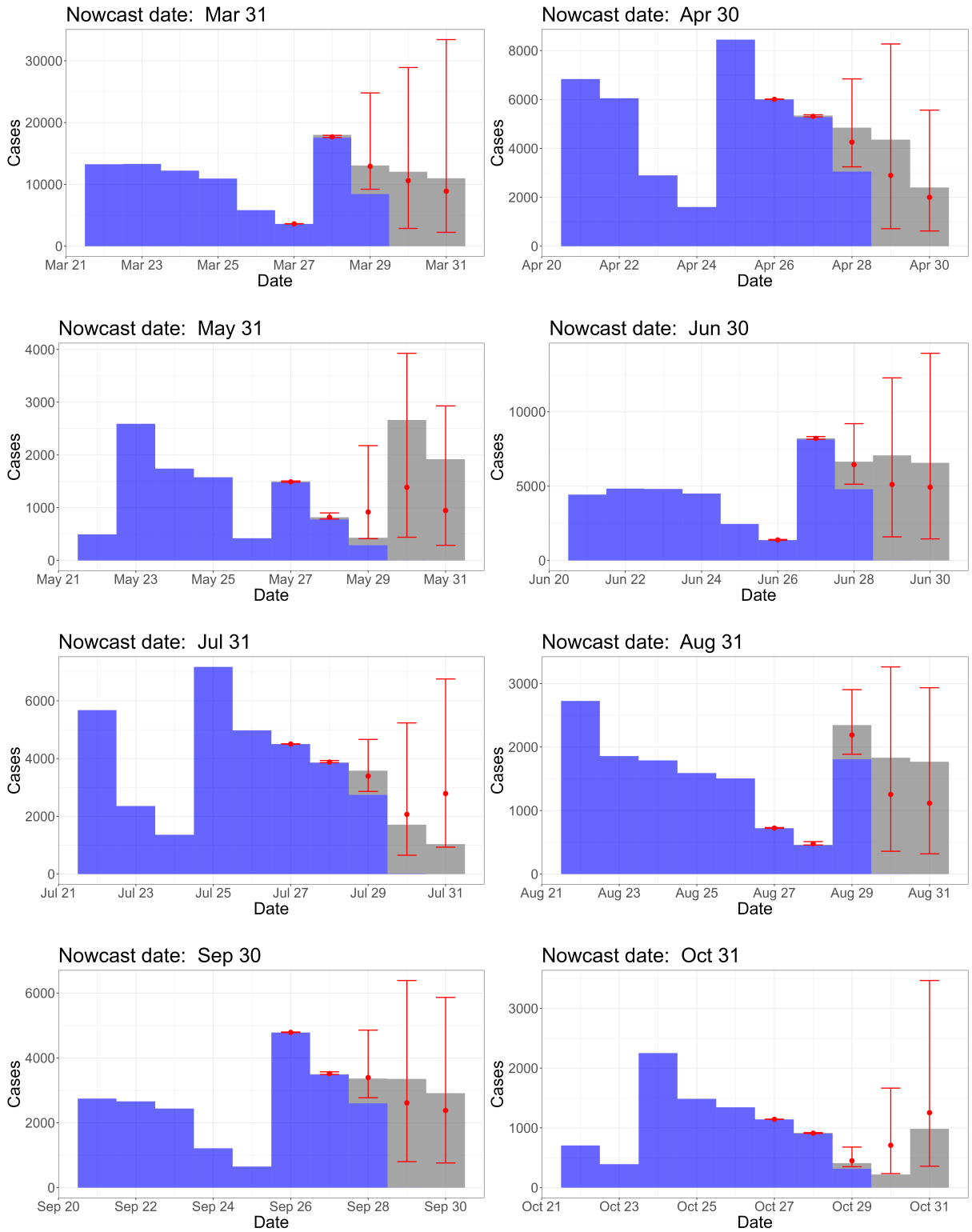


Figure 6: Nowcast for incidence data with different nowcast dates. Blue - reported cases ; Gray - not-yet-reported cases; Red points - nowcast estimates; Red error bar - 95% nowcast prediction interval.

5 Conclusion

This paper presents a novel Bayesian approach of the nowcasting model proposed by Van de Kassteele et al. (2019). During a pandemic or outbreak, it is very important to get an idea on the number of actual cases in real-time. One advantage of using the Bayesian framework is that prediction intervals can be naturally obtained. Our proposed methodology is based on a combination of Laplace approximations and P-splines, making it both fast and flexible.

Based on the results presented in this paper, Bayesian nowcasting with Laplacian-P-Splines seems to be a promising tool. We performed a simulation study to evaluate the (predictive) performance of our method. Simulation results demonstrate that our proposed method, under the assumption of a negative binomial distribution, consistently produces stable results with excellent prediction interval coverage that is closer to the nominal level and exhibits small bias. Moreover, we also applied our method to the COVID-19 mortality and incidence data in Belgium. The nowcast predictions for the mortality data seem to be fairly close to the actual observed values, and all of the 95% prediction intervals for the different nowcast dates being considered contained the observed values. In the case of incidence data with larger numbers of cases, the nowcast results tend to exhibit varying degrees of accuracy. However, it is important to take note that we have no data for the nowcast day, and there are very few reported cases (0.1%) with a one-day delay for the incidence data. This makes it much more difficult to produce accurate nowcast predictions.

While nowcasting can provide valuable real-time information and predictions, it also has certain drawbacks. Nowcasting heavily relies on real-time data, which may not always be readily available, or of good quality. Biases in the data (such as double-counting of cases) that are corrected at a later time, can impact the nowcasting prediction. In

addition, gaps in data collection, processing, or dissemination can impact the timeliness and effectiveness of nowcasting predictions. Despite these drawbacks, nowcasting has been a very important tool for obtaining timely information and short-term predictions. While the proposed nowcasting model is rather complex, we have coded the *nowcasting()* routine in the EpiLPS package (Gressani (2021)) to provide a user-friendly experience. Finally, one extension to the LPS methodology for nowcasting is to incorporate the nowcasting model into the estimation of the reproduction number in the method proposed by Gressani et al. (2022b). This has been currently implemented by Sumalinab et al. (2023). Another extension is to consider the correlation of cases for each time point (t). By doing this, the model can capture temporal dependencies and may provide better estimates of the variability. Furthermore, accounting for spatial correlation and other covariate effects would be valuable. Incorporating spatial information can account for local variations while considering additional covariates can provide a better understanding of the factors influencing disease spread.

Supplementary materials

Online Appendix: Contains details for the model matrices and parameters, derivations for the gradient and Hessian, hyperparameter optimizations, details for the LPS-Poisson model, supplementary figures, and tables. (supplement.pdf)

R scripts: The R codes used to implement the simulation study are available on GitHub through the link https://github.com/bryansumalinab/Laplacian-P-spline-nowcasting.git.

Funding Statement

VERDI: This project was supported by the VERDI project (101045989), funded by the European Union. Views and opinions expressed are however those of the author(s) only

and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

ESCAPE: This project was supported by the ESCAPE project (101095619), funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

Competing Interest Statement

The authors have declared no competing interest.

References

Agresti, A. (2013). *Categorical Data Analysis*. John Wiley & Sons.

De Angelis, D. and Gilks, W. (1994). Estimating acquired immune deficiency syndrome incidence accounting for reporting delay. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 157(1):31–40.

Donker, T., van Boven, M., van Ballegooijen, W. M., van't Klooster, T. M., Wielders, C. C., and Wallinga, J. (2011). Nowcasting pandemic influenza A/H1N1 2009 hospitalizations in the Netherlands. *European Journal of Epidemiology*, 26(3):195–201.

Durbán, M., Currie, I., and Eilers, P. H. C. (2002). Using P-splines to smooth two-dimensional Poisson data. In *Proceedings of 17th International Workshop on Statistical Modelling, Chania, Crete*, pages 207–214.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121.

Eilers, P. H. C., Marx, B. D., and Durbán, M. (2015). Twenty years of P-splines. *SORT: Statistics and Operations Research Transactions*, 39(2):149–186.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, Methods and Applications*. Springer Science & Business Media.

Glöckner, S., Krause, G., and Höhle, M. (2020). Now-casting the COVID-19 epidemic: The use case of Japan, March 2020. *medRxiv*.

Gressani, O. (2021). *EpiLPS: a fast and flexible Bayesian tool for estimating epidemiological parameters. [Computer Software]. https://epilps.com/*.

Gressani, O., Faes, C., and Hens, N. (2022a). Laplacian-P-splines for Bayesian inference in the mixture cure model. *Statistics in Medicine*, 41(14):2602–2626.

Gressani, O. and Lambert, P. (2018). Fast Bayesian inference using Laplace approximations in a flexible promotion time cure model based on P-splines. *Computational Statistics & Data Analysis*, 124:151–167.

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.

Gressani, O., Wallinga, J., Althaus, C. L., Hens, N., and Faes, C. (2022b). Epilps: A fast and flexible bayesian tool for estimation of the time-varying reproduction number. *PLoS computational biology*, 18(10):e1010618.

Günther, F., Bender, A., Katz, K., Küchenhoff, H., and Höhle, M. (2021). Nowcasting the COVID-19 pandemic in Bavaria. *Biometrical Journal*, 63(3):490–502.

Gutierrez, E., Rubli, A., and Tavares, T. (2020). Delays in death reports and their implications for tracking the evolution of COVID-19. *Available at SSRN 3645304*.

Höhle, M. and an der Heiden, M. (2014). Bayesian nowcasting during the STEC O104: H4 outbreak in Germany, 2011. *Biometrics*, 70(4):993–1002.

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 Gressani, O. (2023). Penalty parameter selection and asymmetry corrections to Laplace approximations in Bayesian P-splines models. *Statistical Modelling*, 23(5-6):409–423.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212.

Lawless, J. (1994). Adjustments for reporting delays and the prediction of occurred but not reported events. *Canadian Journal of Statistics*, 22(1):15–31.

Lindsey, J. (1996). Fitting bivariate intensity functions, with an application to modelling delays in reporting acquired immune deficiency syndrome. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 159(1):125–131.

Noufaily, A., Farrington, P., Garthwaite, P., Enki, D. G., Andrews, N., and Charlett, A. (2016). Detection of infectious disease outbreaks from laboratory data with reporting delays. *Journal of the American Statistical Association*, 111(514):488–499.

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.

Sumalinab, B., Gressani, O., Hens, N., and Faes, C. (2023). An efficient approach to nowcasting the time-varying reproduction number. *medRxiv*, pages 2023–10.

Van de Kassteele, J., Eilers, P. H. C., and Wallinga, J. (2019). Nowcasting the number of new symptomatic cases during infectious disease outbreaks using constrained P-spline smoothing. *Epidemiology*, 30(5):737.