Bayesian inference with Laplacian-P-splines

A methodology for fast and flexible estimation of key epidemiologic parameters

Oswaldo Gressani
Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Data Science Institute,
Hasselt University, Belgium.

Computational Bayesian Statistics Seminar

Flatiron Institute, New York

October 27, 2023

Outline

1. Motivation
2. Laplace approximations and P-splines in a Bayesian context
3. Estimation of the time-varying reproduction number
4. Estimation of the incubation distribution
5. Conclusion

1. Motivation

Motivation

Statistical methods play a determinant role in infectious disease epidemiology, as they provide the mathematical apparatus to bridge the gap between observed data and estimates of key epidemiologic quantities.

In this talk, the focus is on:

1. The time-varying reproduction number. $\mathcal{R}_t \rightarrow_{\text{def}}$ Average number of secondary cases generated by an infectious individual at time t . A key parameter for:
   - Monitoring transmissibility and infectiousness of diseases during outbreaks.
   - Parameterizing models to reach effective control and prevention measures.
   - Quantifying the probability of a pathogen's persistence upon arrival in a new location.

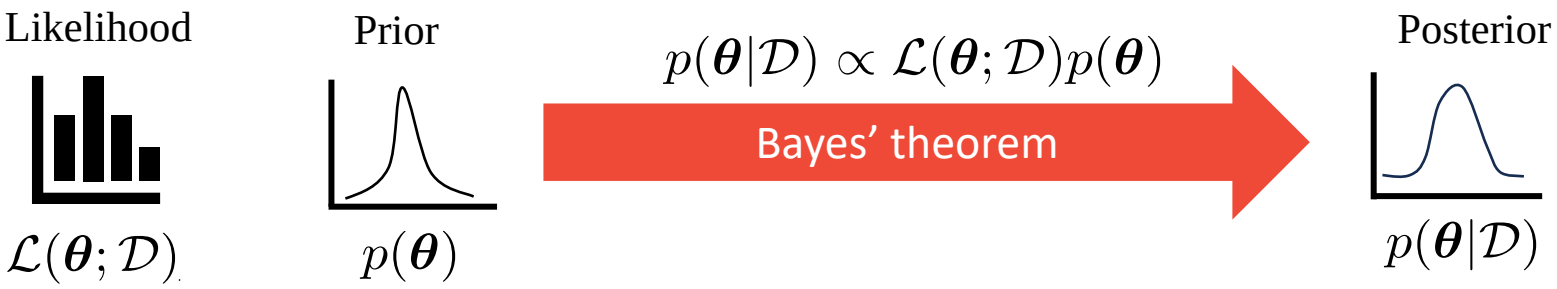
2. The incubation period. $\mathcal{I} \rightarrow_{\text{def}}$ Time between infection and symptom onset. A key quantity to:
   - Globally gauge the epidemic potential of an infectious disease.
   - Help planning optimal quarantine periods.
   - Assess the transmission potential of an infectious disease through the reproduction number.
   - Quantify the size of an epidemic.

Motivation

- Frequentist inferential approach dominated since the birth of epidemiology sparked by Daniel Bernoulli.

- Bayesian methods in “epi” models become increasingly popular:
  - Inclusion of prior information.
  - Uncertainties governing disease transmission mechanisms.
  - High processing power + multi-core architectures → facilitate implementation of MCMC methods.

The Bayesian philosophy



Motivation for Laplacian-P-splines

- Laplacian-P-splines (LPS) combine Laplace approximations and penalized B-splines or P-splines (Eilers and Marx, 1996; Lang and Brezger, 2004) in a unified framework.
  - Laplace approximations to posterior distributions are essentially Gaussian approximations.
  - P-splines allow for flexible estimation of functional components.
- Originally developed in the class of survival models and generalized additive models:

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 and 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 and Data Analysis*, **154**, 107088.

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

- Recent extensions to epidemic models (→ EpiLPS):

Gressani, O., Wallinga, J., Althaus, C., Hens, N. and Faes, C. (2022). EpiLPS: A fast and flexible Bayesian tool for estimation of the time-varying reproduction number. *PLoS Computational Biology*, **18**(10): e1010618.

Gressani, O., Faes, C. and Hens, N. (2023). An approximate Bayesian approach for estimation of the instantaneous reproduction number under misreported epidemic data. *Biometrical Journal*, **65**(6): 2200024.

Gressani, O., Torneri, A., Hens, N. and Faes, C. (2023). Flexible Bayesian estimation of incubation times. *MedRxiv preprint*.

2. Laplace approximations and P-splines in a Bayesian context

The Laplace approximation (1/3)

Born during Enlightenment period in Laplace's Mémoire sur la probabilité des causes par les événements (1774).

Mostly silent in statistical literature until its revival by Tierney and Kadane (1986) and INLA (Rue et al. 2009).

Consider a posterior distribution $p(\boldsymbol{\theta}|\mathcal{D})$ with $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{\dim(\boldsymbol{\theta})}$ and observables $\mathcal{D}$.

Second-order Taylor expansion of $h(\boldsymbol{\theta}) := \log p(\boldsymbol{\theta}|\mathcal{D})$ around $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$:

$$
\begin{aligned}
h(\boldsymbol{\theta}) &\approx h(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \nabla h(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \nabla^2 h(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\
&\approx c(\boldsymbol{\theta}_0) + \boldsymbol{\theta}^\top \left(\nabla h(\boldsymbol{\theta}_0) - \nabla^2 h(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0\right) + \frac{1}{2}\boldsymbol{\theta}^\top \nabla^2 h(\boldsymbol{\theta}_0)\boldsymbol{\theta}. \\
p(\boldsymbol{\theta}|\mathcal{D}) &\approx \exp\left\{\boldsymbol{\theta}^\top \left(\nabla h(\boldsymbol{\theta}_0) - \nabla^2 h(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0\right) + \frac{1}{2}\boldsymbol{\theta}^\top \nabla^2 h(\boldsymbol{\theta}_0)\boldsymbol{\theta}\right\}. \\
&\sim_a \mathcal{N}_{\dim(\boldsymbol{\theta})}\left(\boldsymbol{\theta}_0 - (\nabla^2 h(\boldsymbol{\theta}_0))^{-1}\nabla h(\boldsymbol{\theta}_0), -(\nabla^2 h(\boldsymbol{\theta}_0))^{-1}\right).
\end{aligned}
$$

Around the modal value $\boldsymbol{\theta}^*$ of $p(\boldsymbol{\theta}|\mathcal{D})$, we have $\widetilde{p}_G(\boldsymbol{\theta}|\mathcal{D}) \sim_a \mathcal{N}_{\dim(\boldsymbol{\theta})}\left(\boldsymbol{\theta}^*, -(\nabla^2 h(\boldsymbol{\theta}^*))^{-1}\right)$.

The Laplace approximation (2/3)

In practice, the Laplace approximation is sequentially implemented (e.g. Newton-Raphson, Levenberg-Marquardt).

Beware numerical pitfalls/convergence issues when locating the mode $\boldsymbol{\theta}^*$:

- Step-halving.
- Ascent direction.
- Positive definiteness of negative Hessian.
- Sensitivity to initial values.

Accuracy? Typically OK in "medium/large" samples due to the Bernstein-von Mises theorem and the Gaussian Markov field prior on latent variables.

If needed, use asymmetry corrections, see e.g.:

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

Need to compute analytical gradient/Hessian only once!

Much faster than MCMC.

The Laplace approximation (3/3)

Assume a bivariate-t distribution $(\boldsymbol{\theta}|\mathcal{D}) \sim t_\nu(\boldsymbol{\mu}, S)$ with $\nu = 4$ degrees of freedom.

$$h(\boldsymbol{\theta}) \doteq -\frac{(\nu+K)}{2}\log\left(1+\frac{1}{\nu}\boldsymbol{\theta}^\top S^{-1}\boldsymbol{\theta}-\frac{2}{\nu}\boldsymbol{\theta}^T S^{-1}\boldsymbol{\mu}+\frac{1}{\nu}\boldsymbol{\mu}^\top S^{-1}\boldsymbol{\mu}\right).$$

Gradient $\nabla h(\boldsymbol{\theta}) = -\frac{(\nu+K)}{\nu}\frac{1}{g(\boldsymbol{\theta})}S^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})$ and Hessian entries:

$$\frac{\partial^2 h(\boldsymbol{\theta})}{\partial\theta_1^2} = -\frac{(\nu+K)}{\nu}\frac{1}{g^2(\boldsymbol{\theta})}\left(S_{11}^{-1}z_1^{\partial}(\boldsymbol{\theta})-S_{12}^{-1}z_2^{\partial}(\boldsymbol{\theta})\right).$$

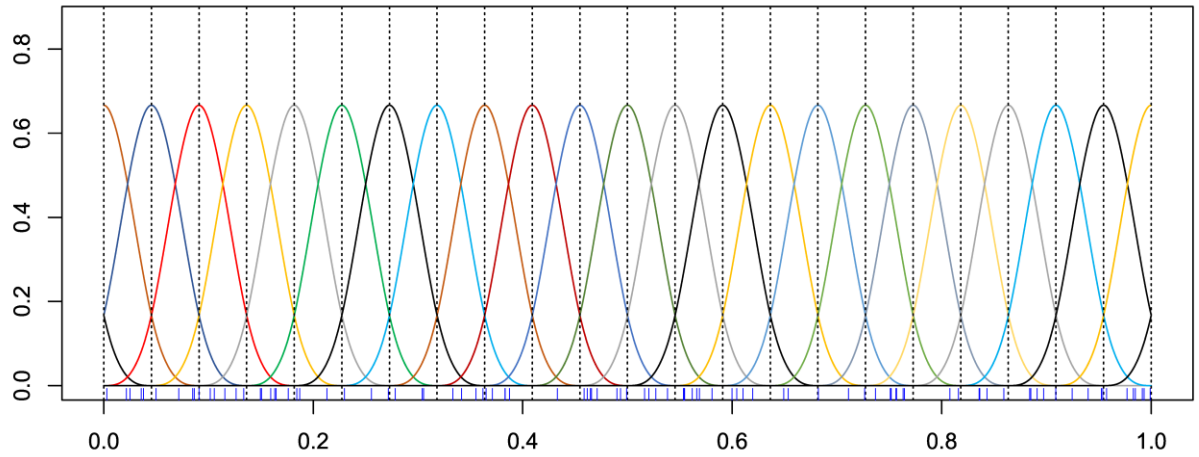
$$\frac{\partial^2 h(\boldsymbol{\theta})}{\partial\theta_2\partial\theta_1} = \frac{\partial^2 h(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2} = -\frac{(\nu+K)}{\nu}\frac{1}{g^2(\boldsymbol{\theta})}\left(-S_{11}^{-1}z_3^{\partial}(\boldsymbol{\theta})+S_{12}^{-1}z_4^{\partial}(\boldsymbol{\theta})\right).$$

$$\frac{\partial^2 h(\boldsymbol{\theta})}{\partial\theta_2^2} = -\frac{(\nu+K)}{\nu}\frac{1}{g^2(\boldsymbol{\theta})}\left(-S_{21}^{-1}z_3^{\partial}(\boldsymbol{\theta})+S_{22}^{-1}z_4^{\partial}(\boldsymbol{\theta})\right).$$

B-splines and Bayesian P-splines

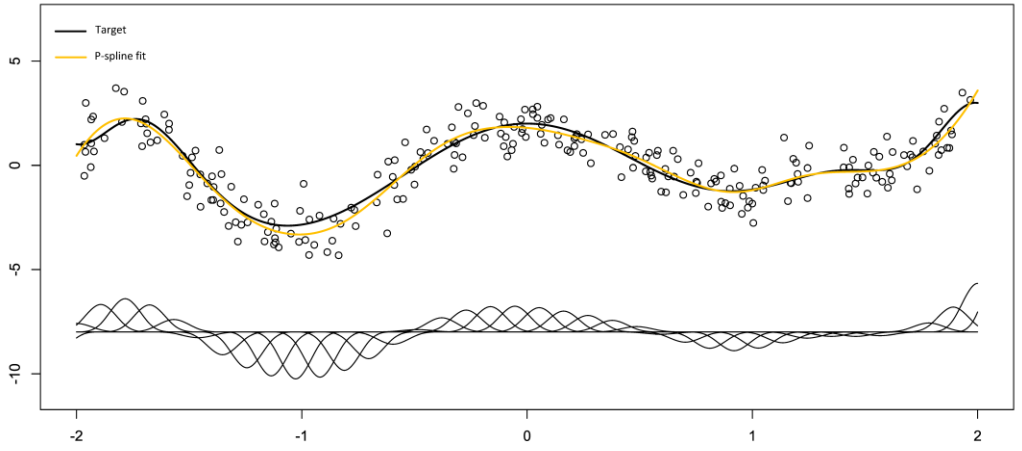
B-splines

- B-splines are used in a regression context to model smooth functional components.
- A B-spline of degree *d*:
  - Polynomial pieces joining together at *d* inner knots.
  - Each piece is a polynomial of degree *d*.
  - + on domain spanned by *d* + 2 knots.
  - Below: a cubic B-spline basis in [0, 1].



P-splines

- Eilers & Marx (1996): Use a large number of B-splines and balance the flexibility of the fit by a roughness penalty based on finite difference of contiguous B-spline coefficients.
- Lang & Brezger (2004): Bayesian version. Penalty is translated into a Gaussian prior for the B-spline coefficients.



The recipe of LPS

Latent Gaussian models: a vector of latent variables $\boldsymbol{\xi}$ (Gaussian prior $\boldsymbol{\xi}|\boldsymbol{\eta} \sim \mathcal{N}$) and hyperparameters $\boldsymbol{\eta}$.

Objective is to approximate the joint posterior of the latent vector:

$$p(\boldsymbol{\xi}|\mathcal{D}) = \int p(\boldsymbol{\xi}, \boldsymbol{\eta}|\mathcal{D})\, d\boldsymbol{\eta} = \int p(\boldsymbol{\xi}|\boldsymbol{\eta}, \mathcal{D})\, p(\boldsymbol{\eta}|\mathcal{D})\, d\boldsymbol{\eta}.$$

1. Laplace approximation to the conditional posterior:

$$p(\boldsymbol{\xi}|\boldsymbol{\eta}, \mathcal{D}) \rightarrow \widetilde{p}_G(\boldsymbol{\xi}|\boldsymbol{\eta}, \mathcal{D}).$$

2. Approximation of the hyperparameter vector:

$$p(\boldsymbol{\eta}|\mathcal{D}) = \frac{p(\boldsymbol{\xi}, \boldsymbol{\eta}|\mathcal{D})}{p(\boldsymbol{\xi}|\boldsymbol{\eta}, \mathcal{D})} \rightarrow \widetilde{p}(\boldsymbol{\eta}|\mathcal{D}) = \left.\frac{p(\boldsymbol{\xi}, \boldsymbol{\eta}|\mathcal{D})}{\widetilde{p}_G(\boldsymbol{\xi}|\boldsymbol{\eta}, \mathcal{D})}\right|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}.$$

3. Approximation of the posterior latent vector (grid-based, Maximum *a posteriori* MAP):

$$\begin{aligned}\widetilde{p}(\boldsymbol{\xi}|\mathcal{D}) &\approx \int \widetilde{p}_G(\boldsymbol{\xi}|\boldsymbol{\eta}, \mathcal{D})\, \widetilde{p}(\boldsymbol{\eta}|\mathcal{D})\, d\boldsymbol{\eta} \\ &\approx \sum_m \widetilde{p}_G\left(\boldsymbol{\xi}|\boldsymbol{\eta}^{(m)}, \mathcal{D}\right)\, \widetilde{p}\left(\boldsymbol{\eta}^{(m)}|\mathcal{D}\right)\, \Delta_m.\end{aligned}$$

3. Estimation of the time-varying reproduction number

Modeling the time-varying reproduction number

$\mathcal{R}_t \rightarrow_{\text{def}}$ Average number of secondary cases generated by an infectious individual at time t.

Gostic et al. (2020) give an elegant overview of existing methods and recommend the **EpiEstim** methodology (Cori et al., 2013) for real-time estimation of $\mathcal{R}_t$.

EpiEstim is a Bayesian approach assuming a Gamma distributed prior on the reproduction number with a Poisson likelihood on case counts (conjugacy eases generation of Markov chains).

Other recent tools for estimating $\mathcal{R}_t$:

- EpiNow2 package (Abbott et al., 2020).
- EpiFilter (Parag, 2021) recursive Bayesian smoother (Kalman filter).
- Pircalabelu (2021) builds an approach based on truncated polynomials and radial basis splines.

EpiLPS → a new Bayesian approach (**Epi**demiological modeling with **L**aplacian-**P**-**S**plines).

EpiLPS model

Let $\mathcal{D} = \{y_t, t = 1, \ldots, T\}$ denote the number of cases by reporting date.

Negative binomial assumption for y_t → accounts for overdispersion.

Gaussian prior on B-spline coeffs. and robust penalty priors (Jullion and Lambert, 2007).

$$
\begin{aligned}
y_t|\mu(t), \rho &\sim \text{NegBin}(\mu(t), \rho), \\
\log(\mu(t)) &= \boldsymbol{\theta}^\top b(t), \\
\boldsymbol{\theta}|\lambda &\sim \mathcal{N}_{\dim(\boldsymbol{\theta})}(0, Q_\lambda^{-1}), \\
\lambda|\delta &\sim \mathcal{G}\left(\phi/2, (\phi\delta)/2\right), \\
\delta &\sim \mathcal{G}(a_\delta, b_\delta), \\
\rho &\sim \mathcal{G}(a_\rho, b_\rho).
\end{aligned}
$$

Hyperparameter vector is $\boldsymbol{\eta} = (\lambda, \rho)^\top$.

EpiLPS has a two-step engine; in step (1) smooth case counts $\widehat{\mu}(t) = \exp(\widehat{\boldsymbol{\theta}}^\top b(t))$ and in step (2) plug the latter in a renewal equation model for $\mathcal{R}_t$.

Smoothing case counts

The estimated mean number of cases at a given time point t is obtained with Laplacian-P-splines.

First, write the log-likelihood of the model:

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

Gradient and Hessian of the (log) conditional posterior of the spline components are obtained analytically:

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

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

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

Newton-Raphson + MAP → $\widetilde{p}_G(\boldsymbol{\theta}|\boldsymbol{\eta}^*, \mathcal{D}) = \mathcal{N}_{\dim(\boldsymbol{\theta})} \left(\boldsymbol{\theta}^*(\boldsymbol{\eta}^*), \Sigma^*(\boldsymbol{\eta}^*) \right)$ → Estimator for B-spline coeffs. → $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^*(\widetilde{\boldsymbol{\eta}}^*)$.

Renewal equation model

Serial interval (SI) $=_{\text{def}}$ the time elapsed between the onset of symptoms in an infector and the onset of symptoms in the secondary cases generated by that infector.

Denote by $\boldsymbol{\varphi} = \{\varphi_1, \ldots, \varphi_k\}$ the (discrete) serial interval distribution, assumed known here.

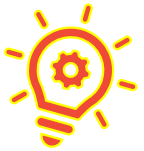
Let $\varphi_s = \mathbb{P}(SI = s)$ be the probability that the SI is equal to s day(s) and assume $\sum_{s=1}^{k} \varphi_s = 1$ and $\mathbb{P}(SI \leq 0) = \mathbb{P}(SI > k) = 0$.

Use renewal equation model $\mathbb{E}(y_t) = \mathcal{R}_t \left(\sum_{s=1}^{t-1} \varphi_s y_{t-s} \right)$, rearrange and plug-in:

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

Using the delta method, we show that $\left(\log R(t) | \mathcal{D} \right) \approx \mathcal{N}_1 \left(h(\boldsymbol{\theta}^* | t), \nabla^\top h(\boldsymbol{\theta} | t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \Sigma^* \nabla h(\boldsymbol{\theta} | t)|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right)$.

This is an entirely sampling-free methodology to obtain point and set estimates of $\mathcal{R}_t$.



A Langevinized Gibbs sampler

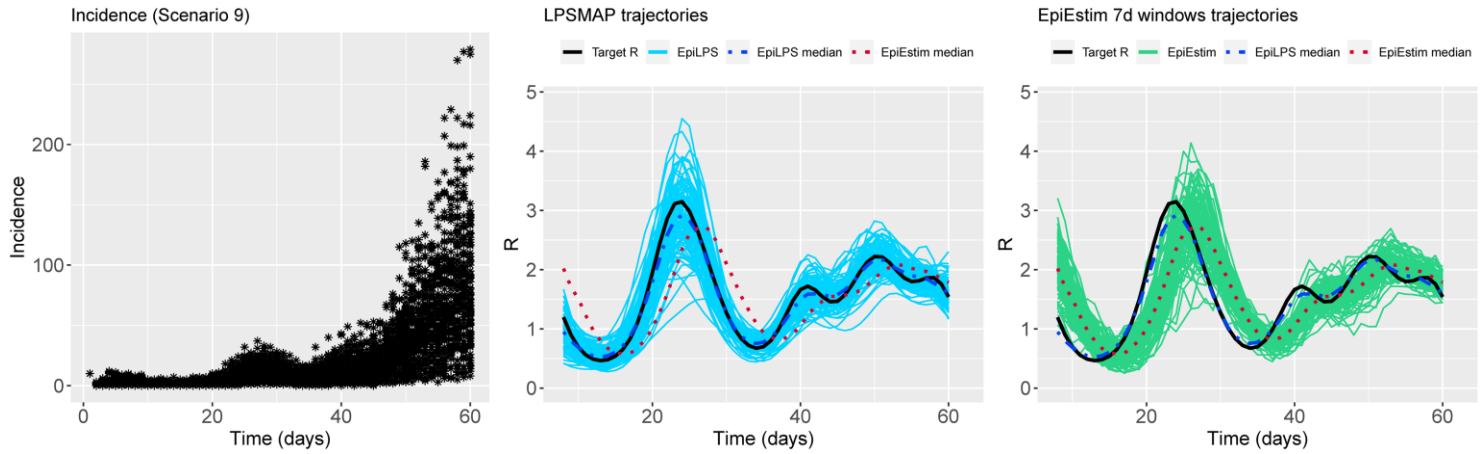
LPSMALA algorithm to sample the posterior $p(\boldsymbol{\theta}, \rho, \lambda, \delta|\mathcal{D})$.

1: Fix initial values $m = 0$, $\lambda^{(0)}$, $\delta^{(0)}$, $\varrho^{(0)}$ and $\widetilde{\boldsymbol{\zeta}}^{(0)} = (\boldsymbol{\theta}^{(0)\top}, w^{(0)})^\top$.
2: **for** $m = 1, \ldots, M$ **do**
3: (*Langevin-Hastings*)
4: Compute Langevin diffusion: $\mathcal{E}\left(\widetilde{\boldsymbol{\zeta}}^{(m-1)}\right) = \widetilde{\boldsymbol{\zeta}}^{(m-1)} + 0.5\varrho^{(m-1)}\Sigma_{LH}\nabla_{\widetilde{\boldsymbol{\zeta}}} \log p(\widetilde{\boldsymbol{\zeta}}|\lambda^{(m-1)}, \delta^{(m-1)}, \mathcal{D})\big|_{\widetilde{\boldsymbol{\zeta}}=\widetilde{\boldsymbol{\zeta}}^{(m-1)}}$
5: Generate a proposal: $\widetilde{\boldsymbol{\zeta}}^{(\text{prop})} \sim \mathcal{N}_{(K+1)}\left(\mathcal{E}\left(\widetilde{\boldsymbol{\zeta}}^{(m-1)}\right), \varrho^{(m-1)}\Sigma_{LH}\right)$.
6: Compute acceptance probability $\pi\left(\widetilde{\boldsymbol{\zeta}}^{(m-1)}, \widetilde{\boldsymbol{\zeta}}^{(\text{prop})}\right) = \min\left\{1, \frac{p(\widetilde{\boldsymbol{\zeta}}^{(\text{prop})}|\lambda^{(m-1)}, \delta^{(m-1)}, \mathcal{D})}{p(\widetilde{\boldsymbol{\zeta}}^{(m-1)}|\lambda^{(m-1)}, \delta^{(m-1)}, \mathcal{D})} \frac{q\left(\widetilde{\boldsymbol{\zeta}}^{(\text{prop})}, \widetilde{\boldsymbol{\zeta}}^{(m-1)}\right)}{q\left(\widetilde{\boldsymbol{\zeta}}^{(m-1)}, \widetilde{\boldsymbol{\zeta}}^{(\text{prop})}\right)}\right\}$
7: Draw $u \sim \mathcal{U}(0, 1)$.
8: **if** $u \leq \pi$ set $\widetilde{\boldsymbol{\zeta}}^{(m)} = \widetilde{\boldsymbol{\zeta}}^{(\text{prop})}$ (accept), **else** $\widetilde{\boldsymbol{\zeta}}^{(m)} = \widetilde{\boldsymbol{\zeta}}^{(m-1)}$ (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 $\varrho^{(m)} = \mathcal{H}^2\left(\sqrt{\varrho^{(m-1)}} + m^{-1}\left(\pi\left(\widetilde{\boldsymbol{\zeta}}^{(m-1)}, \widetilde{\boldsymbol{\zeta}}^{(\text{prop})}\right) - 0.57\right)\right)$.
14: **end for**

Simulations

Scenario	Time domain of epidemic curve	$\mathcal{R}_t$ target function	Serial Interval Mean (SD), days	Reference for Serial Interval
1	$\mathcal{T} = [1, 40]$	$\mathcal{R}_t = 1.3$	φ_{FLU} 2.6 (1.5)	Ferguson et al. (2005) [38] Cori et al. (2013) [3]
2		$\mathcal{R}_t = 2\ \mathbb{I}(t < 20) + 0.9\ \mathbb{I}(t \geq 20)$		
3		$\mathcal{R}_t = 0.25 + \exp(\cos(t/7))$		
4		$\mathcal{R}_t = \exp(\cos(t/15))$		
5	$\mathcal{T} = [1, 40]$	$\mathcal{R}_t = 1.3$	φ_{SARS} 8.4 (3.8)	Lipsitch et al. (2003) [39] Cori et al. (2013) [3]
6		$\mathcal{R}_t = 2\ \mathbb{I}(t < 20) + 0.9\ \mathbb{I}(t \geq 20)$		
7		$\mathcal{R}_t = 0.25 + \exp(\cos(t/7))$		
8		$\mathcal{R}_t = \exp(\cos(t/15))$		
9	$\mathcal{T} = [1, 60]$	$\mathcal{R}_t = 0.5\ (\exp(\sin(\pi t/9)) + 1.5\exp(\cos(4/t)))$	φ_{MERS} 6.8(4.1)	Cauchemez et al. (2016) [40]

https://doi.org/10.1371/journal.pcbi.1010618.t001

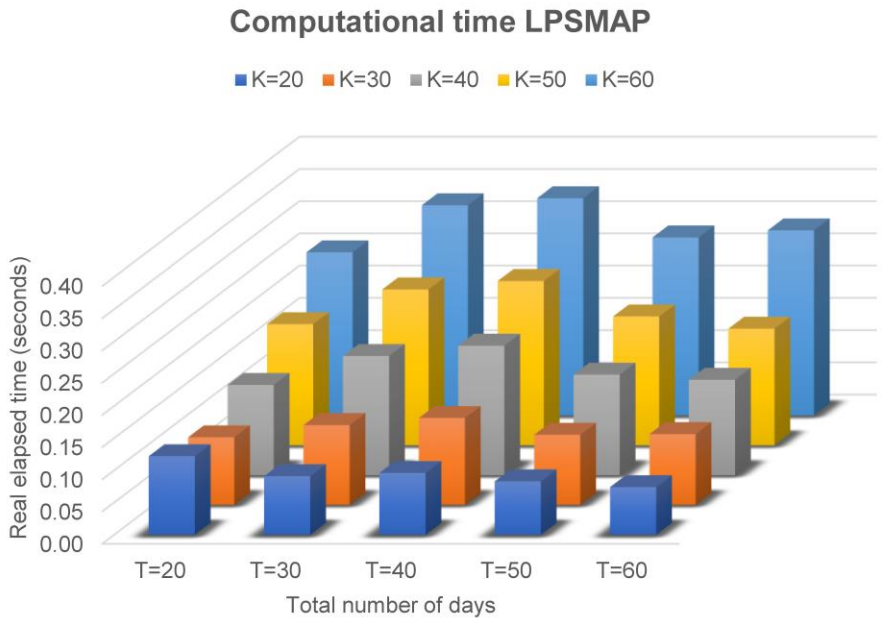


Computational time and real data application

Computational time

- EpiLPS relies on efficient algorithms and low computational cost is required to estimate the reproduction number.

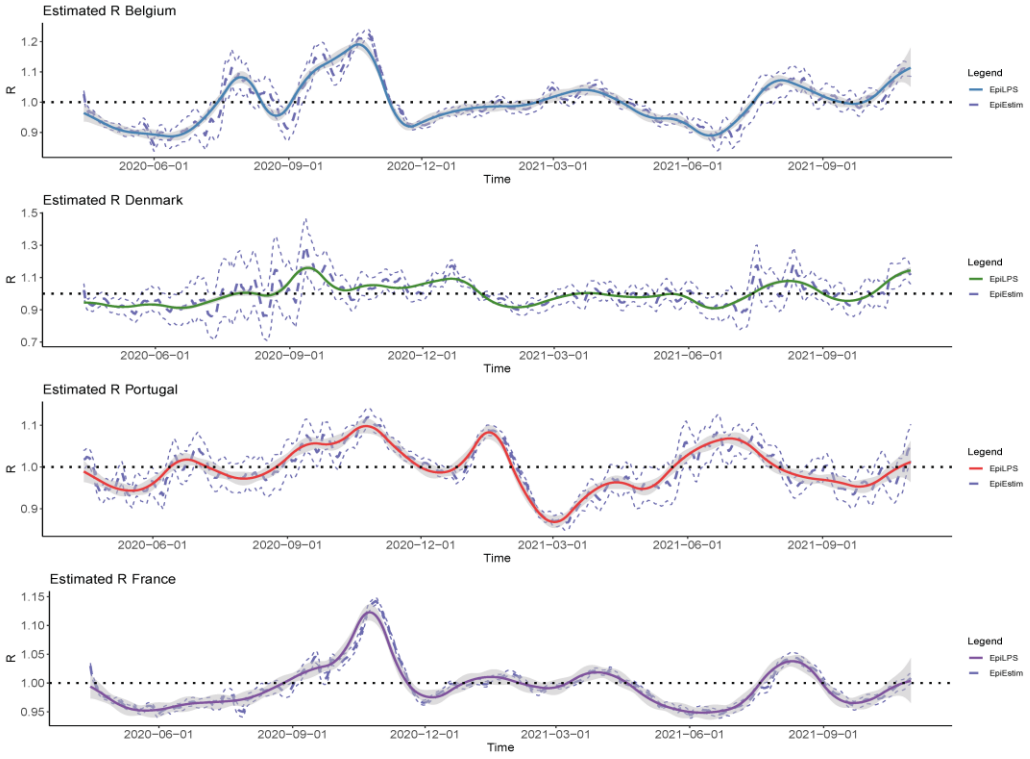
- Source code:
  - Modular structure.
  - Computational intensive routines coded in C++.
  - Integration via the Rcpp package.
  - LPSMAP faster than LPSMALA.



Application to SARS-CoV-2 data

- LPSMAP with K = 30 B-splines an second-order penalty.

- Estimated reproduction number from April 2020 to October 2021 for BEL, DEN, POR and FRA.



4. Estimation of the incubation distribution

Modeling the incubation distribution

The incubation period. $\mathcal{I} \rightarrow_{\text{def}}$ Time between infection and symptom onset.

Important contributions in modeling approaches dealing with interval-censored data:

- Peto (1973): Maximum likelihood with constrained Newton-Raphson.
- Turnbull (1976): EM algorithm to build a non-parametric estimate of the cdf under interval censoring.
- Sinha and Dey (1997): Review of semi-parametric Bayesian methods for interval-censored survival data.

More directly related to infectious disease epidemiology:

- Reich et al. (2009): Frequentist context to estimate the incubation distribution through AFT models.
- Backer et al. (2020) and Miura et al. (2022): Bayesian parametric approaches to estimate the incubation period of COVID-19 and Mpox, respectively.
- Kreiss and Van Keilegom (2022): Semi-parametric method via Laguerre polynomials.

Data

- Observed exposure interval for individual i denoted by $\mathcal{E}_i = [t_i^{E_L}, t_i^{E_R}]$.
- Model in continuous time with $0 \leq t_i^{E_L} < t_i^{E_R} < t_i^S < +\infty$.
- Data at resolution of individual i is $\mathcal{D}_i = \{t_i^{\mathcal{I}_L}, t_i^{\mathcal{I}_R}\}$ and for an information set of size n, $\mathcal{D} = \cup_{i=1}^n \mathcal{D}_i$.

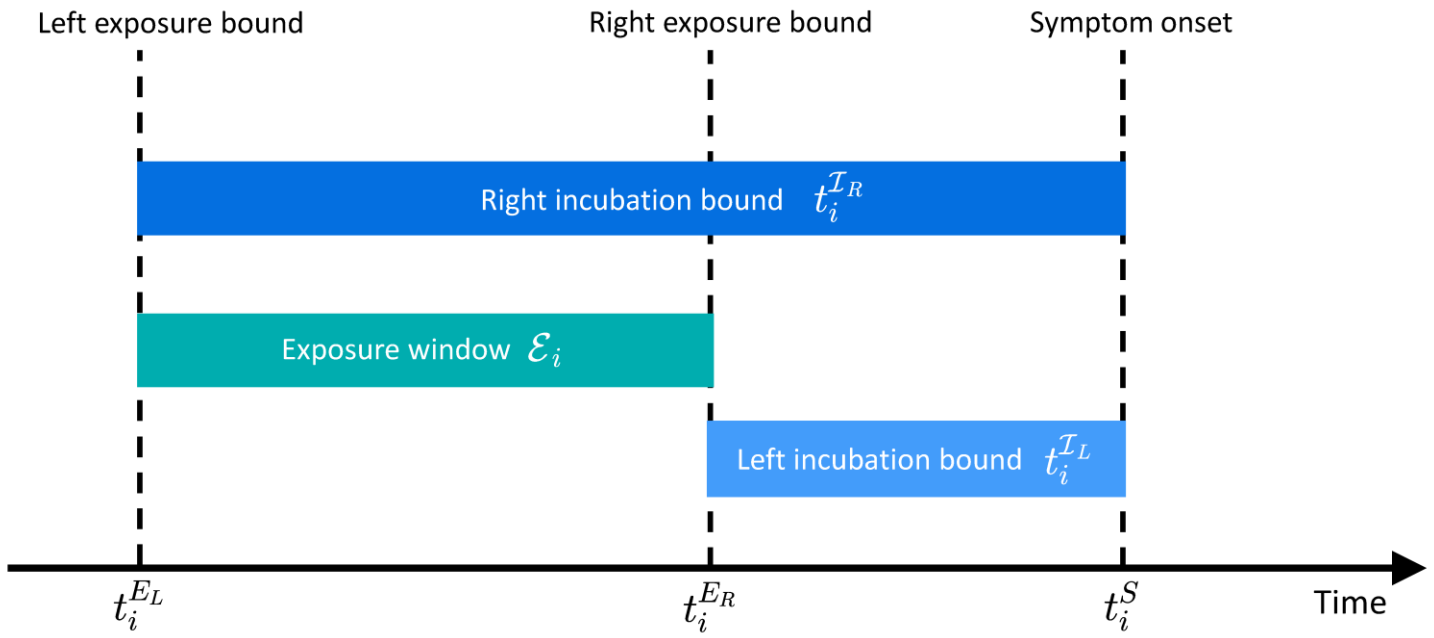


Figure. Relation between exposure window, incubation bounds and symptom onset time.

Model

Incubation time $\mathcal{I}$ → non-negative (continuous) random variable with:

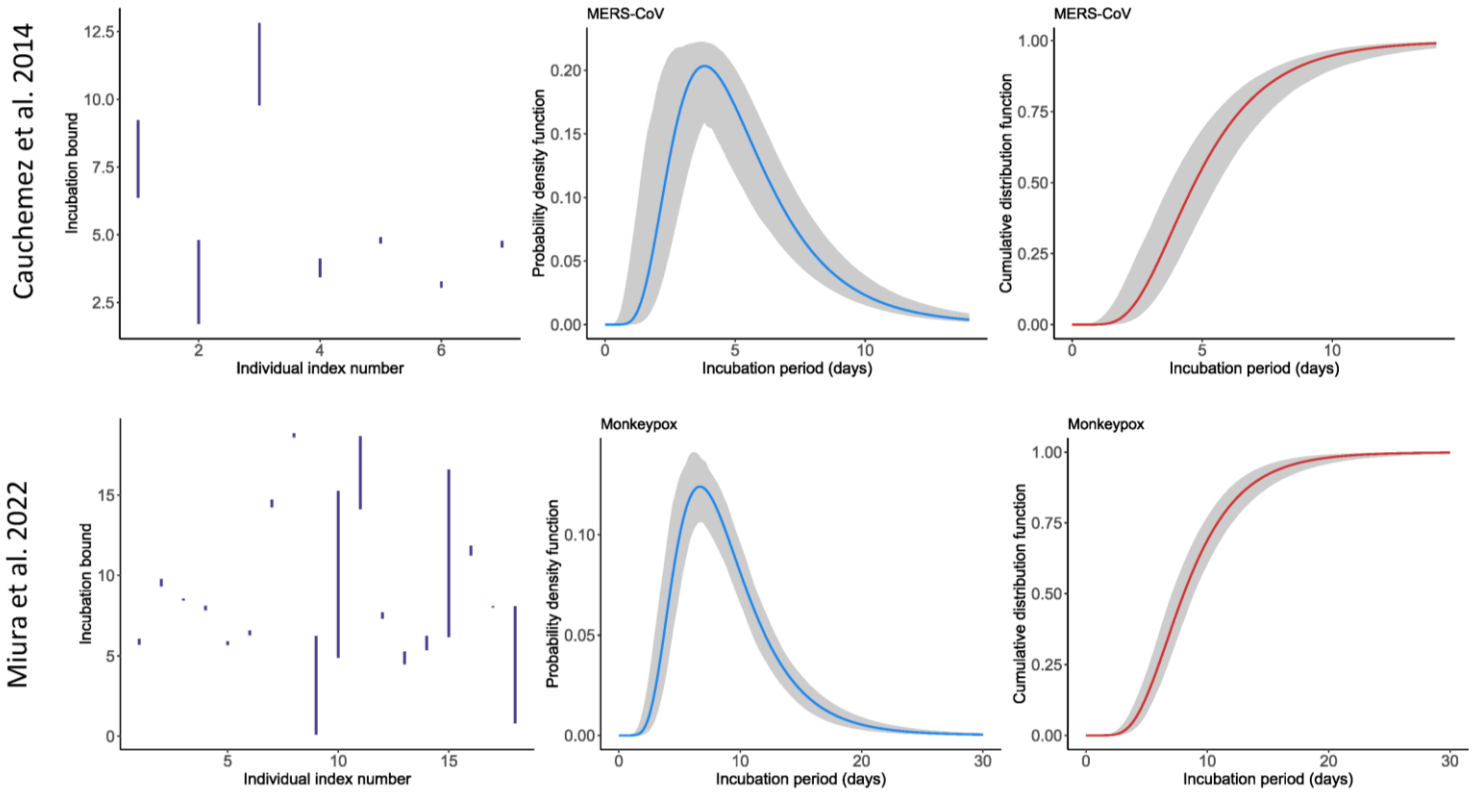
- probability density function $\varphi(\cdot)$, hazard function $h(\cdot)$ and survival function $S(\cdot)$.

Based on data $\mathcal{D}$, we propose a semi-parametric two-component mixture density estimator for $\varphi(\cdot)$ at $t \geq 0$:

$$\widehat{\varphi}_{SP}(t) = \omega\widehat{\varphi}_{IC}(t) + (1-\omega)\widehat{\varphi}_{HS}(t) \quad \text{with} \quad 0 \leq \omega \leq 1\,.$$

- The density estimator $\widehat{\varphi}_{IC}(\cdot)$ is based on the single interval-censored (IC) data.
- The density estimator $\widehat{\varphi}_{HS}(\cdot)$ uses a histogram smoother (HS) based on data from midpoint imputation.

Real data applications (MERS and Mpox)



	Mean incubation (95% CI)	95th percentile (95% CI)
Cauchemez et al. (2014)	5.5 (3.6-10.2)	10.2 (NA)
LPS	**5.3 (4.5-6.2)**	**10.1 (9.2-12.1)**
Miura et al. (2022)	9.0 (6.6-10.9)	17.3 (13.0-29.0)
LPS	**8.9 (7.9-9.9)**	**16.6 (14.7-19.1)**

5. Conclusion

Conclusion

Take home message

- Laplacian-P-splines (LPS) combine Laplace approximations to selected posterior distributions and Bayesian P-splines for fast and flexible inference in the class of latent Gaussian models.
- Attractive tool in infectious disease modeling (estimation of $\mathcal{R}_t$, estimation of the reproduction number under misreported data,...)
- LPS can also be used for nowcasting:

  🗒️ Sumalinab, B. Gressani, O., Hens, N. and Faes, C. (2023). Bayesian nowcasting with Laplacian-P-splines. *MedRxiv preprint*.

LPS vs INLA (Rue et al., 2009)

- INLA → marginal posterior distributions / LPS → joint posterior distributions.
- Algorithmic structure differs totally.
- LPS → analytical gradient/Hessian available, while INLA relies on numerical differentiation.
- LPS is exclusively built around P-splines smoothers (at least for the moment).

Conclusion

In “epi” models

- Fast and flexible Bayesian approach for estimating the time-varying reproduction number based on daily case count data and the incubation period distribution based on coarse data.
- Laplace approximation plays a central role → allows for sampling-free schemes + more efficient MCMC.
- Methodology is nested in the EpiLPS ecosystem https://epilps.com/.
- Routines already available in EpiLPS package (CRAN).
- Efficient algorithms and use of C++ (via Rcpp) improves speed.

EpiLPS (an acronym for **Epi**demiological modeling with **L**aplacian-**P**-**S**plines) is a tool for fast and flexible Bayesian estimation of epidemiological parameters. It can be used (among others) to estimate the epidemic cruve, the instantaneous reproduction number R_t and the incubation period of an infectious disease.

The methodology behind the R package can be found in [1] Gressani et al. 2022. The aim of this website is to give a short overview of the functionalities of EpiLPS.

The in-development version of the package is available on this **GitHub** repository. The stable version is available on **CRAN** downloads 10K downloads 240/month .

Associated literature

[1] Gressani O, Wallinga J, Althaus CL, Hens N, Faes C (2022) EpiLPS: A fast and flexible Bayesian tool for estimation of the time-varying reproduction number. PLoS Comput Biol 18(10): e1010618. 10.1371/journal.pcbi.1010618

[2] Gressani, O., Torneri, A., Hens, N. and Faes, C. (2023). Flexible Bayesian estimation of incubation times. MedRxiv preprint. 10.1101/2023.08.07.23293752

Selected references

[1] Reich, N. G., Lessler, J., Cummings, D. A., and Brookmeyer, R. (2009). Estimating incubation period distributions with coarse data. *Statistics in Medicine*, **28**(22): 2769–2784.

[2] Kreiss, A. and Van Keilegom, I. (2022). Semi-parametric estimation of incubation and generation times by means of Laguerre polynomials. *Journal of Nonparametric Statistics*, **34**(3): 570–606.

[3] 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.

[4] Gressani, O., Wallinga, J., Althaus, C. L., Hens, N., and Faes, C. (2022). EpiLPS: A fast and flexible Bayesian tool for estimation of the time-varying reproduction number. *PLOS Computational Biology*, **18**(10): e1010618.

[5] Gressani, O., Torneri, A., Hens, N. and Faes, C. (2023). Flexible Bayesian estimation of incubation times. *MedRxiv preprint*.

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

[7] Gressani, O., Faes, C. and Hens, N. (2023). An approximate Bayesian approach for estimation of the instantaneous reproduction number under misreported epidemic data. *Biometrical Journal*, **65**(6): 2200024.

[8] 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.