EPIDEMIC LOG GAUSSIAN COX PROCESS: LOCAL LAPLACE METHOD AND CLOSED-FORMS

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper introduces an epidemic multi-output Log Gaussian Cox Process ("LGCP") to model an epidemic's spread. It mixes a compartmental setting's attributes, which defines a class of workhorse models in epidemiology, such as the Susceptible-Infectious-Recovered-Deceased ("SIRD") model, and a non-parametric calibration of quantities that govern transmission. We consider multi-output LGCPs and introduce numerical approximations to compute posterior distributions extremely quickly and in a completely transparent and reproducible fashion. The model is tested on German Covid-19 data and shows good performance.

1 INTRODUCTION

The literature on modelling the spread of the Covid 19 epidemic is quickly growing, encompassing contributions set in a deterministic setting Cont et al. (2020) to probabilistic and deep setups Qian et al. (2020), using local or multi-country datasets Dong et al. (2020). Our paper builds on Hensman & Kypraios (2016) and considers Log Gaussian Cox Processes to model the changes in the number of people in each "compartment" (i.e., the number of susceptible, infectious, recovered or deceased people).

Others have considered the application of Bayesian non-parametrics to epidemiological modelling, for example in Vanhatalo et al. (2010); Xu et al. (2016); Teng et al. (2017). The main emphasis of these works is on numerical techniques to derive the posterior distribution. Our viewpoint is however mostly practical, as it is key to have access to closed-form and numerically stable expressions that can be implemented, run and checked extremely quickly by specialists and non-specialists alike.

Contributions First, we define a compartmental Log Gaussian Cox Process, which is an application of multi-output LGCPs and extend results from the single-compartment case. Second, we derive pseudo-closed form expressions for the posterior distribution of the compartmental LGCP, thus enabling quick and reproducible computations by policy makers. Third, we show that parsimony leads to good results and that more sophisticated techniques or calibrations may not be warranted on epidemiological datasets that are usually of small or medium size and may contain noisy data.

2 POISSON REGRESSION AND LOG GAUSSIAN COX PROCESS

Based on the Susceptible-Infectious-Recovered-Deceased model (Brauer et al. (2019); Brauer (2010); Li (2018) for textbook introductions to compartmental models), we consider time series of count data, S_t, I_t, R_t and D_t, representing, respectively, susceptible, infectious, recovered and deceased. In addition, we consider the daily *changes* in those quantities and denote those by lowercase letters, so that $i_t := I_t - I_{t-1}$ represents the number of new infectious people.

In keeping with the literature (we follow Hensman & Kypraios (2016) in particular), we posit Poisson-type dynamics for the daily changes in compartments' populations:

$$\begin{aligned} i_t &\sim \text{Poisson}(\lambda_t) \\ r_t &\sim \text{Poisson}(\gamma_t) \\ d_t &\sim \text{Poisson}(\delta_t), \end{aligned}$$

where $\lambda_t, \gamma_t, \delta_t$ are (random) hazard rate functions. This is a macro-level model in the sense that it ignores more refined segments that could be location or age-based. However, our setup can be extended to additional compartments.

We now specify the functions $\lambda_t, \gamma_t, \delta_t$ as exponentials of Gaussian processes (see Rasmussen & Williams (2006) for an overview of Gaussian Processes):

- $\log \lambda_t = f_t^{\lambda} \sim$ GP$(t, \log I_{t-1} + \log S_{t-1})$
- $\log \gamma_t = f_t^{\gamma} \sim$ GP$(t, \log I_{t-1})$
- $\log \delta_t = f_t^{\delta} \sim$ GP$(t, \log I_{t-1})$[1].

In other words, we model the latent features as Gaussian Process and use time as well as log infections and susceptible to build the kernel (for the sake of simplicity, we choose an RBF kernel and a zero mean function for the GPs but obtained very similar results with a Matern 3/2 kernel). We denote by $\mathbf{f}^{\xi}$ the $T \times 1$ vector of latent features for each $\xi = \lambda, \gamma, \delta$, by $\mathbf{y}^{\xi}$ the related vector of counts (i.e., new infections, new recoveries or new deaths), and the corresponding $T \times T$ kernel matrix by $\mathbf{K}^{\xi}$. At first, we assume that the three Gaussian processes are independent of each other. We relax this assumption later on by introducing a multi-output LGCP.

3 Deriving the posterior distribution

For all $\xi = \lambda, \gamma, \delta$, since $\mathbf{y}_t^{\xi} | \xi_t \sim$ Poisson(ξ_t) for each $t = 1, \cdots, T$, the posterior distribution of the latent feature vector $\mathbf{f}^{\xi}$, conditional on the observed vector of counts $\mathbf{y}^{\xi}$, is simply

$$p(\mathbf{f}^{\xi} | \mathbf{y}^{\xi}) \propto \exp\left({\mathbf{y}^{\xi}}^T \mathbf{f}^{\xi} - \sum_{i=1}^{T} e^{f_i^{\xi}} - \frac{1}{2} \left(\mathbf{f}^{\xi} - \mathbf{m}^{\xi}\right)^T {\mathbf{K}^{\xi}}^{-1} \left(\mathbf{f}^{\xi} - \mathbf{m}^{\xi}\right) \right), \tag{1}$$

as indicated in Diggle et al. (2013); Hensman & Kypraios (2016); Aglietti et al. (2019). Because this posterior distribution is not tractable, one must resort to numerical techniques to derive it.

3.1 Traditional techniques

While we introduce a method that has the benefit of being available in closed-form, other techniques such as the Laplace method, variational inference ("VI") or Markov Chain Monte Carlo ("MCMC") are usually employed (one is referred to Rasmussen & Williams (2006); Chatzilena et al. (2019); Hensman & Ghahramani (2015); Donner & Opper (2018); Xu et al. (2016); Teng et al. (2017) for further discussions and comparisons of these techniques for Gaussian Processes). We have thus run MCMC, the Laplace approximation and VI on the German Covid 19 data (see Section E for a brief overview).

3.2 A closed-form approximation

To increase speed and reproducibility, we introduce a technique related to the Laplace approximation, but allowing to bypass any optimisation step to derive closed-form expression. Indeed, we show in Section B of the Appendix that we can approximate the posterior distribution $p(\mathbf{f}^{\xi} | \mathbf{y}^{\xi})$ with

$$p(\mathbf{f}^{\xi} | \mathbf{y}^{\xi}) = \phi_{\boldsymbol{\mu}^{\xi}, \boldsymbol{\Sigma}^{\xi}}(\mathbf{f}^{\xi}), \tag{2}$$

where $\phi_{\boldsymbol{\mu}^{\xi}, \boldsymbol{\Sigma}^{\xi}}$ is the multivariate Gaussian density with parameters

$$\begin{aligned} \boldsymbol{\mu}^{\xi} &= \boldsymbol{\Sigma}^{\xi} \left({\mathbf{K}^{\xi}}^{-1} \mathbf{m}^{\xi} + \mathbf{H}^{\xi} \hat{\mathbf{f}}^{\xi} \right) \\ \boldsymbol{\Sigma}^{\xi} &= \left({\mathbf{K}^{\xi}}^{-1} + \mathbf{H}^{\xi} \right)^{-1}. \end{aligned}$$

Here, $\hat{\mathbf{f}}_t^{\xi} = \log(\mathbf{y}_t^{\xi})$ and $\mathbf{H}^{\xi} = \text{diag}(\mathbf{y}^{\xi})$.

Comparing results available in Figures 1 and 3, as well as the performance metrics in Section G

[1] We use GP$(t, \log I_{t-1})$ as shorthand notation for a Gaussian Process with zero mean whose kernel matrix entries depend on the feature vector (t, $\log I_{t-1}$)

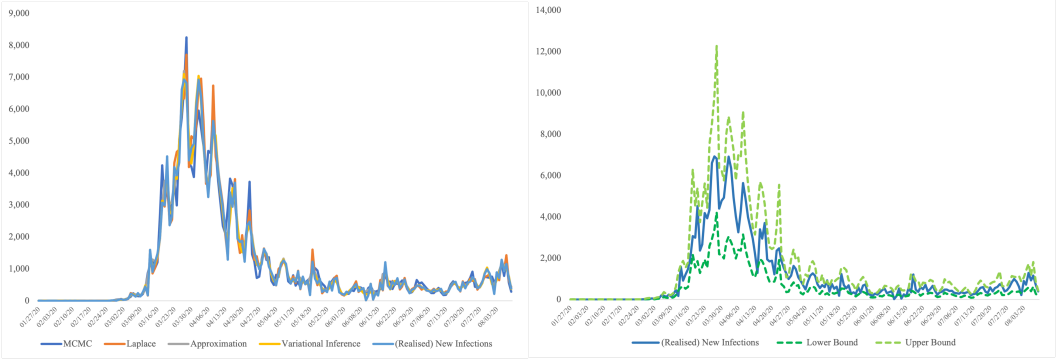

Figure 1: **(a)** Time series of new infections, realised and estimated via MCMC, Laplace method, our approximation and VI. **(b)** Time series of realised new infections with upper and lower bounds derived via VI as mean plus or minus two standard deviations.

4 Deriving posterior probabilities

In this section, we suppose that we have derived a (Gaussian) posterior distribution thanks to variational inference, the Laplace method or our proposed *local* Laplace method. Based on this posterior distribution, we know that $f_t^{\xi} = \log \xi_t \sim N(\mu_{\xi,t}, \sigma^2_{\xi,t})$, which is equivalent to $\xi_t \sim \text{Lognormal}(\mu_{\xi,t}, \sigma^2_{\xi,t})$, for $\xi = \lambda, \gamma, \delta$.

Since $\mathbf{y}_t^{\xi} | \xi_t \sim \text{Poisson}(\xi_t)$, by the tower property of conditional expectation and the law of total variance, it is straightforward to check that

$$\begin{aligned} \mathbb{E}[\mathbf{y}_t^{\xi}] &= e^{\mu_{\xi,t} + \frac{\sigma^2_{\xi,t}}{2}} \\ \mathbb{V}[\mathbf{y}_t^{\xi}] &= e^{\mu_{\xi,t} + \frac{\sigma^2_{\xi,t}}{2}} + e^{2\mu_t + \sigma^2_{\xi,t}} \left(e^{\sigma^2_{\xi,t}} - 1 \right). \end{aligned}$$

It is worth mentioning that the variance can be rewritten as

$$\mathbb{V}[\mathbf{y}_t^{\xi}] = \mathbb{E}[\mathbf{y}_t^{\xi}] + \mathbb{E}[\mathbf{y}_t^{\xi}]^2 \cdot \left(e^{\sigma^2_{\xi,t}} - 1 \right),$$

so that the second term –stemming from the uncertain value of ξ_t– is usually the main contributor to the total variance. Thus, using the plug-in estimator $\mathbb{E}[\mathbf{y}_t^{\xi}]$ for the variance (as is the case of a Poisson distribution with known paramter) can massively underestimate variance in LGCPs.

Lognormal and gamma distributions are fairly similar since they have the same domain (the non-negative real line) and have a constant coefficient of variation, and we thus approximate the former by the latter via moment-matching. In other words, suppose that $\xi_t \sim \text{Gamma}(\alpha_t, \beta_t)$, by performing moment-matching, we obtain that $\alpha_t = 1/(e^{\sigma_t^2} - 1)$ and $\beta_t = \exp(-\left(\mu_t + \frac{\sigma_t^2}{2}\right))/(e^{\sigma_t^2} - 1)$, for $\xi = \lambda, \gamma, \delta$.

This then enables us to derive the total distribution of $\mathbf{y}_t^{\xi}$ as

$$\mathbb{P}\left(\mathbf{y}_t^{\xi} = k\right) = \frac{\Gamma(\alpha_t + k)}{\Gamma(k+1)\Gamma(\alpha_t)} \left(\frac{\beta_t}{\beta_t + 1}\right)^{\alpha_t} \left(\frac{1}{\beta_t + 1}\right)^{k}, \tag{3}$$

which is a negative binomial distribution (see Section C for the full derivation). Note that combining our proposed local Laplace approximation with this expression implies that all quantities of interest are available in closed-form.

5 From multiple outputs to multi-output LGCP

We can combine the (vector) time series $\mathbf{f}^{\xi} \in \mathbb{R}^T$, for $\xi = \lambda, \gamma, \delta$ as a $T \times 3$ matrix

$$\mathbf{F} = \left[\mathbf{f}^{\lambda}, \mathbf{f}^{\gamma}, \mathbf{f}^{\delta}\right], \tag{4}$$

and define an *inter*-compartmental dependence structure complementing the covariances already inherited from the *intra*-compartmental kernels. In other words, information gathered from the dynamics of a particular compartment should be helpful in predicting the dynamics of other compartments; the reader is referred to Aglietti et al. (2019) for a description of multi-output and multi-task LGCPs.

Note that the case of independent LGCPs that we have studied so far is a particular instance of multi-output LGCP where the kernel matrix of $\mathbf{F}$ is block diagonal. We calibrated via VI a multi-output LGCP to the same dataset (see Section F of the Appendix for details). The results in Figure 2 show that the multi-output has similar but slightly smoother results

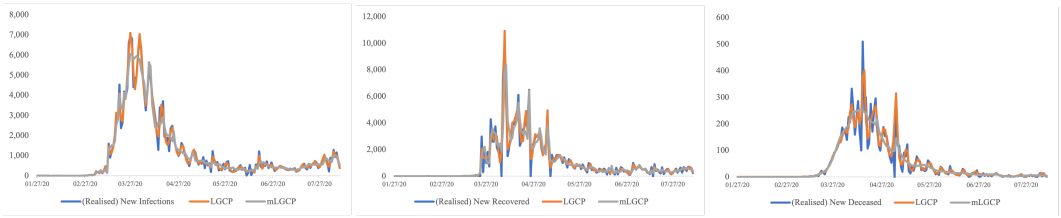

Figure 2: **(a)** Time series of realised new infections, estimated new infections via single-output LGCP, and estimated new infections via multi-output LGCP. **(b)** Time series of realised new recoveries, estimated new recoveries via single-output LGCP, and estimated new recoveries via multi-output LGCP. **(c)** Time series of realised new deaths, estimated new deaths via single-output LGCP, and estimated new deaths via multi-output LGCP.

We conjecture that the relatively poorer out-of-sample performance (see Section G) may be due to two different effects:

- Data is imperfect (and quite small) but reporting biases are often different from one compartment to the next (e.g., deaths reported more infrequently or with corrections);
- Independent LGCPs induce a sparse covariance matrix for $\mathbf{F}$ given its block-diagonal structure so that there are only $3T^2$ entries, whereas a full-blown multi-output LGCP requires $9T^2$ kernel entries, which may be more prone to over-fitting on small samples or less straightforward to fine-tune.

6 Discussion

Log Gaussian Cox Processes are useful tools for epidemiological modelling. Combining well-known multi-output compartmental models such as SIRD with a Bayesian non-parametric calibration results in a flexible and yet parsimonious tool. In the setting of one country only, Germany in our case,

From a technical standpoint, it seems that a multi-output kernel can be a useful tool, but given the relatively small size of epidemiological samples and their (sometimes) low data quality, it does not make a large difference in our setting. Multi-Output LGCPs could prove more effective in dealing with high-dimensional and spatial settings, for instance in tracking the spread of an epidemic such as Covid 19 across different regions.

Whether a model should reflect some of the data "quirks", for instance in terms of reporting frequency and behaviour seasonality (e.g., no reporting of deceased on Sundays, higher mobility during the working week, etc.), or should focus on trends (and be "smoother"), is a matter of choice. We do not believe that there is a purely data-based answer and this amounts to a socio-techical choice on the part of policy makers.

Finally, we have introduced a number of approximations to allow for the extremely fast and reproducible computation of the posterior distribution of log hazard rates, and thus the distribution of the stochastic epidemic process.

REFERENCES

Virginia Aglietti, Theodoros Damoulas, and Edwin Bonilla. Efficient Inference in Multi-task Cox Process Models. *PMLR*, 89, March 2019. URL http://arxiv.org/abs/1805.09781. arXiv: 1805.09781.

Fred Brauer (ed.). *Mathematical Epidemiology*. Springer Berlin Heidelberg, Berlin, 2008th edition edition, February 2010. ISBN 978-3-540-78910-9.

Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical Models in Epidemiology*. Springer, 1st ed. 2019 edition edition, October 2019. ISBN 978-1-4939-9826-5.

Anastasia Chatzilena, Edwin van Leeuwen, Oliver Ratmann, Marc Baguelin, and Nikolaos Demiris. Contemporary statistical inference for infectious disease models using Stan. *Epidemics*, 29: 100367, December 2019. ISSN 1755-4365. doi: 10.1016/j.epidem.2019.100367. URL https://www.sciencedirect.com/science/article/pii/S1755436519300325.

R Cont, R Xu, and A Kotlicki. Modelling covid-19 contagion: Risk assessment and targeted mitigation policies. 2020. doi: 10.1101/2020.08.26.20182477. URL http://dx.doi.org/10.1101/2020.08.26.20182477.

Peter J. Diggle, Paula Moraga, Barry Rowlingson, and Benjamin M. Taylor. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. *Statistical Science*, 28(4):542–563, November 2013. ISSN 0883-4237, 2168-8745. doi: 10.1214/13-STS441. URL https://projecteuclid.org/journals/statistical-science/volume-28/issue-4/Spatial-and-Spatio-Temporal-Log-Gaussian-Cox-Processes--Extending/10.1214/13-STS441.full. Publisher: Institute of Mathematical Statistics.

Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5): 533–534, May 2020. ISSN 1473-3099, 1474-4457. doi: 10.1016/S1473-3099(20)30120-1. URL https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30120-1/abstract. Publisher: Elsevier.

Christian Donner and Manfred Opper. Efficient Bayesian Inference of Sigmoidal Gaussian Cox Processes. *Journal of Machine Learning Research*, 19(67):1–34, 2018. URL http://jmlr.org/papers/v19/17-759.html.

James Hensman and Zoubin Ghahramani. Scalable Variational Gaussian Process Classification. *JMLR: W&CP*, 38:10, 2015.

James Hensman and Theodore Kypraios. Variational Bayesian non-parametric inference for infectious disease models. In *Machine Learning for Healthcare Technologies*, pp. 181–202. October 2016. URL https://digital-library.theiet.org/content/books/10.1049/pbhe002e_ch9. Publisher: IET Digital Library.

James Hensman, Nicolo Fusi, and Neil Lawrence. Gaussian processes for big data. *Uncertainty in Artificial Intelligence - Proceedings of the 29th Conference, UAI 2013*, 09 2013.

Michael Li. *An Introduction to Mathematical Modeling of Infectious Diseases: 2*. Springer, New York, NY, 1st ed. 2018 edition edition, February 2018. ISBN 978-3-319-72121-7.

Dong C. Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45:503–528, 1989.

Zhaozhi Qian, Ahmed M Alaa, and Mihaela van der Schaar. When and How to Lift the Lockdown? Global COVID-19 Scenario Analysis and Policy Assessment using Compartmental Gaussian Processes. *Advances in Neural Information Processing Systems*, pp. 12, 2020.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Mass, January 2006. ISBN 978-0-262-18253-9.

Ming Teng, Farouk Nathoo, and Timothy D. Johnson. Bayesian computation for Log-Gaussian Cox processes: a comparative analysis of methods. *Journal of Statistical Computation and Simulation*, 87(11), 2017. URL https://www.tandfonline.com/doi/abs/10.1080/00949655.2017.1326117.

Mark van der Wilk, Vincent Dutordoir, ST John, Artem Artemev, Vincent Adam, and James Hensman. A framework for interdomain and multioutput Gaussian processes. *arXiv:2003.01115*, 2020. URL https://arxiv.org/abs/2003.01115.

Jarno Vanhatalo, Ville Pietiläinen, and Aki Vehtari. Approximate inference for disease mapping with sparse Gaussian processes. *Statistics in Medicine*, 29(15):1580–1607, 2010. ISSN 1097-0258. doi: https://doi.org/10.1002/sim.3895. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3895. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.3895.

Xiaoguang Xu, Theodore Kypraios, and Philip D O'Neill. Bayesian non-parametric inference for stochastic epidemic models using gaussian processes. *Biostatistics*, 17(4):619–633, October 2016. ISSN 1465-4644. doi: 10.1093/biostatistics/kxw011. URL https://doi.org/10.1093/biostatistics/kxw011.

A Results for other compartments

We provide here the results obtained by fitting independent LGCPs to new recoveries and new deaths. We note that the performance of the model is similar to what we obtained in the case of new infections.

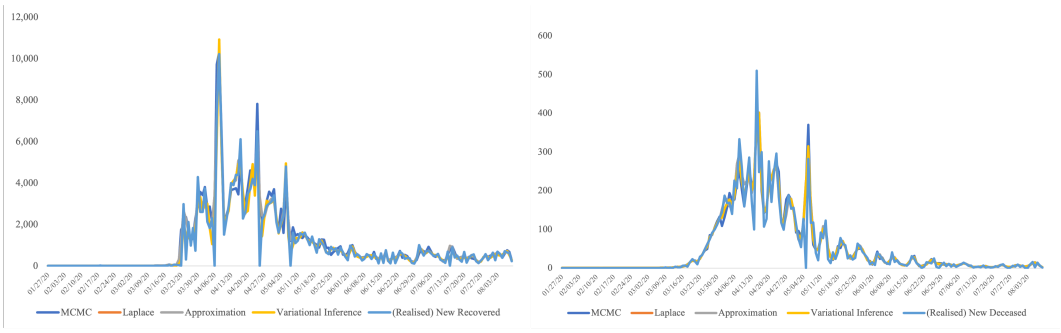

Figure 3: **(a)** Time series of new recoveries, realised and estimated via MCMC, Laplace method,our approximation and VI. **(b)** Time series of new deaths, realised and estimated via MCMC, Laplace method,our approximation and VI.

B Local Laplace method

We notice that the posterior distribution can be expressed as

$$p(\mathbf{f}|\mathbf{y}) \propto \exp\left(\mathbf{y}^T\mathbf{f} - \sum_{i=1}^{T} e^{f_i} - \frac{1}{2}\left(\mathbf{f}-\mathbf{m}\right)^T \mathbf{K}^{-1}\left(\mathbf{f}-\mathbf{m}\right)\right). \tag{5}$$

Now, if we consider the function $g(\mathbf{f}) = \sum_{i=1}^{T} e^{f_i} - \mathbf{y}^T\mathbf{f}$, we notice that it admits an optimum at $\hat{\mathbf{f}}$ defined as $\hat{f}_i = \log(y_i)^2$. The Hessian matrix is correspondingly given by $\mathbf{H} = \text{diag}\left(\mathbf{y}\right)$.

We can thus approximate the posterior density as

$$p(\mathbf{f}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{f}-\hat{\mathbf{f}}\right)^T \mathbf{H}\left(\mathbf{f}-\hat{\mathbf{f}}\right) - \frac{1}{2}\left(\mathbf{f}-\mathbf{m}\right)^T \mathbf{K}^{-1}\left(\mathbf{f}-\mathbf{m}\right)\right), \tag{6}$$

[2]If $y_i = 0$, one can simply set a lower threshold $\varepsilon > 0$

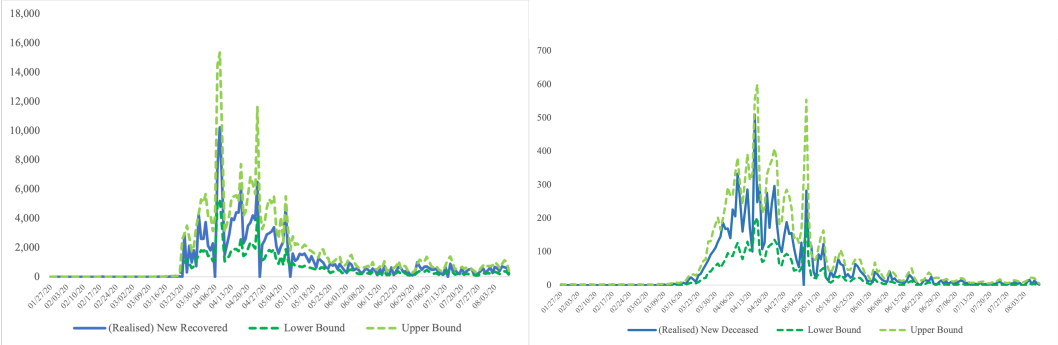

Figure 4: **(a)** Time series of realised new recoveries with upper and lower bounds, derived via VI as mean plus or minus two standard deviations. **(b)** Time series of realised new infections with upper and lower bounds, derived via VI as mean plus or minus two standard deviations.

which is equivalent, by completing the square, to

$$p(\mathbf{f}|\mathbf{y}) = \phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{f}), \tag{7}$$

where $\phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ is the multivariate Gaussian density with parameters

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{\Sigma}\left(\mathbf{K}^{-1}\mathbf{m} + \mathbf{H}\hat{\mathbf{f}}\right) \\ \boldsymbol{\Sigma} &= \left(\mathbf{K}^{-1} + \mathbf{H}\right)^{-1}. \end{aligned}$$

We thus have a closed-form approximation to the posterior distribution of $\mathbf{f}$. We call it a *local* Laplace method in the sense that we apply the Laplace method to the likelihood directly and not to the posterior density.

The interest of applying the local Laplace method to the LGCP framework comes from the fact that –in that case– $\mathbf{H}$ is diagonal, hence very easy to manipulate.

C GAMMA APPROXIMATION

Replacing the lognormal density with the moment-matching Gamma density leads to

$$\begin{aligned} \mathbb{P}\left(\mathbf{y}_t = k\right) &= \frac{1}{k!}\int_0^{+\infty} e^{-\lambda}\lambda^k p_{\lambda_t}(\lambda)d\lambda \\ &= \frac{1}{k!}\frac{\beta_t^{\alpha_t}}{\Gamma(\alpha_t)}\int_0^{+\infty} \lambda^{\alpha_t+k-1}e^{-(\beta_t+1)\lambda}d\lambda \\ &= \frac{\Gamma(\alpha_t+k)}{\Gamma(k+1)\Gamma(\alpha_t)}\left(\frac{\beta_t}{\beta_t+1}\right)^{\alpha_t}\left(\frac{1}{\beta_t+1}\right)^{k}, \end{aligned}$$

which is the desired result and corresponds to a negative binomial distribution with $r = \alpha_t$ number of failures, and a probability of success of $p_t = \frac{1}{\beta_t+1}$.

D FEATURE SELECTION

Our choice of functional form for the hazard rates is driven by the fact that we wish to remain close to proposed models of epidemics, in particular SIRD (cf. Brauer et al. (2019); Li (2018)). In this context, λ_t is the force of infection and specified as $\lambda_t = \beta S_t I_t$, which can be rewritten as $\log(\lambda) = \log(\beta)+\log(S_t)+\log(I_t)$. We thus choose to make λ_t a function of $\log(S_{t-1})+\log(I_{t-1})$.

We add time as a feature as it enables us to add a flexible dependence structure accounting for changes in policies (lockdowns, confinements, quarantines, vaccination programmes, etc.) and behaviours (reduced mobility for example). In this way, two days with the same number of susceptible and infectious people will not necessarily be similar. A discussion of alternative formulations is available in Hensman & Kypraios (2016).

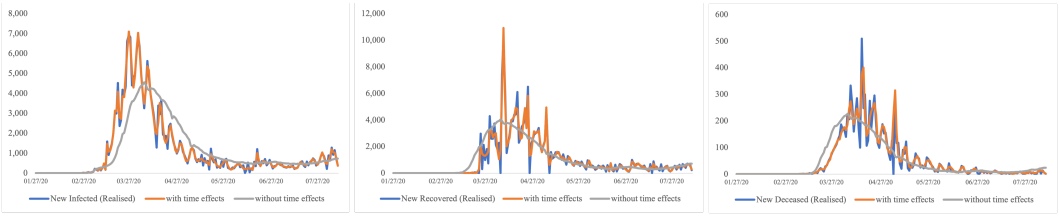

Figure 5: **(a)** Time series of realised new infections, estimated new infections via single-output LGCP *without* time as input to the kernel, and estimated new infections via single-output LGCP *with* time as input to the kernel. **(b)** Time series of realised new recoveries, estimated new recoveries via single-output LGCP *without* time as input to the kernel, and estimated new recoveries via single-output LGCP *with* time as input to the kernel. **(c)** Time series of realised new deaths, estimated new deaths via single-output LGCP *without* time as input to the kernel, and estimated new deaths via single-output LGCP *with* time as input to the kernel.

E DATA SPECIFICATION

The data that we have considered in this paper comes from the John Hopkins University (Dong et al. (2020)) repository for Covid 19 cases (including cumulative infections, recoveries and deaths) in Germany from 27 January 2020 until 10 August 2020. We have used as susceptible the whole German population of ca. 83 million people and accordingly reconstructed the time series for new infections, new recoveries and new deaths.

F ALGORITHMS

F.1 VARIATIONAL INFERENCE FOR SINGLE- AND MULTI-OUTPUT LGCPS

All VI results were obtained the python gaussian processes package *GPFlow v*2.1.3.

For all single output LGCP, we used a radial basis kernels (RBF) for the inputs variables. The VI model was a stochastic variational inference for a gaussian process (SVGP), for details see Hensman et al. (2013) and Hensman & Ghahramani (2015).

The SVGP variables were optimized using the tensorflow v2.3.1 Adam optimizer. The optimizer call was run for a maximum number of iterations of 20000, with every iteration applied to the entire data set.

The optimization made use of an early stop when the latest evidence lower bound (ELBO) did not improve from the ELBO as observed 10 iteration steps before. The early stop was invoked in all cases, though sometimes as late as past step 18000. The RBF kernel parameters obtained for each regression can be found in the table underneath (rounded to 2 decimal places):

regression	variance	lengthscale
$\log \lambda_t = f_t^\lambda \sim \mathrm{GP}(t, \log I_{t-1} + \log S_{t-1})$	7.63	3.13
$\log \gamma_t = f_t^\gamma \sim \mathrm{GP}(t, \log I_{t-1})$	8.55	2.74
$\log \delta_t = f_t^\delta \sim \mathrm{GP}(t, \log I_{t-1})$	4.35	2.80

The multi output LGCP is an extension of the SVGP implementation van der Wilk et al. (2020). Multi output models use latent variables through which correlation is modelled (Aglietti et al. (2019)). We made use of two latent gaussian processes and shared the kernel optimizer across all outputs. The optimizer itself was the non-stochastic L-BFGS-B optimizer (Liu & Nocedal (1989)) from the *scipy* package applied to the entire data set which was allowed to run for 4000 iterations. Kernel optimization results were:

regression	variance	lengthscale
$[\log \lambda_t, \log \gamma_t, \log \delta_t] = \mathrm{GP}(t, \log I_{t-1} + \log S_{t-1}, \log I_{t-1})$	21.34	8.89

F.2 MARKOV CHAIN MONTE CARLO

We performed Markov Chain Monte Carlo using a Metropolis algorithm with $n = 1,000$ batches. We repeated each simulation run $M = 100$ times.

G MODEL TESTING

G.1 COMPARATIVE METRICS ACROSS METHODS

We compute the mean absolute error ("MAE") and root mean squared error ("RMSE") by comparing the predicted count for every day to the realised count, in each compartment.

G.1.1 NEW INFECTIONS

Method	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
Markov Chain Monte Carlo	201.71	353.20
Laplace Approximation	86.28	183.50
Local Laplace	127.85	217.73
Variational Inference	102.21	175.18

G.1.2 NEW RECOVERIES

Method	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
Markov Chain Monte Carlo	248.70	522.68
Laplace Approximation	183.51	452.69
Local Laplace	183.51	452.69
Variational Inference	183.53	359.86

G.1.3 NEW DEATHS

Method	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
Markov Chain Monte Carlo	12.47	29.72
Laplace Approximation	10.70	28.00
Local Laplace	10.70	28.00
Variational Inference	11.17	27.98

G.2 OUT-OF-SAMPLE AND OUT-OF-TIME

To test the epidemic LGCP models, we performed an out-of-sample/out-of-time test on the single- and multi-output LGCPs obtained via VI, on the last 20 days (representing 10% of the dataset).

G.2.1 SINGLE-OUTPUT LGCP

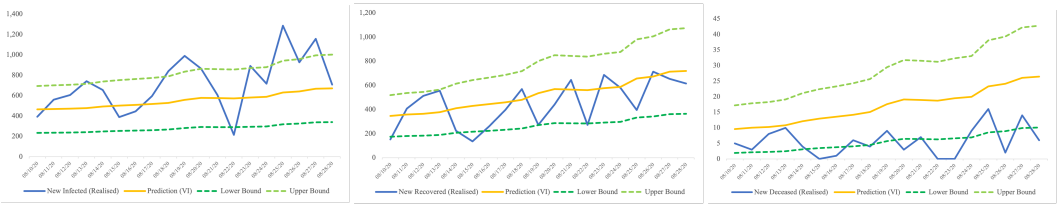

Figure 6: **(a)** Out-of-sample and out-of-time prediction of new infections versus realised new infections. **(b)** Out-of-sample and out-of-time prediction of new recoveries versus realised new recoveries. **(c)** Out-of-sample and out-of-time prediction of new deaths versus realised new deaths.

Method	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
New infections	226.33	281.74
New recoveries	143.39	167.29
New deaths	11.32	12.74

G.2.2 MULTI-OUTPUT LGCP

Method	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
New infections	407.34	498.00
New recoveries	243.89	279.97
New deaths	5.21	6.74