Gaussian Process Nowcasting: Application to COVID-19 Mortality Reporting

Iwona Hawryluk[1]   Henrique Hoeltgebaum[2]   Swapnil Mishra[1]   Xenia Miscouridou[2]   Ricardo P Schnekenberg[3]

Charles Whittaker[1]   Michaela Vollmer[1]   Seth Flaxman[2]   Samir Bhatt[1]   Thomas A. Mellan[1]

[1]Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, UK
[2]Department of Mathematics, Imperial College London, UK
[3] Nuffield Department of Clinical Neuroscience, University of Oxford, UK

Abstract

Updating observations of a signal due to the delays in the measurement process is a common problem in signal processing, with prominent examples in a wide range of fields. An important example of this problem is the nowcasting of COVID-19 mortality: given a stream of reported counts of daily deaths, can we correct for the delays in reporting to paint an accurate picture of the present, with uncertainty? Without this correction, raw data will often mislead by suggesting an improving situation. We present a flexible approach using a latent Gaussian process that is capable of describing the changing auto-correlation structure present in the reporting time-delay surface. This approach also yields robust estimates of uncertainty for the estimated nowcasted numbers of deaths. We test assumptions in model specification such as the choice of kernel or hyper priors, and evaluate model performance on a challenging real dataset from Brazil. Our experiments show that Gaussian process nowcasting performs favourably against both comparable methods, and against a small sample of expert human predictions. Our approach has substantial practical utility in disease modelling — by applying our approach to COVID-19 mortality data from Brazil, where reporting delays are large, we can make informative predictions on important epidemiological quantities such as the current effective reproduction number.

1   INTRODUCTION

In many real-world settings, current observations from a noisy signal can be systematically biased, with these biases only being corrected after subsequent updates create more complete data. Often, these updates occur much later in the future due to data processing or reporting delays. Not accounting for these delays would result in biased predictions, while waiting for updates would result in a lack of timely estimates. The need for timely estimates to predict the present is colloquially known as nowcasting, and its importance has been shown in a wide range of fields such as actuarial science, economics, and epidemiology [Kaminsky, 1987, Lawless, 1994, Bastos et al., 2019, McGough et al., 2020].

Nowcasting, as defined by Banbura et al. [2010] at the European Central Bank, is the process of predicting the present, the very recent past, and very near future using time series data known to be incomplete. An example from economics is using monthly data to nowcast the current state of important indicators for an economy such as GDP or income. More broadly, nowcasting is relevant for scenarios not only where the data are incomplete, but when the data are comprised of a biased subsample that will be updated in the future retrospectively, following lengthy delays.

In epidemiology, nowcasting is required due to delays in reporting arising from limitations in testing capacity, data curation, and the requirement for pseudonymisation of patient data [Bastos et al., 2019]. These delays are further compounded by the noise inherent in such data due to limited sampling (typically only a subset of the population is sampled). Throughout this paper, we specifically focus on the delays in the reporting of deaths. An individual dies of a disease on a given day, but the delay between this event and the death being reported (and appearing in the dataset) can be substantial because of the reasons noted above. These reporting delays mask the true *current* state of the epidemic, and have material consequences for our understanding of both present and future evolution of the epidemic. For example, estimation of key epidemiological quantities such as the effective reproduction number (R_t) would be systematically biased. Contemporary, real-time and unbiased estimates are necessary for effective public health planning and policy.
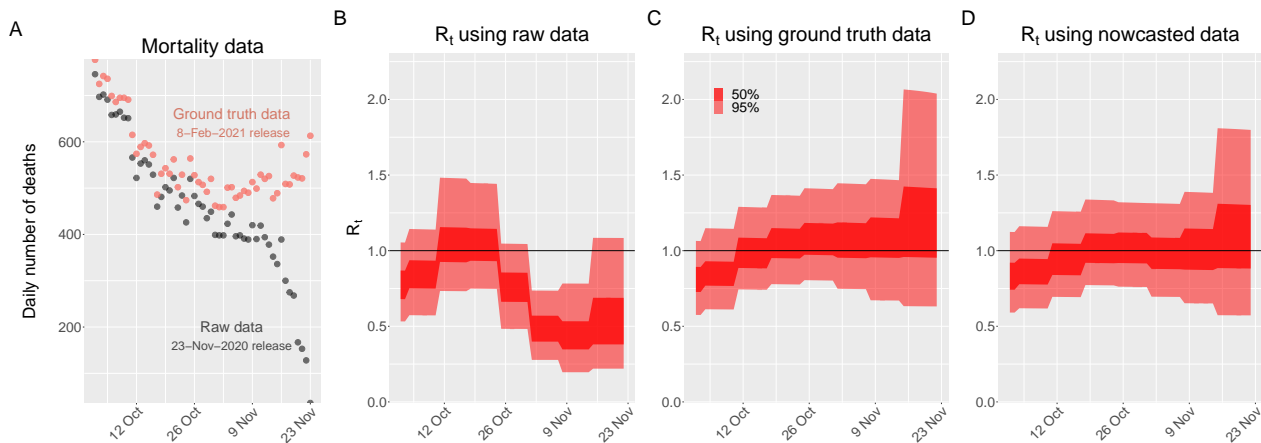
Figure 1: A) Reported daily hospital deaths are censored at recent times due to reporting delays. This can be seen by comparing the raw data with a ground truth from two months in the future, when the records have been backdated. B-C) The effective reproduction number R_t for SARS-CoV-2 infections in Brazil from 30-Jun-2020 to 23-Nov-2020, estimated using deaths from the raw reported data released on the 23-Nov-2020, and using a backdated ground truth based on data released on 08-Feb-2021. D) R_t estimates based on nowcasted mortality data. Whereas the raw data results in misleading estimates of R_t, with the estimated $R_t < 1$, by applying nowcasting to the deaths counts we achieve a picture of the epidemic closer to the truth.

In this paper we propose a nowcasting framework based on latent Gaussian processes (GPs). This methodology is used to address the specific problem of delayed reporting in the true incidence of deaths due to COVID-19 in Brazil.

1.1 RELATED METHODS

Previous methods for nowcasting exist in several different contexts. Bańbura and Modugno [2014] propose a maximum likelihood approach with a dynamic factor model to predict GDP. Shi et al. [2015] use a deep learning approach based on LSTM to nowcast rainfall intensity. Codeco et al. [2018] provide a framework to gather epidemiological information and correct for delays in reporting in Brazilian data. Bastos et al. [2019] present a Bayesian hierarchical model for nowcasting applied on data relating to dengue fever and severe acute respiratory infection cases. In McGough et al. [2020] a Bayesian nowcasting approach is proposed that produces accurate estimates that capture the time evolution of the epidemic curve. Specifically for COVID-19, Bayesian nowcasting approaches have been used to correct for the reporting delays in Bavaria and Sweden [Günther et al., 2020, Altmejd et al., 2020]. Further discussion around the challenges in estimating reporting delays are also addressed in Seaman and De Angelis [2020]. Finally the problem and background context in Brazil for delays in reporting with corrected data are further explained in Bastos et al. [2020], Villela [2020].

Our methods build upon and generalize the NobBS (Nowcasting by Bayesian Smoothing) method originally proposed by McGough et al. [2020]. NobBS is a Bayesian method

that produces smooth and accurate nowcasted estimates in the presence of multiple diseases. NobBS allows for both uncertainty in the delay distribution and the evolution of the epidemic curve. While an effective method, NobBS has several limitations, such as inability to pick up fast-occurring changes in the delay distribution, which we overcome in this paper. The extensions we show result in comparable performance for COVID-19 mortality surveillance in Brazil, but present a better fit to the dynamic delays distribution.

2 OUR CONTRIBUTIONS

The problem tackled in this paper is conceptually illustrated in Figure 1. The black points are the data available to us at a given time, and the red the ground truth that is only available much further in the future. It can be seen that the discrepancy between the presently available data and the underlying ground truth data grows markedly as we approach the present – a distinguishing characteristic of reporting delays. Alongside this, in Figure 1 we also show 3 estimates of the effective reproduction number R_t (defined as the average number of infections an infected individual will go on to infect), obtained using a Bayesian hierarchical renewal-type model [Flaxman et al., 2020, Mellan et al., 2020, Mishra et al., 2020]. Understanding this epidemiological quantity is vital – $R_t > 1$ results in epidemics growing, while $R_t < 1$ results in epidemics declining. Figure 1B shows estimates of R_t derived from the raw data, while Figures 1C and 1D show estimates of R_t derived from the ground truth data and our nowcasting approach respectively. These plots show that not correcting for delays can lead to a fundamentally

different picture of the current epidemic state. Delays in death reporting lead to an underestimation of the true number of deaths in the observed data - the result is a suggestion of a declining epidemic, despite the fact that the epidemic is actually growing.

In this paper we focus on the Brazilian death data from the publicly available hospitalisation database with deaths from both confirmed and suspected COVID-19 diagnostic status [Ministério da Saúde, 2020]. Our central premise is that using these daily death data alone results in policy decisions being made based on false statistics and trends [Villela, 2020]. To facilitate well informed policy making based on unreliable data streams we propose and implement a nowcasting method using latent Gaussian processes. These GPs are capable of capturing the complex correlation structure in delayed data and present an effective means to correct the reporting delays. We use this corrected death data to calculate the effective reproduction number R_t using raw retrospective observed data, nowcasted data and the ground truth updated dataset (Figure 1).

Our contributions are the following:

- We provide a new, flexible and accurate way to correct for delays in reporting. Our framework solves the nowcasting problem through using latent GPs, and provides realistic estimates for the deaths today given incomplete data. Our approach closely predicts the non observed/missing values and simultaneously learns the underlying (latent) data generating mechanisms of the delays.

- We compare our approach to an established alternative method (NobBS), and in a novel contribution, also provide a comparison to a small human expert panel of infectious disease epidemiologists. Domain knowledge is of primary importance for such applications, and is frequently the primary approach taken to interpret data. In generating estimates that are improved over both existing computational methods as well as human experts, we demonstrate the utility of our approach.

- An important contribution of this work are the results and estimates provided. Implementing our approach enables generating of more accurate estimates of the reproduction number; and in turn, a better understanding of the evolution of the COVID-19 epidemic in Brazil. Our framework is implemented in the easy to use probabilistic program PyStan, and therefore facilitates use in low and middle income settings with limited technical expertise.

The structure of the paper is as follows: In section 3 we briefly introduce Gaussian processes and describe the latent GP nowcasting models with several variants. In section 4 we describe the data and perform retrospective tests to evaluate the accuracy of the new models and compare them with a sample of human experts predictions. Finally, we discuss the advantages and limitations of the GP nowcasting framework in section 5.

3 GAUSSIAN PROCESS NOWCASTING

3.1 NOWCASTING

Let n_t denote the response variable of interest that needs to be nowcasted at time t. In this paper n_t represents the reported COVID-19 mortality in week t. The mortality observations, in general, consist of measurements from an online data source, subject to distributed observation delays. The central task of nowcasting approaches is to identify a regular time-delay structure, and use this to estimate n_t, at a time when it has only been partially observed. The structure that nowcasting identifies is the additive decomposition of the observable over the reporting delay d. That is, the true signal at a given time t is the sum over all the delayed partial observations for that time:

$$n_t = \sum_d n_{t,d}. \tag{1}$$

The intuition behind this formulation is that the "true" deaths that occurred at time t are distributed over various delays d due to the delays in reporting them.

A visual example of partial observation at recent times is the right-censored epidemiological data shown in Figure 2A. For all data releases, we observe precipitous declines in contemporary data, which is then subsequently revised upwards as the data becomes more complete. In the COVID-19 context this occurs due to time lags in registering and reporting death certificates [Villela, 2020]. Figure 2B shows that most deaths are reported to near completeness after around 5 weeks, and 90% are reported within 10 weeks. Splitting the data up by delay index we form a 2D array in time and delay, $n_{t,d}$. The filled-in part of $n_{t,d}$, called the *reporting triangle*, is shown in Figure 2C. The lower triangle part of this 2D array is missing, since at any time T only the number of deaths reported with delay $d \leq T - t$ are known for each epidemiological week t.

The representation of the data by time and delay, rather than time and reporting date, induces a regular structure — one that is auto-correlated and approximately monotonically decreasing in delay (Figure 2C). This relatively simple structure makes this problem amenable to statistical modelling. The lower triangle of the $n_{t,d}$ matrix can be predicted with the model, and therefore an estimate of the true signal is available for any time up to the current time by Eqn. 1, by summing over the delays. This is a common theme from which variations of nowcasting branch out.

To model the discrete positive values of $n_{t,d}$, we can use a Poisson or a negative-binomial likelihood for overdispersed data:

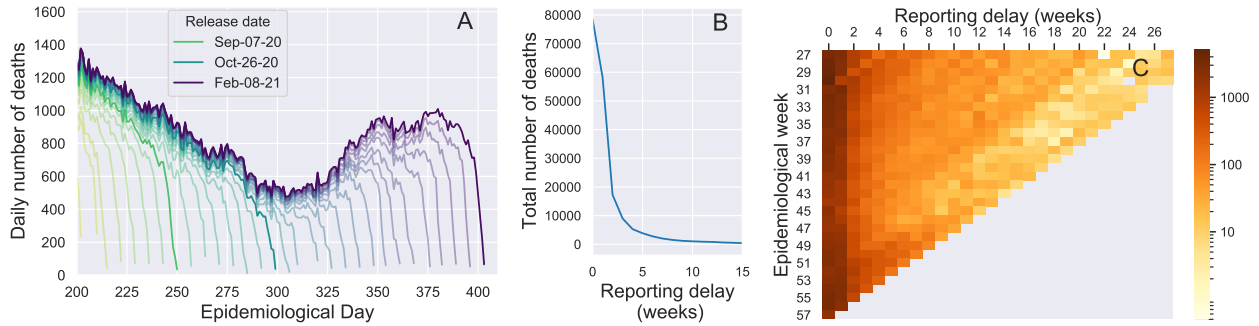$$n_{t,d} \sim \text{NB}(\lambda_{t,d}, r). \tag{2}$$

Figure 2: A) Daily COVID-19 deaths in Brazil as reported in releases of data between July-2020 and Feb-2021. Each line represents a single release. B) Total number of deaths reported per reporting delay in weeks. Most deaths are reported with delay ≤ 5 weeks. C) The reporting triangle, showing the number of COVID-19 deaths reported for each week with specific reporting delay. Mortality data were obtained from the SIVEP-Gripe hospitalisation database [Ministério da Saúde, 2020].

In the negative-binomial case, the dispersion parameter r is a hyperparameter that can be learnt or given an informative prior based on the problem. The latter approach is common among the established Bayesian nowcasting methods [Bastos et al., 2019, Günther et al., 2020, McGough et al., 2020]. The mean of the negative binomial, λ_{t_d}, is often modelled as a random walk [Bastos et al., 2019] or as an auto-regressive process [McGough et al., 2020] along the time dimension, that is joint independent with a learnt vector of delays. The second approach is taken in the NobBS model [McGough et al., 2020], used in this paper as a benchmark. Specifically, the NobBS model describes λ_{t_d} as:

$$\log(\lambda_{t,d}) = \alpha_t + \log(\beta_d) \tag{3}$$

where α_t is a latent signal for week t and β_d is the probability of reporting with delay d. It is worth noting, that with this approach, the distribution of delays β_d is fixed throughout the window of analysis.

This approach has been successful for dengue and influenza surveillance [Codeco et al., 2018, Bastos et al., 2019], but has limitations in terms of the generality of the time-delay covariance structure that can become apparent in more dynamic nowcasting scenarios, such as an evolving epidemic with changing delay distributions (Figure S1 in the Supplement). Such issues can be minimised by tuning the window over which the static delay vector is estimated, or by manually adding cross-term covariates. Here we employ Gaussian processes as a generic flexible alternative to model arbitrarily structured λ_{t_d}. The details of this are set out in the following section.

3.2 LATENT GP

The introductory model we consider consists of a latent GP with a 1D kernel. In general terms, GPs are a class of Bayesian non-parametric models that define a prior over functions. They are a powerful tool in machine learning,

for learning complex functions with applications in both regression and classification problems [Rasmussen and Williams, 2006, Wilson and Adams, 2013]. In recent years GPs have gained popularity in statistics and in machine learning, due to their flexibility and excellent performance for many spatial and spatiotemporal problems [Wilson and Adams, 2013, Flaxman et al., 2015], including COVID-19 modelling [Qian et al., 2020]. The covariance function or kernel together with the mean function completely define a GP. The mean function is the base function around which all of the realizations of the GP are distributed. The *covariance kernel* is a crucial component of the Gaussian process, as it describes the covariance of the Gaussian process random variables i.e. how similar two points are. Therefore, the kernel defines the shape of the distribution and which type of functions are more probable.

One of the most popular choices of covariance kernel, and the one we chose to introduce the model with, is the *squared exponential* kernel, k_{SE}, with entries defined by a covariance function $k_{\text{SE}}(\cdot, \cdot)$ such that

$$k_{\text{SE}}(t_i, t_j) = \alpha^2 \exp\left(-\frac{||t_i - t_j||_2^2}{2\rho^2}\right). \tag{4}$$

The parameter α defines the kernel's variance scale, and ρ is a lengthscale parameter that specifies how nearsighted the correlation between pairs of time points (t_i) is. The kernel results in a prior over a set of functions to describe, $\lambda_{t,d}$, the mean of the statistical model in Eqn. 2. This is modelled as a zero mean log-space latent Gaussian process

$$\log(\lambda_{t,d}) \sim \text{GP}(0, k_{\text{SE}}). \tag{5}$$

Due to weak identifiability [Rasmussen and Williams, 2006], a strategy to identify the hyperparameters ρ and α is to fix the lengthscale ρ to the maximum delay time considered in the nowcasting problem, and learn only the scale parameter α. Markov Chain Monte Carlo (MCMC) is used in order

to generate posterior summaries for arbitrary (non-normal) latent Gaussian processes.

3.3 GENERALISED MODEL

3.3.1 Additive Kernel Model

The basic model introduced above can be extended to provide a more expressive description of the data. The purpose of this is to be able to describe the complex structure in $n_{t,d}$.

Using the compositional kernel approach [Duvenaud et al., 2013, Wilson and Adams, 2013, Wilson et al., 2016], we can create a new additive kernel over multiple lengthscales, indexed s, as

$$k_{\text{add}} = \sum_s k_s \,,$$
$$\log(\lambda_{t,d}) \sim \text{GP}(0, k_{\text{add}}) \,. \tag{6}$$

The lengthscale hyperparameters are fixed or given strongly informative priors, ρ_s, while each α_s is learnt. In the simplest case we consider a kernel with two lengthscale contributions, short- and long-range correlation structure:

$$k_{\text{add}}(t_i, t_j) = k_{\text{long}}(t_i, t_j) + k_{\text{short}}(t_i, t_j) + \sigma^2 \delta_{ij} \,, \tag{7}$$

plus a regularising term with a Kronecker delta function ensuring σ^2 Gaussian noise is only added when $i = j$. The choice of kernel confers bias that can result in a better generalisation. The logic of this kernel is to split the covariance into two components: (a) a smooth long-range component, used to extrapolate the trend into the unknown part of the reporting triangle where large distances from the observed points exist and (b) a part for describing variation in $n_{t,d}$ over shorter lengthscales. Additionally, the separation of kernels provides a generic method to describe more complex data generating processes – for example, the long-range kernel can be squared-exponential, while the short-range can be a less smooth type with a different power spectrum such Matérn (1/2). This can be used to create a general statistical model for all of $n_{t,d}$. Furthermore, in this regard the δ contribution provides a source of regularisation which may be useful if there is reason to believe $n_{t,d}$ values are subject to variation beyond the scope of the basic nowcasting framework. For example, if a death can switch category from a COVID-19 suspected death to a cause other than COVID-19 in later data releases, this could result in a negative $n_{t,d}$ count, which can be modelled as an error to be regularised.

A further modification that can be applied if the time-delay surface $n_{t,d}$ has a complex structure, is to split the data into two components and model each with separate kernels. For example, if delays of 0 or 1 weeks account for a large fraction of total counts, they can be considered separately to delays > 1. This approach is considered later in section 4.2.

But a more generic formulation is to consider a 2D kernel to fully account for the time-delay correlation structure, which is introduced below.

3.3.2 2D Kernel Model

As a further expansion of the approach described before, we introduce a separable two dimensional kernel over time and delay, $k((t_i, t_j), (d_i, d_j)) = k_t(t_i, t_j)k_d(d_i, d_j)$. Separable kernels can be efficiently implemented using Kronecker product algebra as described in Flaxman et al. [2015]. Specifically, individual Gram matrices for time and delay are combined using the Kronecker product such that

$$K_{t,d} = K_t \otimes K_d \,. \tag{8}$$

As before, the kernel can be given an additive structure over multiple lengthscales. For example,

$$k_{\text{long}}(t, d) = k_{\text{long}}^t k_{\text{long}}^d \,,$$
$$k_{\text{short}}(t, d) = k_{\text{short}}^t k_{\text{short}}^d \,,$$
$$\log(\lambda_{t,d}) \sim \text{GP}(0, k_{\text{long}}(t, d) + k_{\text{short}}(t, d)) \,. \tag{9}$$

This approach captures the relationship between t and d. In both 1D and 2D kernel approaches it is possible to perform partial pooling of the models parameters by combining two or more spatial locations with similar features, for example neighbouring states, if limited data is available. In practice however we found that there was a limited gain in doing so as our approach works well with relatively few observations.

4 DATA AND MODEL PROPERTIES, FIT AND TESTING

4.1 DATA

The numbers of deaths per date have been extracted from the Brazilian Ministry of Health's Sistema de Informação de Vigilância Epidemiológica da Gripe (SIVEP-Gripe) database [Ministério da Saúde, 2020]. SIVEP-Gripe is a large publicly available database providing anonymised patient-level records of all individuals who died or were hospitalised with suspected or confirmed COVID-19 in Brazil [Bastos et al., 2020, de Souza et al., 2020, Niquini et al., 2020]. New data have been released regularly online, on a weekly basis, in the second half of 2020 considered here. In this study, we extracted all SIVEP-Gripe data releases from 7-July-2020 to 31-May-2021. We consider all cases of suspected or confirmed COVID-19 (class 4 and 5).

There are a number of potential sources of error in the reported SIVEP data. One is underascertainment — systematic biases which are beyond the scope of correction by this nowcasting methodology. Another source of error

is delayed classification. After the initial input of patient's data into the database (usually at the time of hospitalisation), the entry might be later updated with clinical and laboratory data, including confirmatory COVID-19 testing. Further updates will include the outcome and its date (i.e. date of death or date of hospital discharge) and cases receive a final classification. Cases can be classified as confirmed (class 5) or suspected COVID-19 (class 4), or other causes (classes 1-3). Despite being described as a "final classification", reclassification does occur, and is especially common for unknown cases to be reclassified as COVID-19 once results from confirmatory tests are informed to the health authorities. On the other hand, some deaths attributed to suspected SARS-CoV-2 infection are later "removed" from the SIVEP database, due to duplicate filtering or because they are eventually attributed to other diseases. That can cause the number of deaths on certain days to decrease in consecutive data releases, as shown in Figure S2 in the Supplement.

The number of deaths per day as reported by each release is presented in Figure 2, together with a reporting triangle, showing the distribution of the reporting delays across time. According to the SIVEP-Gripe dataset, over 90% of all deaths have been reported with delay less than 10 weeks (Figure 2B). We therefore choose the maximum reporting delay D for our data to be $D = 10$, and sum up all deaths which were reported with the delay longer than 10 weeks. Finally, to create the reporting triangle appropriate for our model, we aggregate the data into weeks.

4.2 MODEL FIT

We fit and present 7 models. For 1D kernel GPs, we consider a single SE kernel (1D SE), and additive long- and short-range component kernels (1D SE+SE and 1D SE+Mat). The additive long- and short-range component kernels are also considering splitting the data into across delay greater and less than one (1D SE+SE data-split). Finally we consider a 2D kernel GP model with additive long and short range components (2D additive). The NobBS model of McGough et al. [2020] is fitted and presented for reference of the current state-of-the-art. All models are fitted to the SIVEP-Gripe weekly COVID-19 deaths reported in Brazil, currently available until 31-May-2021.

Posterior samples of the parameters in the models were generated using Hamiltonian Monte Carlo with Stan [Hoffman and Gelman, 2014, Carpenter et al., 2017], using the PyStan interface (version 2.19.0.0). For each fit we used 4 chains and 1000 iterations, with 400 iterations dedicated to warm-up. The convergence of each model fit was evaluated by ensuring that $\hat{R} < 1.01$ for each parameter. Traceplots and other MCMC diagnostic measures were also investigated (Supplement, section 3).

Each of the models, characterised by the likelihood given in Eq (2), and a latent GP part for modelling the $\lambda_{t,d}$ (section 3.2) is trained by supplying the reporting triangle $n_{t,d}$ filled with data available up to the point of the nowcast. Each of the parameters governing the model, such as over-dispersion r or lengthscales and variances of the GPs are learnt during the model fit. The best performing hyperparameters of the prior distribution were selected conditioned on the observed results. All of those parameters and their prior densities are given in the Supplement, Table S1. The training of the model and nowcasting through sampling each element of the $n_{t,d}$ matrix is done simultaneously. Specifically, at each iteration parameter values are sampled and immediately used to sample from the negative-binomial distribution to obtain all elements of the $n_{t,d}$ matrix.

Other nowcasting methods, including NobBS, focus primarily on estimating only the "missing" part of the $n_{t,d}$ array and comparing the total numbers n_t, that is sums of each row of the array. Here, we aim to obtain a statistical model explaining all elements of the $n_{t,d}$ matrix. The reason for that is twofold: firstly, having a model that describes the whole $n_{t,d}$ surface well increases the reliability of the model, which is vital in any healthcare setting. Secondly, the SIVEP-Gripe database contains hard to identify errors discussed in section 4.1, therefore it is preferable to treat the reported data with additional statistical uncertainty. The fit of the 2D GP and the NobBS models to the $n_{t,d}$ matrix is presented in Figure 3 and shows that the GP-based nowcasting method fits the time-delay structure much closer than NobBS.

4.3 RETROSPECTIVE TESTING

To evaluate the accuracy of the competing nowcasting models, we fit all of the models to the retrospective data sets available at each week between 05-Oct and 30-Nov-2020, using the numbers of deaths recorded for the whole of Brazil. This way we obtain 63 different sets of nowcasts, which we compare to the numbers of deaths reported by the most recent SIVEP-Gripe data release from 31-May-2021. The start date gives us at least 15 weeks of training data for each of the nowcasts. The end date of 30-Nov is 26 weeks before the most recent release, so the number of deaths reported in the most recent release can be confidently taken as a true value. The comparison is done by calculating the weighted and unweighted rooted mean squared error (RMSE) and the continuous ranked probability score (CRPS) between the "true" values from the most recent release and the nowcasted values. The differences between the ground truth and raw data, as shown in Figure 1A, are used as weights. For each RMSE and CRPS evaluation, we use the mortality data from the 10 weeks leading up to the date of the nowcast.

Out of all tested models, GP models with 1D kernels with 2 components (SE+SE and SE+Mat) performed worse than the benchmark. The predictive accuracy of the other mod-
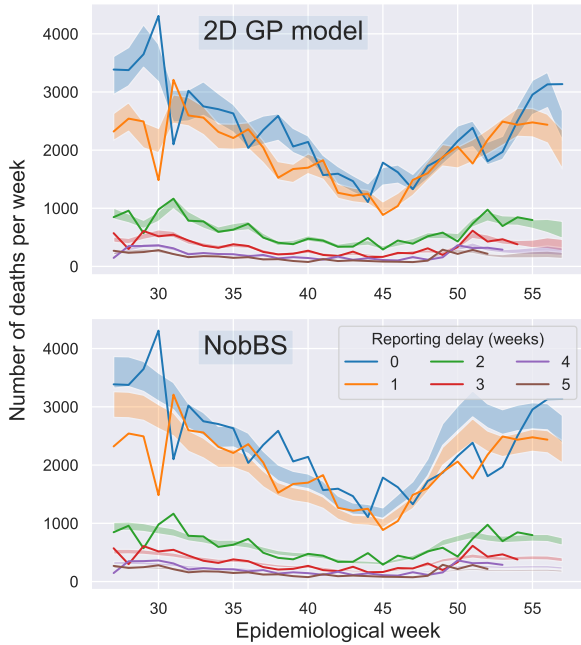
Figure 3: Reported and nowcasted numbers of deaths with reporting delay 0 to 5 weeks generated by the 2D additive GP and NobBS models. These plots show the columns of the reporting triangle $n_{t,d}$. The reported data are shown with solid lines, and the 50% CrI for the nowcasts with the ribbons. 95% CrI are shown in Figure S3.
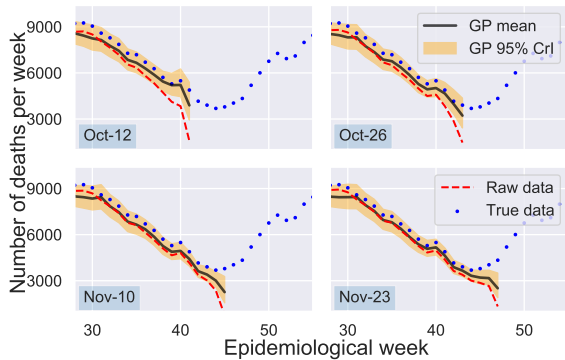

Figure 4: Retrospective testing for the whole of Brazil using the 2D additive GP model.

els was comparable to that of the benchmark, as shown in Figure 5, while also simultaneously giving an appropriate statistical description of the data (Figure 3). These results provide empirical evidence that our proposed method, under correct specification, gives a complete and accurate approach for describing and nowcasting COVID-19 death data. Model fits for the 2D additive GP model, including the 95% credible intervals (CrI) are shown in Figure 4 and the fits for the remaining models are shown in Figures S8-S14.

In addition to the forecasting metrics, as a novelty, we also

Figure 5: RMSE and CRPS evaluated for n_t for tested now-casting methods over the weeks 5-Oct to 30-Nov-2020. For the weighted RMSE, the weights were calculated as a difference between the ground truth and the raw data available up to the nowcasting date.

evaluate how the GP nowcasting model performs when compared to human experts' predictions. We asked a group of infectious disease epidemiology experts to provide a series of nowcasts when presented with time series data up to 12-Oct and 23-Nov-2020. They were asked for their estimates of the true numbers of deaths due to COVID-19 in Brazil on the 08-Oct and 19-Nov-2020 (Figure S15). The dates were specifically chosen to represent different scenarios. In the first one, 08-Oct-2020, both the raw data and the updated numbers of daily deaths were declining. Whereas in the second date, 19-Nov-2020, the raw data were declining while the updated release revealed that the true numbers of daily deaths were actually increasing. For this experiment, 36 anonymous experts provided their point estimates and confidence intervals, which are presented in Figure 6. To extract daily deaths from the model's weekly estimates, we performed a simple interpolation, through setting the nowcasted number of deaths per week divided by 7 to the middle day of the given week, and interpolating the remaining values using splines (example shown in Figure S16).
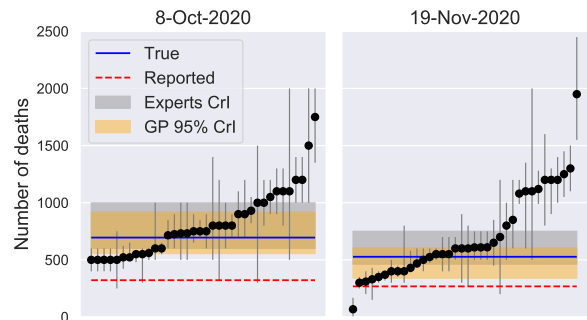

Figure 6: Human experts estimates of the true number of deaths are shown with the black points and errorbars.

Notably in both cases the median value guessed by the human experts was not far from the "true" value (difference of 55 and 73 deaths/day respectively for the human median estimate and 13 and 72 for the nowcasting model mean), however only 36% and 50% of all answers included the true value within the provided credible intervals, respectively. The confidence intervals given by the human experts were comparable with the 95% CrI of the model, with the model confidence narrower by 19 deaths/day for the first date and 27 deaths/day for the second date.

4.4 SENSITIVITY ANALYSIS

We performed basic sensitivity analysis for the 1D SE+SE data-split GP model. We first varied the prior for the over-dispersion parameter r. This parameter is often unidentifiable by the models and has to be chosen based on the data [McGough et al., 2020]. This is confirmed by our sensitivity analysis, where changing the prior for r changed the width of the confidence interval, but did not impact the mean predictions, as shown in Figures S19 and S20. Changing the priors for the scale parameter α to less informative, that is increasing the variance of the priors does not significantly affect the mean predictions (Figures S21, S22). Changing the mean of the priors does however have an impact on the predictions and leads to bi-modality of the posterior distribution if the model is misspecified, e.g. when $N(0, 1)$ priors are used (Figures S23, S24).

5 DISCUSSION

Applying nowcasting to surveillance data suffering from the reporting delays is crucial to accurate tracking of real-time epidemic dynamics. The limitations associated with using non-corrected data in epidemiological analyses is highlighted with our results of the R_t estimates shown in Figure 1. Use of this raw data leads to continued underestimation of R_t and predicts a declining epidemic. Specifically, in the month preceding the nowcast, the relative entropy value for the ground truth and raw data R_t was on average 13.14 (max 43.8) and for ground truth and nowcasted data R_t only 0.26 (max 0.35) (see Figure S4). By contrast, the ground truth results show that the epidemic remains uncontrolled, with R_t remaining above 1 - an important conclusion also captured by our nowcasting approach.

The recent emergence of SAR-CoV-2 variants of concern with altered epidemiological characteristics, such as increased transmissibility [Volz et al., 2021] or partial evasion of immunity [Faria et al., 2021], emphasise the need accurate and continued real-time epidemic tracking. To estimate COVID-19 mortality in Brazil, during a resurgent phase of the epidemic concurrent with the emergence of the P.1 variant, we perform nowcasting using data released on the 8-Feb-2021 and compare the predictions to the updated

numbers of deaths released on the 31-May-2021. The results of the nowcast for the whole of Brazil are shown in Figure 7. The reported data show a decline in the weekly number of deaths since epidemiological week 54 for Brazil, however the nowcasted results show much higher numbers of estimated weekly deaths, closer to the true value.
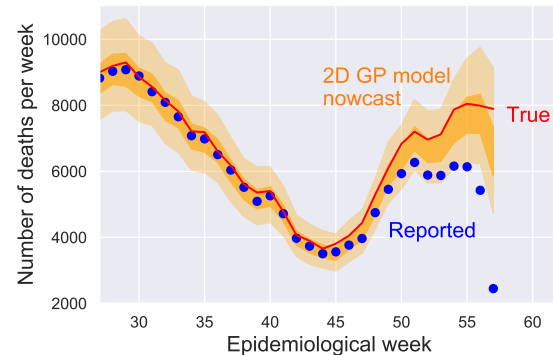

Figure 7: Nowcasted and reported deaths due to COVID-19 death for Brazil up to 8-Feb-2021 generated with the 2D additive GP model. 50% and 95% CrI for the GP model nowcasts are shown with the ribbon. True values were obtained from the SIVEP release on the 31-May-2021.

The GP nowcasting framework we introduced was benchmarked with the established NobBS method [McGough et al., 2020]. The main theoretical difference between the two proposed models is how the mean of the negative-binomial distribution, describing the number of deaths per week, is modelled. Our model uses a latent Gaussian process to do that, whereas NobBS uses first-order random walk. The 2D GP model improves on the RMSE for point predictions and CRPS for the distribution of samples compared to NobBS, as shown in Figure 5, and provides more realistic uncertainty intervals (Figures S13 and S14). As well as improving predictive performance on the missing parts of the reporting triangle, the GP framework provides a more expressive statistical model capable of better explaining the historical reporting data (see Figure 3).

One of the limitations of the approach described here is the dependence on the historical data and the regularity of the data releases, a limitation shared by many other nowcasting approaches. Additional challenges include variability in the distribution of reporting delays over time. For example, during the initial phase of the Brazilian COVID-19 epidemic, reporting delays were particularly severe. Delays in reporting are typically most extensive during outbreaks of a novel pathogen (such as SARS-CoV-2), due to the limitations in diagnostic availability and testing capacity. Relatedly, during epidemic peaks, strain on healthcare systems and administrative staff due to increasing admissions can also result in lengthening of reporting delays.

The GP nowcasting models introduced in this paper can be

readily used for real-time monitoring of the new outbreaks of diseases, as relatively few data points are required to train the model, ca. 3 months here. In other applications it is possible more data may be required, depending on the distribution of reporting delays, variance of counts, and regularity and granularity of data. Although this paper focuses on application of the proposed GP-based nowcasting framework to the death counts for the whole country, the GP models can be applied at finer spatial scales, as illustrated for individual states of Brazil in Figure S5. This flexibility is important due to the large heterogeneity of the healthcare system in the country [Baqui et al., 2020].

6 CONCLUSIONS

We have presented a new approach to modelling time-delay data, which can be used to nowcast online data streams that have statistically distributed delays. Our approach uses latent Gaussian processes with additive kernels, and gives a fully flexible and generic method to describe and predict the data for unknown delays. The method has been demonstrated for assessing mortality and estimating the effective reproduction number for COVID-19 reporting in Brazil, but can be used for other contexts in which delays in a measurement process exist.

7 CODE AND DATA AVAILABILITY

Python, R and Stan code used to analyse the data and fit the nowcasting models is available at `https://github.com/ihawryluk/GP_nowcasting`.
The SIVEP-Gripe database [Ministério da Saúde, 2020] is available to download from Brazil Ministry of Health website `https://opendatasus.saude.gov.br/dataset/bd-srag-2020`.

References

Adam Altmejd, Joacim Rocklöv, and Jonas Wallin. Nowcasting COVID-19 statistics reported with delay: a case-study of Sweden. *arXiv*, 2020. URL `https://arxiv.org/abs/2006.06840`.

Marta Bańbura and Michele Modugno. Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160, 2014. URL `https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1189.pdf`.

Marta Banbura, Domenico Giannone, and Lucrezia Reichlin. Nowcasting. *ECB Working Paper No. 1275*, 2010. URL `https://ssrn.com/abstract=1717887`.

Pedro Baqui, Ioana Bica, Valerio Marra, Ari Ercole, and Mihaela van der Schaar. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *The Lancet Global Health*, 8:e1018 − e1026, 2020. doi: https://doi.org/10.1016/S2214-109X(20)30285-0.

Leonardo S Bastos, Theodoros Economou, Marcelo F C Gomes, Daniel A M Villela, Flavio C Coelho, Oswaldo G Cruz, Oliver Stoner, Trevor Bailey, and Claudia T Codeço. A modelling approach for correcting reporting delays in disease surveillance data. *Statistics in Medicine*, 38(22): 4363–4377, 2019. doi: https://doi.org/10.1002/sim.8303.

Leonardo S Bastos, Roberta P Niquini, Raquel M Lana, Daniel AM Villela, Oswaldo G Cruz, Flávio C Coelho, Claudia T Codeço, and Marcelo FC Gomes. COVID-19 and hospitalizations for SARI in Brazil: a comparison up to the 12th epidemiological week of 2020. *Cadernos de Saúde Pública*, 36, 2020. URL http://www.scielo.

br/scielo.php?script=sci_arttext&pid=
S0102-311X2020000406001&nrm=iso.

Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017. doi: 10.18637/jss.v076.i01.

C. Codeco, F. Coelho, O. Cruz, S. Oliveira, T. Castro, and L. Bastos. Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil. *Revue d'Épidémiologie et de Santé Publique*, 66:S386, 2018. doi: https://doi.org/10.1016/j.respe.2018.05.408.

W.M. de Souza, L.F. Buss, D.d.S. Candido, et al. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nature Human Behavior*, 4:856–865, 2020. doi: https://doi.org/10.1038/s41562-020-0928-4.

David Duvenaud, James Lloyd, Roger Grosse, Joshua Tenenbaum, and Ghahramani Zoubin. Structure Discovery in Nonparametric Regression through Compositional Kernel Search. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1166–1174, Atlanta, Georgia, USA, 2013. PMLR. URL http://proceedings.mlr.press/v28/duvenaud13.html.

Nuno R. Faria, Thomas A. Mellan, Charles Whittaker, Ingra M. Claro, Darlan da S. Candido, Swapnil Mishra, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*, 372(6544):815–821, 2021. doi: 10.1126/science.abh2644.

Seth Flaxman, Andrew Gordon Wilson, Daniel Neill, Hannes Nickisch, and Alex Smola. Fast Kronecker Inference in Gaussian Processes with non-Gaussian Likelihoods. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 607–616, Lille, France, 07–09 Jul 2015. PMLR.

Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Helen Coupland, Thomas A Mellan, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, Imperial COVID19 Reponse Team, Azra Ghani, Christl A. Donnelly, Steven Riley, Lucy C Okell, Michaela A C Vollmer, Neil M. Ferguson, and Samir Bhatt. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584:257—261, 2020. doi: https://doi.org/10.1038/s41586-020-2405-7.

Felix Günther, Andreas Bender, Katharina Katz, Helmut Küchenhoff, and Michael Höhle. Nowcasting the COVID-19 pandemic in Bavaria. *Biometrical Journal*, 63(3):490–502, 2020. doi: https://doi.org/10.1002/bimj.202000112.

Matthew D Hoffman and Andrew Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014. URL https://www.jmlr.org/papers/volume15/hoffman14a/hoffman14a.pdf.

Kenneth S. Kaminsky. Prediction of IBNR claim counts by modelling the distribution of report lags. *Insurance: Mathematics and Economics*, 6(2):151 – 159, 1987. doi: https://doi.org/10.1016/0167-6687(87)90024-2.

J. F. Lawless. Adjustments for Reporting Delays and the Prediction of Occurred but Not Reported Events. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 22(1):15–31, 1994. URL http://www.jstor.org/stable/3315820.

Sarah F McGough, Michael A Johansson, Marc Lipsitch, and Nicolas A Menzies. Nowcasting by Bayesian Smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLoS computational biology*, 16(4): e1007735, 2020. doi: https://doi.org/10.1371/journal.pcbi.1007735.

Thomas A Mellan, Henrique H Hoeltgebaum, Swapnil Mishra, Charlie Whittaker, Ricardo P Schnekenberg, Axel Gandy, H Juliette T Unwin, Michaela AC Vollmer, Helen Coupland, Iwona Hawryluk, et al. Report 21: Estimating COVID-19 cases and reproduction number in Brazil. *medRxiv*, 2020. URL https://doi.org/10.25561/78872.

Ministério da Saúde. SRAG 2020 - Banco de Dados de Síndrome Respiratória Aguda Grave, 2020. URL https://opendatasus.saude.gov.br/dataset/bd-srag-2020.

Swapnil Mishra, Tresnia Berah, Thomas A Mellan, H Juliette T Unwin, Michaela A Vollmer, Kris V Parag, Axel Gandy, Seth Flaxman, and Samir Bhatt. On the derivation of the renewal equation from an age-dependent branching process: an epidemic modelling perspective. *arXiv*, 2020. URL https://arxiv.org/abs/2006.16487.

Roberta Pereira Niquini, Raquel Martins Lana, Antonio Guilherme Pacheco, Oswaldo G Cruz, Flávio Codeço Coelho, Luiz Max Carvalho, Daniel Antunes Maciel Villela, Marcelo Ferreira da Costa Gomes, and Leonardo Soares Bastos. Description and comparison of demographic characteristics and comorbidities in SARI from COVID-19, SARI from influenza, and the Brazilian general population. *Cadernos de Saúde Pública*, 36, 00 2020. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2020000705013&nrm=iso.

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. In *Advances in Neural Information Processing Systems*, volume 33, pages 10729–10740, 2020. URL https://proceedings.neurips.cc/paper/2020/file/79a3308b13cd31f096d8a4a34f96b66b-Paper.pdf.

C. E. Rasmussen and C. K .I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. URL www.GaussianProcess.org/gpml.

Shaun Seaman and Daniela De Angelis. Challenges in estimating the distribution of delay from COVID-19 death to report of death. 2020. URL https://www.mrc-bsu.cam.ac.uk/tackling-covid-19/nowcasting-and-forecasting-of-covid-19/challenges-in-estimating-the-distribution/-of-delay-from-covid-19-death-to-report/-of-death/.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. NIPS'15, page 802–810, 2015. URL https://proceedings.neurips.cc/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf.

Daniel Villela. How limitations in data of health surveillance impact decision making in the COVID-19 epidemic. 2020. doi: https://doi.org/10.1590/SciELOPreprints.1313.

Erik Volz, Swapnil Mishra, Meera Chand, Jeffrey C. Barrett, Robert Johnson, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, 593:266—269, 2021. doi: https://doi.org/10.1038/s41586-021-03470-x.

Andrew Gordon Wilson and Ryan Adams. Gaussian Process Kernels for Pattern Discovery and Extrapolation. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1067–1075, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL http://proceedings.mlr.press/v28/wilson13.html.

Andrew Gordon Wilson, Zhiting Hu, Ruslan Salakhutdinov, and Eric P. Xing. Deep Kernel Learning. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 370–378, Cadiz, Spain, 09–11 May 2016. PMLR. URL http://proceedings.mlr.press/v51/wilson16.htm.