


Bayesian imputation of COVID-19 positive test counts for nowcasting under reporting lag

Radka Jersakova¹  | James Lomax¹ | James Hetherington^{1,2} |
Brieuc Lehmann² | George Nicholson³ | Mark Briers¹ |
Chris Holmes^{1,3}

¹The Alan Turing Institute, London, UK

²University College London, London, UK

³University of Oxford, Oxford, UK

Correspondence

Radka Jersakova, The Alan Turing Institute, London, NW1 2DB, UK.

Email: rjersakova@turing.ac.uk

Abstract

Obtaining up to date information on the number of UK COVID-19 regional infections is hampered by the reporting lag in positive test results for people with COVID-19 symptoms. In the UK, for ‘Pillar 2’ swab tests for those showing symptoms, it can take up to five days for results to be collated. We make use of the stability of the under reporting process over time to motivate a statistical temporal model that infers the final total count given the partial count information as it arrives. We adopt a Bayesian approach that provides for subjective priors on parameters and a hierarchical structure for an underlying latent intensity process for the infection counts. This results in a smoothed time-series representation nowcasting the expected number of daily counts of positive tests with uncertainty bands that can be used to aid decision making. Inference is performed using sequential Monte Carlo.

KEYWORDS

Bayesian inference, COVID-19, nowcasting, reporting lag, sequential Monte Carlo, uncertainty quantification

Radka Jersakova and James Lomax contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* published by John Wiley & Sons Ltd on behalf of Royal Statistical Society.

1 | INTRODUCTION

In light of COVID-19, the UK government tracks the number of lab-confirmed positive tests over time, primarily as a measure of progress against epidemic control.¹ Since it takes time for test results to be reported to their local authority, and subsequently centralised, there is uncertainty on the most recent positive test counts. This uncertainty diminishes over time until all tests for a particular day are eventually reported, whereafter the count remains unchanged. The time taken until the reported counts converge to a final value, here referred to as *reporting lag*, is around four days. News reports and publicly available summaries of the positive tests² ignore the days for which the counts have not yet converged to a final value, and often report a moving average of the positive tests.

We propose here a model on the positive test count with reporting lag which enables ‘nowcasting’ (Banbura et al., 2010) of the true count with uncertainty; in other words, by correcting for the underestimate in live reported data, we are able to suitably estimate and impute the actual positive test count and extend the seven-day moving average to the present moment. We also demonstrate how to incorporate the model into a statistical alerting system which is triggered when there is high confidence the reported positive test counts are above a threshold value.³

Given the pace of the COVID epidemic, there are a number of concurrent works with similar features to our approach (Fronterre et al., 2020; Lee & Robertson, 2020; Matt Keeling and the Warwick COVID modelling team, 2020; PHE Joint Modelling Cell and PHE COVID Outbreak Surveillance Team, 2020). These use either binomial or negative binomial models for test counts combined with spatio-temporal models (an approach widely used in epidemiology for modelling disease risk and spread). In contrast to our model, however, they do not consider reporting lag, and only analyse data once all the results are in.

We demonstrate that the reporting lag is in fact predictable, and include it in our model to return a nowcast that incorporates the most recently available reported data. We model the reporting lag using binomial thinning; whilst there already exist well-founded mechanisms for building auto-regressive binomial thinning models on count data (Jung & Tremayne, 2006), we choose instead to estimate distributions on the thinning rates directly from empirical data to avoid restricting the dynamics of the lag to a particular form or order of auto-regressive prior. With this approach we gain the additional benefit of finite-time saturation to a fixed value, which is a property observed in all sequences of lagged reports. We combine empirical priors on the under-reporting and a generative model in a time series prior, providing a posterior distribution on the underlying intensity of the disease. In Figure 1 we show the posterior distribution on the time series of counts for Leeds in December alongside the true and reported counts as an example of nowcasting with uncertainty quantification.

Other works that model a lagged reporting process for count data applied to COVID deaths (Seaman et al., 2020) and cases of dengue fever (Stoner & Economou, 2020) introduce a smooth temporal structure with cubic splines and encode week-day and calendar effects into the reporting lag. Schneble et al. (2021) also choose to apply temporal smoothing by including penalized

¹Tests here refer to PCR swab tests which detect presence of the virus. We analyse number of tests carried out through the National Health Service (NHS) and Public Health England (PHE) as well as commercial partners (pillars 1 and 2).

²Summary statistics and visualisations of the latest available data are presently available as a dashboard at <https://coronavirus.data.gov.uk>.

³The model code and scripts for reproducing all of our results can be accessed here: <https://github.com/alan-turing-institute/jbc-turing-rss-nowcasting>.

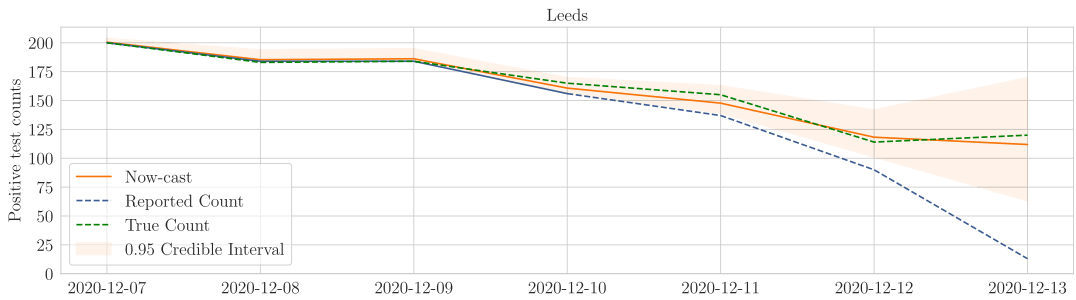


FIGURE 1 Example nowcast with uncertainty for Leeds in December 2020. Reported counts (blue line) up to the 9th of December correspond to the final true count (green line) but for more recent dates the reported count is incomplete. The nowcast (orange line) uses the incomplete reports to estimate what the true count will be once all tests are processed. The nowcast uncertainty is greatest for most recent dates and decreases with time. All uncertainty is eliminated once all tests have been reported

splines in their regression for a negative binomial model on the mortality count, as well as using age and gender as covariates in their nowcast of deaths in regions of Germany. McGough et al. (2020) encode temporal smoothness in modelling dengue fever and influenza-like illnesses with a geometric random walk prior, and introduce a Poisson model on case count which is relaxed to a negative binomial to control for under-dispersion. They introduce reporting delay through a weakly-informative Dirichlet prior on the reporting probabilities across a finite window between zero and some maximum possible delay. They choose further to infer the scale of the random walk under a diffuse gamma prior and note contention between learning stable random walk scales and under-dispersed delay distributions in the posterior. Hawryluk et al. (2021) replace the geometric random walk in McGough et al. (2020) with a Gaussian process where the length scale is controlled by the maximum reporting delay. This allows changes in the auto-correlation structure of the under-reporting to be learnt as time progresses, leading to improved nowcast estimates. Both McGough et al. (2020) and Hawryluk et al. (2021) use the expected posterior rate as their nowcast. The approach taken by Günther et al. (2021) also uses a negative binomial emission model, but models the delay between disease onset and case report as a Weibull distribution with shape and scale parameterized by a generalized additive model containing weekday and age effects. These effects are smoothed temporally with cubic splines, with the sequence of negative binomial rates also subject to a random walk prior.

These works have demonstrated the value of nowcasting in a number of contexts. Our main contribution is therefore the construction of a model which is tailored to perform well in nowcasting of positive tests for UK Lower-Tier Local Authorities (LTLAs). The key structural differences which have enabled this compared to the other works discussed are as follows: We choose to estimate the under-reporting process empirically, and sequentially, rather than learn it directly under the model we build, which lends robustness to changes in the reporting process. We also assume a smooth temporal random walk over the first-order differences of a latent Poisson process which models the positive test counts for each reporting day, rather than smoothing the rate directly. Instead of tackling potential under-dispersion by broadening a Poisson likelihood to a negative binomial, we note that uncertainty on the Poisson rate induced by the random walk prior will permit an infinite Poisson mixture, and let this implicit mixture handle dispersion control. Furthermore, we pose the unobserved final counts explicitly as latent random variables to be inferred under our model, meaning that we can use the marginal posterior for these final counts

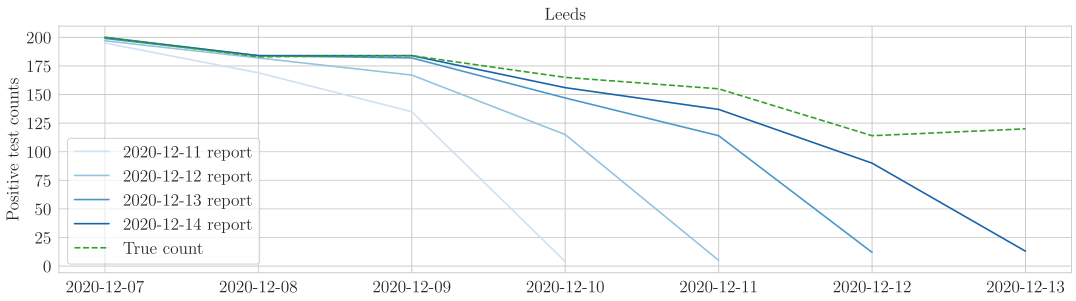


FIGURE 2 Four consecutive reports of positive test counts for Leeds in December. The report on 14th December updates counts reported on 13th for test dates from 10th December onwards, before which they each agree with the true count. Both underestimate the true count for the most recent dates, and will therefore continue to be updated in subsequent reports until the true count is reached

to construct estimates of the current number of positive tests. As reporting proceeds, this marginal posterior converges to the true count, eliminating any uncertainty in the nowcast once all tests are reported. The combination of these components leaves the random walk scale as the single free parameter in the model, which we select in an empirical Bayes manner, by optimising with respect to the model evidence.

2 | DATA

The collection of daily lab-confirmed COVID-19 positive test counts are available as open data (see Footnote 2). The data are stratified, so that the count for a nation is equal to the sum of counts for its regions. These regions are themselves divided into Upper-Tier Local Authorities (UTLAs), and each of these UTLAs is covered by a set of LTLAs, with the highest resolution count data available at the LTLA-level. In England, there are 9 regions, 150 UTLAs, and 316 LTLAs.

On each day, every LTLA reports a sequence of positive test counts for all test dates up to but not including the current day, allowing for updates to previously reported values (Figure 2). The most recently reported counts consistently underestimate the true count due to the lag in reporting. As time progresses and more tests are processed, the reported value for a given test date approaches the true count with increasing certainty (Figure 2). As a result, for each test date we observe a sequence of monotone increasing reports which eventually converges to the correct count for that particular date.⁴

3 | NOTATION

Let $i \in \{1, \dots, 316\}$ index the collection of LTLAs. Let $t \in \{0, \dots, T\}$ index the number of days for which data is available so that x_{it} is an unobserved random variable corresponding to the true count for LTLA i on day t . Let $j \in \{1, \dots, T - t\}$ index the reporting lag so that $y_{it}^{(j)}$ denotes the report for day t on day $t + j$. Each true count x_{it} is associated with a sequence of observed, reported

⁴In rare cases an error can lead to an over-estimate report followed by a lower count update.

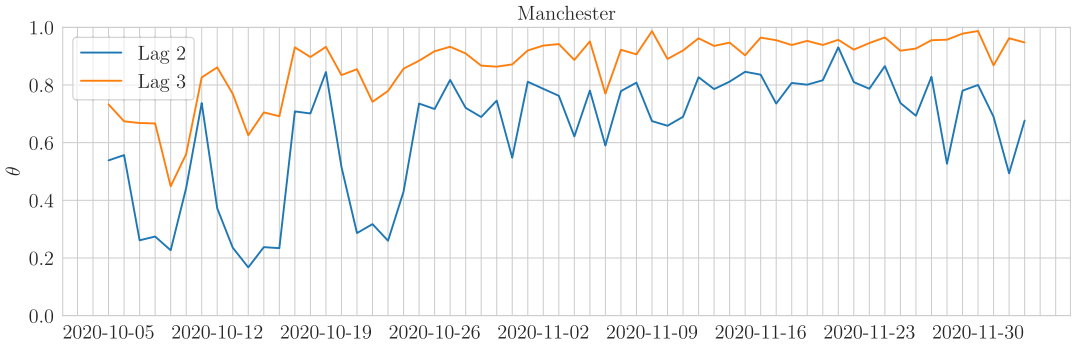


FIGURE 3 Time series of the true reporting rate θ at lag 2 and 3 for Manchester in October and November. The reporting rate is not temporally stable in October but becomes more predictable in November

counts $\mathbf{y}_{it} = (y_{it}^{(1)}, \dots, y_{it}^{(T-t)})$. For some finite but unknown maximum reporting lag τ_{it} , we observe $y_{it}^{(j)} = x_{it}$ for $j > \tau_{it}$.

Our aim was to specify a model on the true counts x_{it} , given reported counts $y_{it}^{(j)}$ and without knowledge of τ_{it} , in order to infer a distribution on the x_{it} which concentrates on the true value as T increases. We further define the reporting rate at lag j , $\theta_{it}^{(j)} := \frac{y_{it}^{(j)}}{x_{it}}$, to be the proportion of the true count that is reported at lag j . For historical data such that $y_{it}^{(T-t)}$ has converged to x_{it} , we can study $\theta_{it}^{(j)}$ in order to characterise the behaviour of the reporting lag.

4 | REPORTING LAG

We develop our nowcasting model on the basis of the empirical behaviour of the reporting rates $\theta_{it}^{(j)}$, which we now describe. Since July 2020, we have observed two distinct modes of predictable reporting behaviour which we refer to as spatially and temporally stable.

4.1 | Temporally stable reporting

If the reporting rates for an LTLA at each lag do not change much over some time window (see Figure 3) we say that they are temporally stable across such a window. When this reporting behaviour occurs, we may construct a temporally local set $S_{it}^{(j)} := \{t' : t - W \leq t' \leq t - \tau_{it'}\}$ where W is the length of the stable interval⁵ so that $\theta_{it}^{(j)}$ is estimated by

$$\hat{\theta}_{it}^{(j)} = \frac{1}{W - \tau_{it}} \sum_{k \in S_{it}^{(j)}} \theta_{ik}^{(j)}. \quad (1)$$

Figure 4 shows empirical distributions for $\theta_{it}^{(j)}$ marginally across the day-index t for lags $j = \{1, \dots, 5\}$. We observe that $\mathbb{E}[\theta^{(j)}]$ is increasing in j .

⁵The reporting rates can only be learned from data that has already converged to the true count, restricting the interval to dates past the maximum reporting lag τ_{it} .

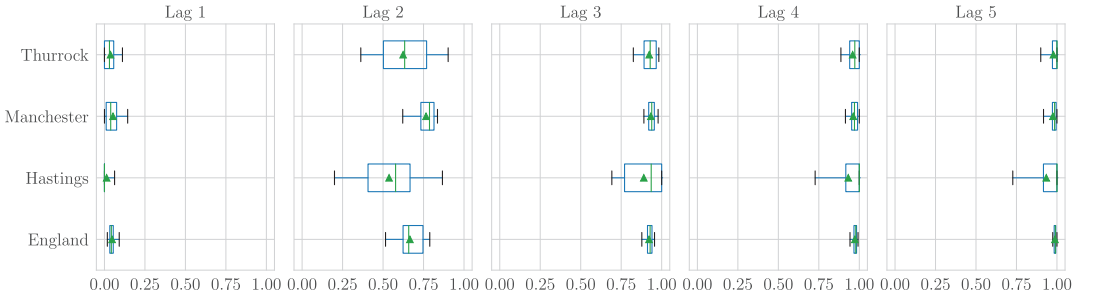


FIGURE 4 Empirical distributions of the reporting rates $\theta_i^{(j)}$ in November for a selection of Lower-Tier Local Authorities. We also include England to show that the marginal $\theta^{(j)}$ obeys the intuition encoded by observations in Section 4; the mean of $\theta_i^{(j)}$ increases with increasing lag, indicating that more reports are accounted for as time progresses

The reporting rates are predictable in a fashion supported by our intuition; as lag increases we are increasingly confident that $\theta_{it}^{(j)} = 1$ across all LTLAs, and this state is reached by a coherent sequence of updates to an initial underestimate. It is also clear from Figure 4 that there is enough variation in $\theta_i^{(j)}$ between LTLAs to warrant their separate consideration for modelling.

4.2 | Spatially stable reporting

Let $N_n(i)$ be the n -hop neighbourhood of LTLA i on the adjacency graph of LTLAs.⁶ When reporting is spatially stable we observe that the reporting rates of LTLAs which are close to one-another are similar and so we may estimate a reporting rate for an LTLA from those of its neighbours by

$$\hat{\theta}_{it}^{(j)} = \frac{1}{|N_n(i)|} \sum_{k \in N_n(i)} \theta_{kt}^{(j)}. \quad (2)$$

In the left panel of Figure 5 we show the performance of a 2-hop neighbourhood estimator for Manchester in October; the reporting rates of neighbouring LTLAs track one another. It is clear though that the reporting is not temporally stable, and so we must rely on spatial estimates alone. In the right panel we measure the performance of the 2-hop estimates (2) against the truth for all LTLAs where we have observed at least one report $x_{it} > 100$ marginally across the dates between the 14th and 30th October. There is a clear linear relationship ($R = 0.9$) with the truth.

4.3 | Empirical Bayes prior specification

In each of Sections 4.1 and 4.2, we demonstrate that noisy estimates of the reporting rates $\theta_{it}^{(j)}$ can be constructed. From here onwards we proceed only with temporal estimates; in a number of time periods throughout the analysis limited spatial correlations were observed. In order to

⁶The 1-hop neighbours of LTLA i are its immediate neighbours, 2-hop neighbours are the neighbours of those neighbours.

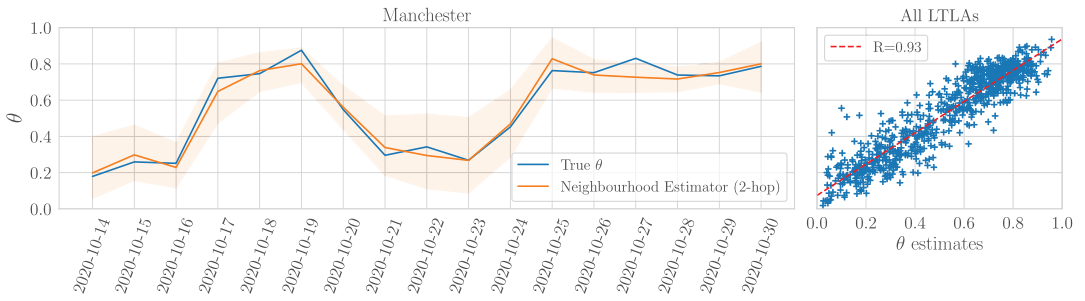


FIGURE 5 Lag-2 spatial neighbourhood priors. *Left*: True reporting rate $\theta_{it}^{(2)}$ for Manchester, and an estimator built from its 2-hop LTLA neighbours for each day from the 14th to the 30th of October and *Right*: 2-hop LTLA neighbour estimator against the true $\theta_{it}^{(2)}$ across all LTLAs with at least one reported count $x_{it} > 100$, combined over all days between the 14th to the 30th of October

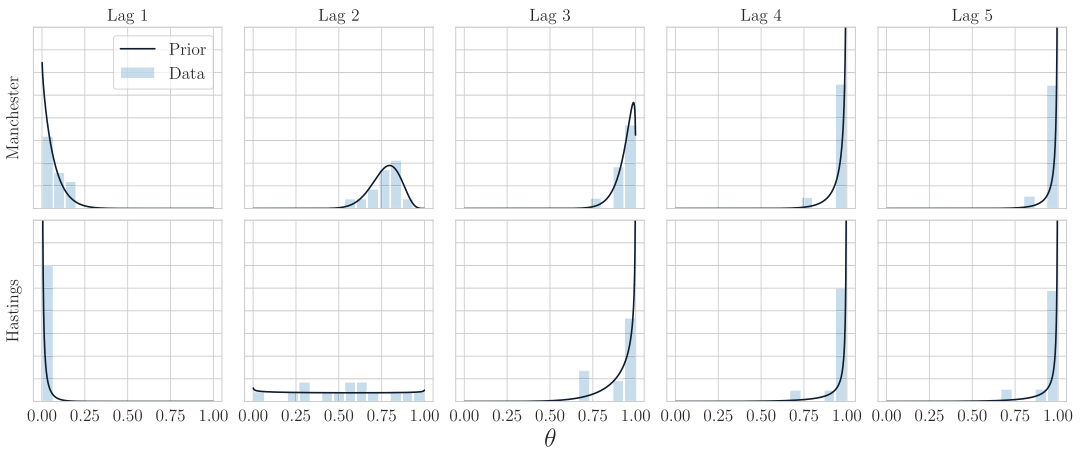


FIGURE 6 Empirical distributions and moment matched Beta priors $\theta_i^{(j)}$ for Manchester and Hastings in November

avoid underdispersion in models on x_{it} we must capture uncertainty in these estimates. We therefore propose that $\theta_{it}^{(j)} \sim \text{Beta}(\alpha_i^{(j)}, \beta_i^{(j)})$ and estimate $\alpha_i^{(j)}$ and $\beta_i^{(j)}$ by moment matching with the empirical distribution measured on the most recent 2 weeks of converged reports.⁷ Denote by $m_{ij} = \mathbb{E}[\theta_i^{(j)}]$ and $v_{ij} = \text{Var}[\theta_i^{(j)}]$ the means and variances of these empirical distributions and let $v_{ij} = \min\{v_{ij}, m_{ij}(1 - m_{ij}) - \epsilon\}$ with $\epsilon > 0$ but arbitrarily small, then

$$\alpha_i^{(j)} = \frac{m_{ij}^2 (1 - m_{ij})}{v_{ij}} - m_{ij} \quad \beta_i^{(j)} = \alpha_i^{(j)} \left(\frac{1 - m_{ij}}{m_{ij}} \right) \quad (3)$$

are chosen. Figure 6 illustrates the empirical distribution and resulting fit for two LTLAs in November - as time progresses, the moment matching produces Beta distributions with most of their mass on large values of $\theta_{it}^{(j)}$ as expected.

⁷Figure E1 in Appendix E illustrates that within a temporally stable interval, the prior is fairly insensitive to the number of data points used in its estimation. This is not the case during temporally unstable reporting.

5 | MODELLING

Based on the empirical observations made in Section 4, we now describe a model on reports $y_{it}^{(j)}$ which accounts for a predictable reporting lag in order to impute up-to-date predictive distributions on the current true count x_{it} . We model the reported counts $y_{it}^{(j)}$ as a random variable following a binomial distribution conditional on the (unobserved) true count x_{it} and the reporting rate $\theta_{it}^{(j)}$, placing a beta prior on $\theta_{it}^{(j)}$ with parameters $\alpha_i^{(j)}, \beta_i^{(j)}$ determined by (3). Modern research on epidemic monitoring typically poses negative binomial models on count data, with spatio-temporal regularization for smoothness; here we capture the same effect by proposing that the x_{it} result from a Poisson process and integrating over the rate parameter under a Gaussian random-walk time-series prior. We introduce the model in a modular fashion.

5.1 | Binomial thinning

To describe the relationship between the true counts x_{it} and the reported counts y_{it} , we treat each $y_{it}^{(j)}$ as a binomial random variable with success probability $\theta_{it}^{(j)}$ over x_{it} independent Bernoulli trials. Following the arguments of Section 4, we place beta priors on the $\theta_{it}^{(j)}$ with moment-matched $(\alpha_i^{(j)}, \beta_i^{(j)})$. Noting that the most recent available report $y_{it}^{(T-t)}$ is a sufficient statistic for x_{it} (Appendix C) we specify an observation model (5) in which only $y_{it}^{(T-t)}$ yields information on $(\theta_{it}^{(T-t)}, x_{it})$, via a binomial likelihood.⁸ This leads to the following hierarchical model:

$$\theta_{it}^{(T-t)} \sim \text{Beta} \left(\alpha_i^{(T-t)}, \beta_i^{(T-t)} \right) \quad (4)$$

$$y_{it}^{(T-t)} | \theta_{it}^{(T-t)}, x_{it} \sim \text{Binomial} \left(\theta_{it}^{(T-t)}, x_{it} \right). \quad (5)$$

Integrating out each $\theta_{it}^{(T-t)}$ yields the following joint distribution on $(x_{it}, y_{it}^{(T-t)})$:

$$p(x_{it}, y_{it}^{(T-t)}) = p(x_{it}) p(y_{it}^{(T-t)} | x_{it}) \quad (6)$$

$$= p(x_{it}) \binom{x_{it}}{y_{it}^{(T-t)}} \frac{\text{B} \left(y_{it}^{(T-t)} + \alpha_i^{(T-t)}, x_{it} - y_{it}^{(T-t)} + \beta_i^{(T-t)} \right)}{\text{B} \left(\alpha_i^{(T-t)}, \beta_i^{(T-t)} \right)} \quad (7)$$

where $p(x_{it})$ is a prior on x_{it} which may, for example, be flat across a set of feasible positive test counts given the population of the LTLA under consideration. This joint distribution may be used to draw samples from the marginal posterior $p(x_{it} | \mathbf{y}_{it})$ with Metropolis Hastings (MH).

5.2 | Latent intensity process

While estimating the true count directly is clearly important in monitoring the epidemic, a more compelling epidemiological variable is the latent rate which controls emission of these counts. To

⁸Note that all of day t 's earlier reports $y_{it}^{(T-1-t)}, y_{it}^{(T-2-t)}$, etc. are ignored for inference on x_{it} , because the sufficiency argument implies they add no extra information on x_{it} in addition to what $y_{it}^{(T-t)}$ provides.

extend the model, we therefore assume that each x_{it} is the result of a Poisson process with rate λ_{it} so that the hierarchy is now given by

$$\lambda_{it} \sim \text{Gamma}(a_{it}, b_{it}) \quad (8)$$

$$\theta_{it}^{(T-t)} \sim \text{Beta}\left(\alpha_i^{(T-t)}, \beta_i^{(T-t)}\right) \quad (9)$$

$$x_{it} \mid \lambda_{it} \sim \text{Poisson}(\lambda_{it}) \quad (10)$$

$$y_{it}^{(T-t)} \mid \theta_{it}^{(T-t)}, x_{it} \sim \text{Binomial}\left(\theta_{it}^{(T-t)}, x_{it}\right). \quad (11)$$

When λ_{it} is integrated out under a gamma prior distribution this is equivalent to proposing a negative-binomial distribution on x_{it} , which is a common choice in modelling count data for epidemic monitoring. The joint distribution on $(\lambda_{it}, y_{it}^{(T-t)})$ is given by:

$$p(\lambda_{it}, y_{it}^{(T-t)}) = p(\lambda_{it}) \sum_{x_{it}} \frac{\lambda_{it}^{x_{it}} e^{-\lambda_{it}}}{x_{it}!} \binom{x_{it}}{y_{it}^{(T-t)}} \frac{\text{B}\left(y_{it}^{(T-t)} + \alpha_i^{(T-t)}, x_{it} - y_{it}^{(T-t)} + \beta_i^{(T-t)}\right)}{\text{B}\left(\alpha_i^{(T-t)}, \beta_i^{(T-t)}\right)} \quad (12)$$

which may be used in MH to generate samples from the new posterior of interest $p(\lambda_{it} \mid y_{it}^{(T-t)})$.

5.3 | Temporal smoothing

Estimates of the latent rates λ_{it} may suffer from high variance, particularly in the early stages of reporting when the reported $y_{it}^{(T-t)}$ likely constitute underestimates of x_{it} . To reduce this variance, we encode time dependence in the latent rates, as follows. Let κ_{it} be the difference between Poisson rates so that

$$\lambda_{it} = \kappa_{it} + \lambda_{i,t-1}. \quad (13)$$

Further impose an AR1 prior with scale σ_i on the sequence $\kappa_{i,0:T}$ so that given a standard normal random variable $\epsilon_t \sim \mathcal{N}(0, 1)$ we may write

$$\kappa_{it} = \kappa_{i,t-1} + \sigma_i \epsilon_t \iff p(\kappa_{it} \mid \kappa_{i,t-1}) = \mathcal{N}(\kappa_{i,t-1}, \sigma_i^2). \quad (14)$$

This represents an intuitive assumption of local temporal smoothness in the epidemic rate and is a common choice of temporal regularisation; the Kalman Filter (Kalman, 1960), to which our model bears close resemblance, operates under a similar latent transition distribution. The prior induces dependence between each $\lambda_{it}, \kappa_{it}$ and the observed data $\mathbf{y}_{i,0:T} = \mathbf{y}_{i,0}, \dots, \mathbf{y}_{i,T}$. The key distributions of interest are then the joint *filtering* and *smoothing* distributions given by $p(\lambda_{it}, \kappa_{it} \mid \mathbf{y}_{i,0:t})$ and $p(\lambda_{it}, \kappa_{it} \mid \mathbf{y}_{i,0:T})$ respectively.

5.4 | Weekend effects

The smoothness assumption described in Section 5.3 constrains the sequence $\lambda_{i,0:T}$ to vary in proportion to the random walk scale σ_i . The weekend effects demonstrated in Figure 7 break

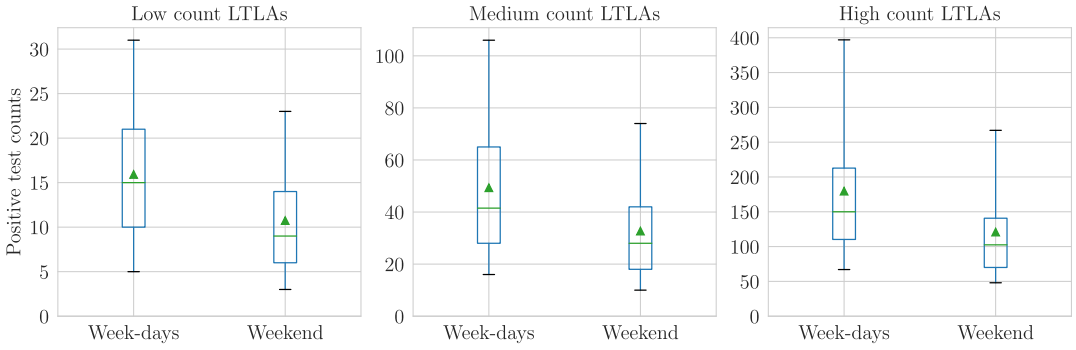


FIGURE 7 Daily positive test counts reported for week-day and weekend test dates in November across all Lower-Tier Local Authorities (LTLAs). The LTLAs are classified based on whether the mean number of tests is <20 (low count LTLAs), <100 (medium count LTLAs) or ≥ 100 (high count LTLAs). We observe some reduction in tests on weekends relative to week-days across all LTLAs

this smoothness assumption; when there are predictable drops in the test count at weekends it is because fewer tests are taken, rather than any true decrease in the underlying incidence rate of the disease.⁹ To capture this effect, we introduce latent random variables $z_{i,t} \in [0, 1]$ with prior distribution

$$p(z_{i,t}) = \begin{cases} \delta(z_{i,t} - 1) & \text{if } \text{day}(t) \in \{\text{Monday}, \dots, \text{Friday}\} \\ \text{Beta}(a, b) & \text{if } \text{day}(t) \in \{\text{Saturday}, \text{Sunday}\} \end{cases} \quad (15)$$

then let the emission distribution on x_{it} be

$$x_{it} \mid \lambda_{it}, z_{it} \sim \text{Poisson}(z_{it} \lambda_{it}) \quad (16)$$

so that smoothness in $\lambda_{i,0:T}$ is maintained by allowing z_{it} to capture these observed weekend effects by emitting counts at a reduced rate. In practice a flat $\text{Beta}(1, 1)$ prior allows the smoothness assumption to be maintained, though selecting a stronger a and b may be possible if weekend effects are predictable in their magnitude. We can measure the strength of these effects by examining the posterior smoothing distributions $p(z_{it} \mid \mathbf{y}_{0:T})$ on weekend days. In Appendix A.2 we give details on how to evaluate this posterior in the case where each weekend day has its own unique latent effect, but share prior parameters, as well as demonstrating how this addition alters the procedure for determining $p(x_{it} \mid \mathbf{y}_{i,0:T})$. From now on we omit the LTLA index i for brevity.

5.5 | The complete model

Together Sections 5.1–5.4 specify a smooth time-dependent model for count data exhibiting weekend effects and a lagged reporting process. The Directed Acyclic Graph (DAG) in Figure 8

⁹There is some variability in tests day to day as well as on the weekend. However, these effects are more subtle and less predictable, whilst the dip in testing on weekends is clear and consistent. We therefore allow the model to account for this persistent weekend effect but avoid including the additional parameters required to handle day-of-week effects explicitly, which are small by comparison.

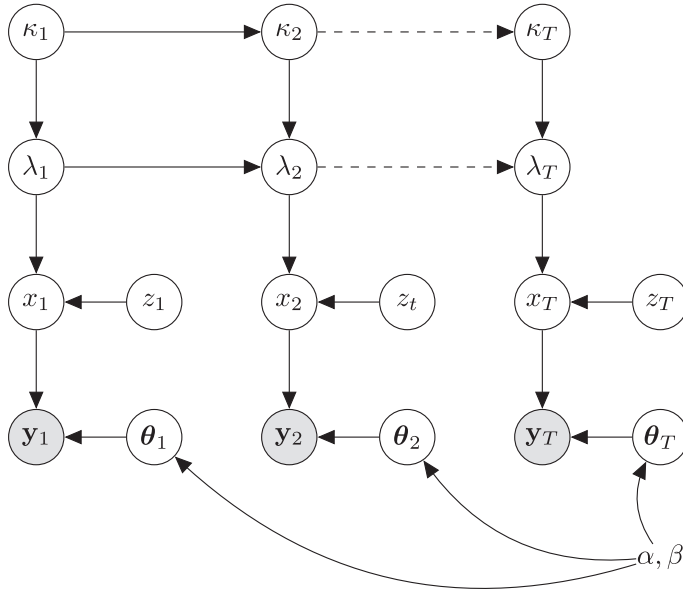


FIGURE 8 DAG depicting the dependencies in the model with joint distribution (17). The latent Poisson rates λ_t are coupled by the AR1 drift κ_t (13). Each x_t arises from a Poisson process with rate $\lambda_t z_t$, where $z_t = 1$ on weekdays. The report counts follow a beta-binomial distribution with parameters x_t, y_t, α, β which results from integrating out each $\theta_t^{(j)}$ under a $\text{Beta}(\alpha^{(j)}, \beta^{(j)})$ prior

shows the full conditional dependency structure for this model, encoded equivalently by the joint distribution

$$p(\lambda_0) \prod_{t=0}^T \left[p(x_t | \lambda_t, z_t) p(\lambda_{t+1} | \lambda_t, \kappa_{t+1}) p(\kappa_{t+1} | \kappa_t) p(z_t) p(y_t^{(T-t)} | x_t, \theta_t^{(j)}) p(\theta_t^{(j)}) \right]. \quad (17)$$

We can derive conditional independence relations between variables at different time points as follows. For day t , denote by $\theta_t = \{\theta_t^{(j)}\}_{j=t+1}^T$ the collection of under-reporting rates for each lagged report and by $\Omega_t = \{\lambda_t, \kappa_t, x_t, z_t, \theta_t, \mathbf{y}_t\}$ the collection of latent variables and observations, then applying the d -separation criteria Pearl (2009) to the DAG in Figure 8 we have the conditional independence relation

$$\Omega_t \perp\!\!\!\perp \Omega_{\neq t} | \lambda_t, \kappa_t. \quad (18)$$

We will use these conditional independence relations as the basis for drawing samples sequentially from the posterior.

6 | POSTERIOR INFERENCE

6.1 | Metropolis Hastings samplers

For each of the submodels discussed in Section 5 we draw posterior samples with standard Markov Chain Monte Carlo (MCMC) methods. For the time-independent submodels, Equations (7) and

(12) serve as potential functions for simple MH sampling. In Equation (7) the sum over x_t can be performed directly for benign¹⁰ choices of prior on x_t , when the true x_t is expected to be small, or by numerical integration when the prior is concentrated on a region of its support. To sample from $p(\lambda_t | y_t^{(T-t)})$ we use standard normal proposals and initialise the sampler to the mean of the prior $\mathbb{E}_{p(\lambda_t)}[\lambda_t]$. Since the expected distance for an n -step 1-dimensional Gaussian random walk is \sqrt{n} , we can be confident that the posterior is well explored by choosing an n such that our worst-case estimate of the absolute error $|x_t - \mathbb{E}_{p(\lambda_t)}[\lambda_t]|$ is well explored by \sqrt{n} . In all cases we apply thinning and burn in to the sample chain, though no hard convergence checks are made.

6.2 | Marginalized particle filtering and smoothing

In the time-dependent case, we make use of the conditional independence relation (18) to construct an algorithm for filtering and smoothing which builds on the simple MH sampling schemes as those used for inference in time-independent models. Inference mechanisms under the conditional dependencies induced by the random-walk prior in Section 5.3 are well studied (Doucet & Johansen, 2012) and so we give here the equations necessary for sequential posterior sampling of the latent λ_t, κ_t and x_t . At its foundation the complete model of Section 5.5 is Markov on a continuous latent state space with a non-conjugate emission distribution. We therefore employ a forward-backward algorithm, and deal with the non-conjugate structure by performing MH sampling at each time step. The atomic distributions constructed from these samples are propagated forwards as a prior to the subsequent time step, and re-sampled backwards for smoothing; this simple particle (SMC) algorithm differs from the canonical bootstrap filter since each filtering distribution is sampled directly rather than re-sampled from the previous predictive distribution.¹¹

In what follows we dispense with the LTLA index for brevity and outline the strategy for sequential sampling without weekend effects. The interested reader may refer to Appendices A.1 and A.2 for derivations of marginal filtering and smoothing distributions for all latent variables and including all effects proposed in Section 5. Consider that with N samples $\Lambda_{t-1}^{(i)}, K_{t-1}^{(i)} \sim p(\lambda_{t-1}, \kappa_{t-1} | \mathbf{y}_{0:t-1})$ from the joint filtering distribution at time $t - 1$ we may evaluate

$$p(\lambda_t, \kappa_t, \mathbf{y}_t | \mathbf{y}_{0:t-1}) \approx \frac{p(\mathbf{y}_t | \lambda_t)}{N} \sum_{i=1}^N p(\lambda_t | \kappa_t, \Lambda_{t-1}^{(i)}) p(\kappa_t | K_{t-1}^{(i)}) \quad (19)$$

where the sum over the atoms $\Lambda_t^{(i)}, K_{t-1}^{(i)}$ results from approximating $p(\lambda_{t-1}, \kappa_{t-1} | \mathbf{y}_{0:t-1})$ by the atomic distribution on MH samples. Sampling from $p(\lambda_0 | \mathbf{y}_0)$ is done exactly as in the time independent case using a potential given by Equation (12); by induction we can therefore compute a sequence of joint distributions of the form Equation (19) such that we may draw samples from the sequence of joint filtering distributions $p(\lambda_0, \kappa_0 | \mathbf{y}_0), \dots, p(\lambda_t, \kappa_t | \mathbf{y}_{0:t})$. We can construct the smoothing distributions by re-sampling these filtering atoms in a backward pass. Write the j th

¹⁰No prior could, for example, place any mass on counts in excess of the number of tests taken in its LTLA on the day under consideration.

¹¹We also implemented a standard bootstrap particle filter for the forward pass, but found that the accuracy of the posterior approximation worsened when the filtering predictive had poor coverage, due to high degeneracy in re-sampling particles in the tails.

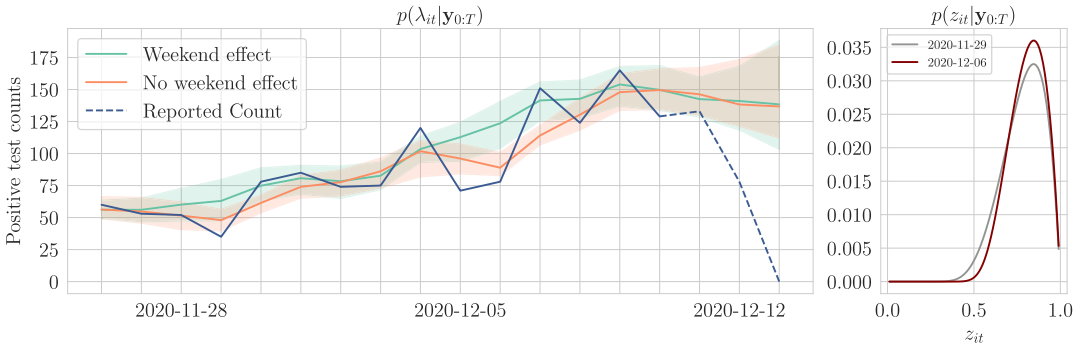


FIGURE 9 Smoothing distributions for λ_t and z_t in December for Canterbury with and without weekend effects. Inclusion of weekend effects allows λ_t to remain smooth over periodic decreases in the positive test count at weekends. Each $p(z_t | \mathbf{y}_{0:T})$ measures the strength of the weekend effect for day t ; on the right we show these posterior distributions for Sunday 29th November and 6th December

smoothing atoms as $L_{t+1}^{(j)}, \zeta_{t+1}^{(j)} \sim p(\lambda_{t+1}, \kappa_{t+1} | \mathbf{y}_{0:T})$. Then the re-sampling probabilities given M samples are

$$p(\lambda_t = \Lambda_t^{(i)}, \kappa_t = K_t^{(i)} | \mathbf{y}_{0:T}) \approx \frac{1}{M} \sum_{j=1}^M \frac{w_{ij}}{w_{*j}} \quad (20)$$

where the weights and normaliser are given by $w_{ij} = \delta \left(L_{t+1}^{(j)} - (\zeta_{t+1}^{(j)} + \Lambda_t^{(i)}) \right) \mathcal{N} \left(\zeta_{t+1}^{(j)}; K_t^{(i)}, \sigma^2 \right)$ and $w_{*j} = \sum_{i=1}^N w_{ij}$ respectively. The full procedure for inference is given by Algorithm 1. In Figure 9 we show the result of learning this smoothing distribution with and without weekend effects included. Given the smoothing atoms for each λ_t we can compute an approximate smoothing distribution for each x_t by

$$p(x_t | \mathbf{y}_{0:T}) \approx \frac{p(\mathbf{y}_t | x_t)}{M} \sum_{j=1}^M \frac{p(x_t | \zeta_t^{(j)})}{p(\mathbf{y}_t | \zeta_t^{(j)})} \quad (21)$$

which for time T gives us our required nowcast of the true count in the face of reporting lag. See Figure 14 for example nowcast showing uncertainty existing on recent counts and diminishing with time.

Algorithm 1: Particle SMC for smoothing distributions

Sample: N atoms $\{\Lambda_0^{(i)}\}_{i=1}^N$ from $p(\lambda_0 | \mathbf{y}_0)$

for $t \leftarrow 1$ **to** T **do**

$f(\lambda_t) = p(\mathbf{y}_t | \lambda_t) \sum_{i=1}^N p(\lambda_t | \Lambda_{t-1}^{(i)});$
Sample: N atoms $\{\Lambda_t^{(i)}\}_{i=1}^N$ by MH on $f(\lambda_t)$

Set: $\zeta_T^{(0)}, \dots, \zeta_T^{(N)} \leftarrow \Lambda_T^{(0)}, \dots, \Lambda_T^{(N)}$

for $t \leftarrow T - 1$ **to** 0 **do**

Sample: N atoms $\{\zeta_t^{(i)}\}_{i=1}^N$ by re-sampling $\{\Lambda_t^{(i)}\}_{i=1}^N$ with probabilities (A12)

return $p(\lambda_t | \mathbf{y}_{0:T}) \approx \frac{1}{N} \sum_{i=1}^N \delta(\lambda_t - \zeta_t^{(i)}) \quad \forall t \in 0, \dots, T$

6.3 | Kalman gain and moving averages

In public reporting of the positive test count, the UK government uses a windowed average computed over a seven-day period. This simple mechanism is to an extent well motivated as an approximation to linear dynamical models. For simplicity, consider modelling only the latest reported counts for times where $y_t^{(T-t)} = x_t$ by a Kalman Filter with no drift so that

$$p(x_0) = \mathcal{N}(\mu, \sigma^2) \quad (22)$$

$$p(x_t | x_{t-1}) = \mathcal{N}(x_{t-1}, \sigma^2) \quad (23)$$

$$p(y_t^{(T-t)} | x_t) = \mathcal{N}(x_t, \sigma_y^2). \quad (24)$$

Let the Kalman Gain (see Appendix D) for time-step t be $K_t \in [0, 1]$. The expectation of the Kalman filtering distribution at time-step t may be written as

$$\mathbb{E}_{p(x_t | y_{0:t})} [x_t] = K_t y_t + (1 - K_t) \mathbb{E}_{p(x_{t-1} | y_{0:t-1})} [x_{t-1}] \quad (25)$$

which expands recursively to give

$$\mathbb{E} [x_t] = \mu \prod_{i=1}^t (1 - K_i) + \sum_{i=1}^t K_i y_i \prod_{j=i}^t (1 - K_j) \quad (26)$$

$$= \text{const} + \sum_{i=1}^t y_i w_i \quad (27)$$

where $w_i = K_i \prod_{j=i}^t (1 - K_j)$. Equation (26) is a weighted average of the observed data. It is clear that $w_t > w_{\leq t-2}$ and further that if $K_t > 1/2$ then $w_t > w_{\leq t-1}$ so that the most recent observation has the largest weight. This gives a rough interpretation of the filtering distribution as a weighted average of the data with most weight given to the most recent observations.

In the absence of implementing a full model, we therefore suggest that a windowed average with decaying weights constitutes a type of posterior point-estimate for the Kalman Filter, which is in itself a reasonable approximation to our model for times where reported counts have converged to the truth; Figure 10 shows a comparison between our model, and a number of weighted moving averages which track the expected value of the smoothing posterior on λ_t .

6.4 | Model selection

Our model includes one free parameter: the scale σ of temporal smoothing applied by the random walk prior on the sequence $\kappa_{0:T}$ (see Equation 13). The choice of σ influences the set of feasible posterior distributions on λ_t ; in the absence of prior information on the rate of change of the disease it is important to choose σ to best explain the data observed. We take an empirical Bayes approach, and maximise the model evidence (Fong & Holmes, 2020)

$$p(\mathbf{y}_{1:T}) \approx p(\mathbf{y}_0) \prod_{t=1}^T \frac{1}{N} \int p(\mathbf{y}_t | \lambda_t) \sum_{i=1}^N \delta \left(\lambda_t - (\Lambda_{t-1}^{(i)} + \kappa_t) \right) p(\kappa_t | K_{t-1}^{(i)}) d\lambda_t d\kappa_t \quad (28)$$

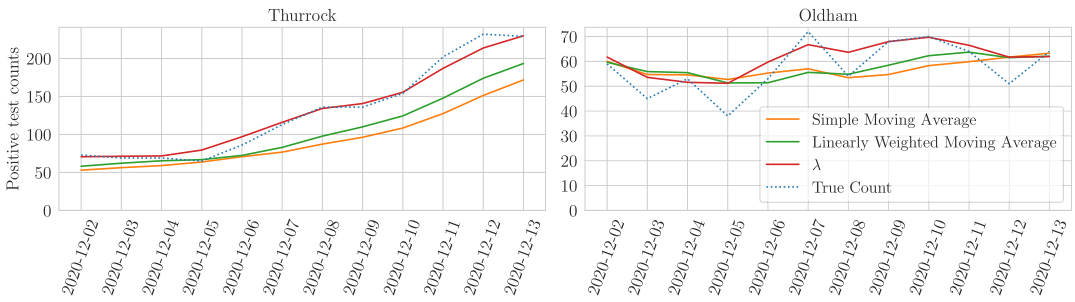


FIGURE 10 Smoothed counts for Thurrock and Oldham using seven day moving averages. The simple moving average weights all observations equally whereas the linearly weighted moving average gives more weight to recent observations. These are plotted alongside the expected value of the λ smoothing posterior distribution

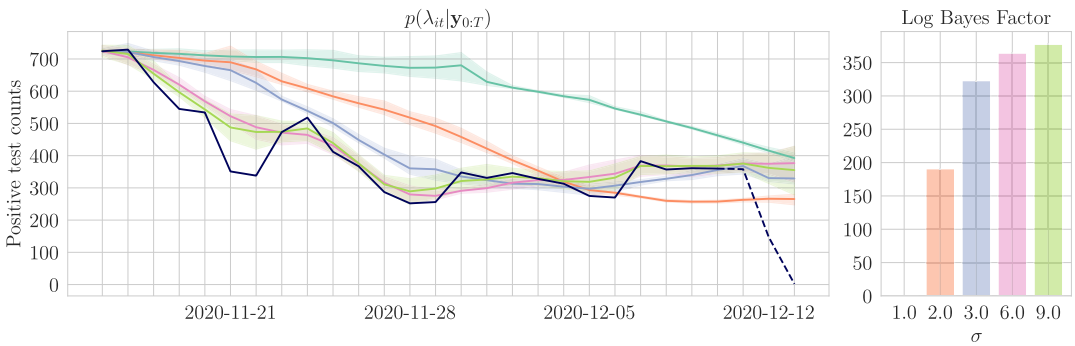


FIGURE 11 Smoothing distributions on the rates λ_{it} for Birmingham and their corresponding (relative) evidence given different values of σ . For example, the orange line is the λ_{it} posterior when setting $\sigma = 2$ whereas the purple posterior is given by setting $\sigma = 3$. The random-walk scale which maximises the evidence is $\sigma = 9$

over a feasible set of σ . From the perspective of a windowed average, this roughly corresponds to choosing window length and weights that best fit the observed data. Figure 11 demonstrates the influence of σ on the smoothing posterior. When the scale is too small relative to the variation in the counts, the posterior intensity cannot change quickly enough to capture well the most recent reported counts and risks missing changes in trend; when the scale is increased, the posterior intensity can decrease quickly enough to explain the observations. It is also possible to infer a smoothing posterior on σ under any prior with support on the positive reals. Figure E2 in Appendix E demonstrates the effect of inferring sigma under a half-normal prior for Thurrock.

7 | MODEL MONITORING AND ALERTING

One of the UK government’s main strategies for pandemic control has been to impose restrictions at the LTLA level in response to increasing case numbers. For our model to be useful in informing these decisions we must therefore have confidence in its predictions. Moreover, we would like to make assertions of probability on whether or not case numbers are following an increasing

trend, or have exceeded a threshold of concern. We now describe how our model can be used to (1) monitor systematic reporting errors, and (2) act as an alert system at the LTLA level.

7.1 | Monitoring

During the course of positive-test reporting thus far, there has been one major reporting error. A technical oversight resulted in the omission of 15,841 positive tests recorded between the 25th of September and 2nd of October, prior to a corrective backdate commenced on the 3rd of October. This fundamental change in reporting should be detectable under a well-calibrated model. The posterior predictive distribution $p(x_{t+1} | \mathbf{y}_{0:t})$ and lag- j smoothing distribution $p(x_{t+1} | \mathbf{y}_{0:t+j})$ each assign mass to a set of positive test counts at time $t + 1$ which are credible according to the model, but accounting for observations up to times t and $t + j$ respectively.

When the reporting process is well matched with the assumptions of our model, we expect these distributions to be more consistent than when there are large systematic reporting errors. As a mechanism for detecting these errors, we propose the following consistency statistic:

$$C_t^{(j)} = \mathbb{E}_{p(x_{t+1} | \mathbf{y}_{0:t+j})} [p(x_{t+1} | \mathbf{y}_{0:t})]. \quad (29)$$

When j is large, the smoothing distribution reduces to a delta function on the truth, and the statistic amounts to measuring how well the predictive distribution captures the true value. Extending this reasoning to the sequence of lagged reports for time $t + 1$, we recover the conditional evidence $p(\mathbf{y}_{t+1} | \mathbf{y}_{0:t})$ by integrating out all unobserved random variables. Since $y_t^{(j)}$ under-reports x_t when j is small, we may choose to evaluate

$$p(\mathbf{y}_{t-k:t+1} | \mathbf{y}_{0:t-k-1}) = \prod_{i=t-k}^{t+1} p(\mathbf{y}_{i+1} | \mathbf{y}_{0:i}) \quad (30)$$

as an aggregated measure of how well the model has captured reports which have not yet converged to a fixed value, which may constitute early warning signs against systematic reporting errors.

7.2 | Alerting

A fundamental assumption of the model (17) is that the positive test-counts x_t at time t are a-priori Poisson distributed with rate λ_t on weekdays,¹² and so $\text{Var}[x_t | \lambda_t] = \lambda_t$. Although windowed average techniques (see Section 6.3) may prove to be good estimators of the expected posterior on λ_t for times t such that we have observed $y_t^{(j)} = x_t$, these methods cannot include under-reports, or make statements of probability. For an alerting system that is both stable and probabilistic, we can examine the posterior behaviour of the $\lambda_{0:T}$ and $\kappa_{0:T}$. The difference $|\kappa_t - \kappa_{t+1}|$ may only substantially exceed σ when there is sufficient posterior evidence to do so after observing $\mathbf{y}_{0:T}$. This makes the marginal smoothing posteriors $p(\lambda_t | \mathbf{y}_{0:T})$ and $p(\kappa_t | \mathbf{y}_{0:T})$ ideal for establishing if the intensity of the epidemic has crossed a threshold, and whether or not the sequence of intensities

¹²On weekends, the rate is $\lambda_t z_t$ as discussed in Section 5.4.

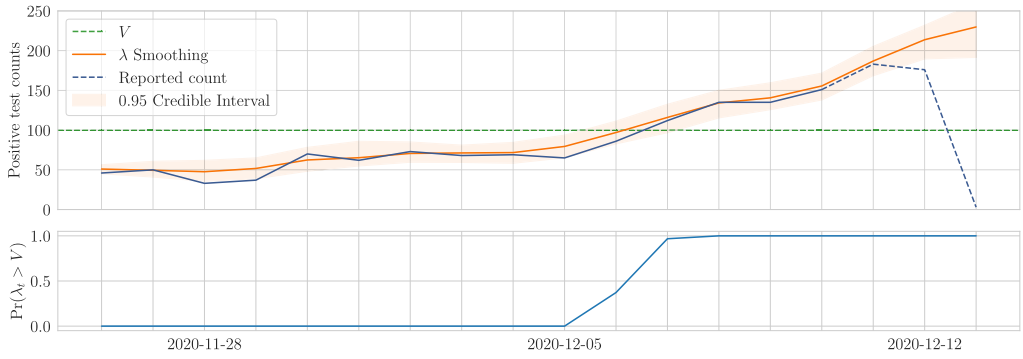


FIGURE 12 Example alert mechanism for Thurrock detecting if λ is above a threshold value. *Top*: the smoothing posterior on λ as well as the reported counts and a threshold V . *Bottom*: Probability that λ is above the threshold value at each time step

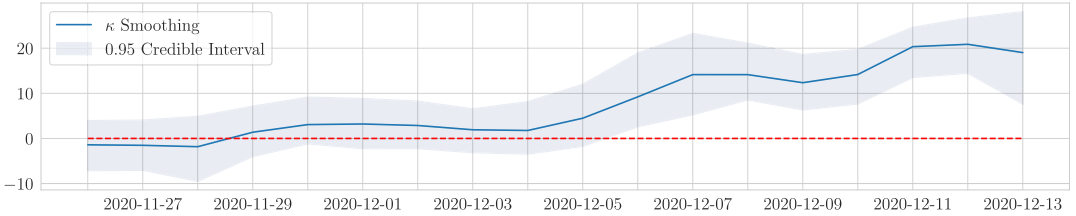


FIGURE 13 The smoothing posterior on the drift κ_t for Thurrock showing λ is following an increasing trend from 6th December onwards

is increasing. For some threshold value V we can use the fraction of smoothing particles which are larger than V

$$p(\lambda_t > V | \mathbf{y}_{0:T}) = \int_V^\infty p(\lambda_t | \mathbf{y}_{0:T}) d\lambda_t \tag{31}$$

$$\approx \frac{1}{M} \sum_{j=1}^M \mathbb{1} \left[\zeta_t^{(m)} > V \right] \tag{32}$$

to estimate an alerting probability. In Figure 12 we give an example of an alert for Thurrock between the 26th of November and the 13th of December and in Figure 13 we show how the smoothing posterior on $\kappa_{0:T}$ is easy to interpret visually as a description of whether the intensity is increasing.

8 | RESULTS

Under the model described in this report, the nowcast on x_T inferred from all available data is given by the smoothing distribution $p(x_T | \mathbf{y}_{0:T})$. In Figure 14 we display this nowcast for two LTLAs, as well as the smoothing distribution for times $t < T$; the uncertainty on x_t reduces with lag so for times far enough in the past the nowcast reflects our confidence that the reported count is the truth.

We show results over a 2-week period in December 2020 when the UK was transitioning out of lockdown, where the model we propose may be the most useful in helping to monitor changes in

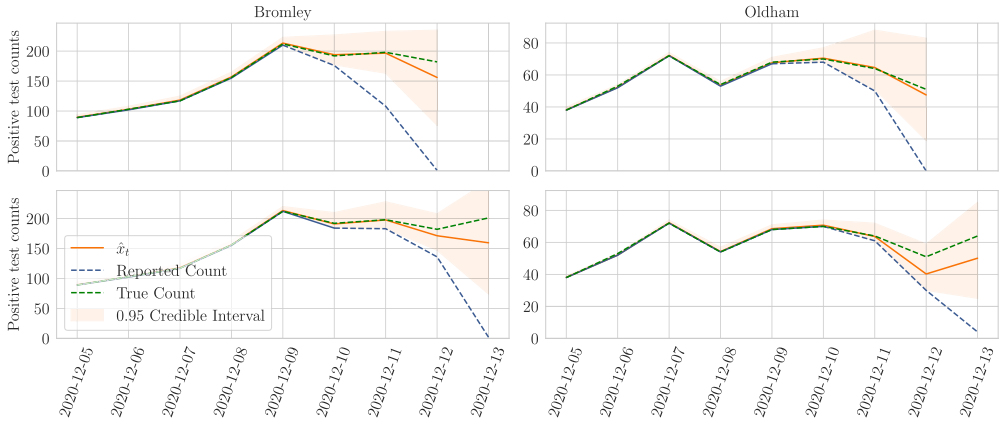


FIGURE 14 Example nowcasts $p(x_t | y_{0:T})$ for Bromley and Oldham made on the 13th of December and the 14th of December. The posterior is exactly the true count for times long in the past. For recent times where we don't trust the reports, there is uncertainty as captured by the posterior distribution. As new and updated data is reported, all nowcasts may be adjusted to account for this new information

the total number of positive tests. For each LTLA, we learn the reporting priors from the 2 weeks of most recently converged data within the given LTLA as described in Section 4.3 and we select the random walk scale that optimises the evidence as suggested by Section 6.4. For each model, we obtain 250 posterior samples (with a burn in of 750 iterations and thin rate of 20).¹³

8.1 | Moving average baseline

Although any statistical model on positive test cases provides the benefit of uncertainty quantification over non-statistical methods, the success of our model in deployment depends on its ability to improve upon simple techniques for estimating key indicators on the state of COVID-19. The 7-day windowed average on reports at the LTLA level, taken up to 4 days¹⁴ before present day, is such a simple and publicly reported baseline. We may use the latest week of smoothing distributions to estimate the 7-day windowed average up to present day, rather than 4 days ago, so that decisions on local restrictions may incorporate information from under-reported count data.

In absence of the full joint distribution $p(x_{t-7:T} | y_{0:T})$ we estimate the lag- j 7-day windowed average for day T $WA_T^{(j)}(7)$ from the marginals by

$$WA_T^{(j)}(7) = \frac{1}{7} \sum_{t=T-7}^T \mathbb{E}_{p(x_t | y_{0:T+t})} [x_t]. \quad (33)$$

To measure the performance of Equation (33) we assume the lag-7 reports have converged to x_t and measure the absolute error

$$AE_T^{(j)}(7) = \left| WA_T^{(j)}(7) - \frac{1}{7} \sum_{t=T-7}^T y_t^{(7)} \right|. \quad (34)$$

¹³With these values it takes just under 100 s to run our implementation of the model on a single time series with 14 time-steps using a MacBook Pro with Quad-Core Intel i7 processor. See Figure E3 for example traceplots.

¹⁴This is the minimum omission period employed. When the reliability of recent counts has decreased, this window is extended.

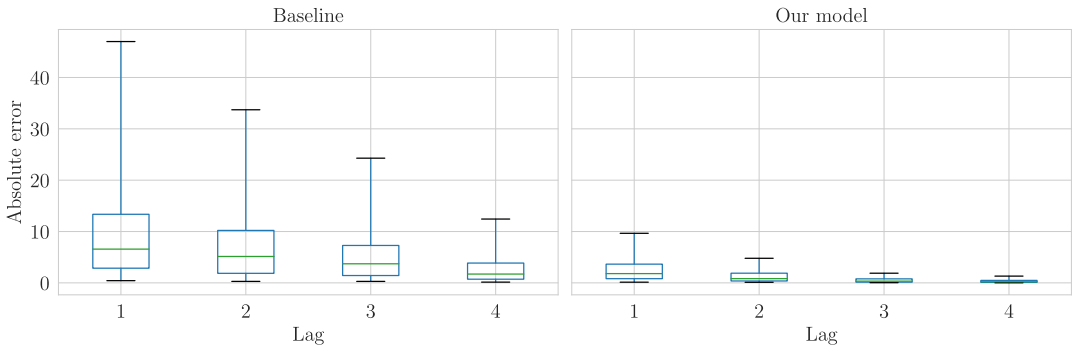


FIGURE 15 Distribution of absolute error for the imputed windowed average across all Lower-Tier Local Authorities by lag. The box plot whiskers cover the 5th and 95th percentiles. Model run on the 14th of December reports

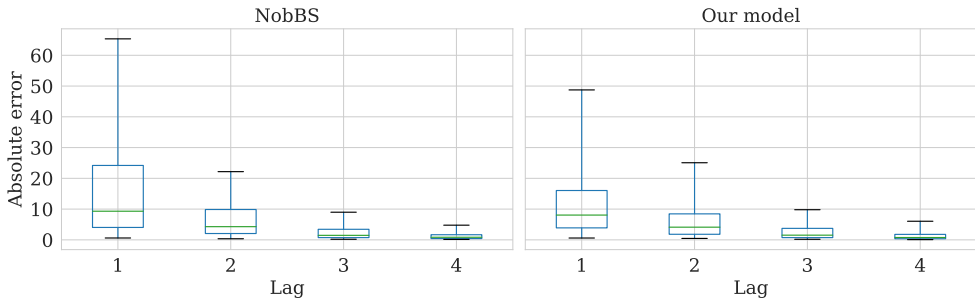


FIGURE 16 Distribution of absolute error for daily nowcasts made by our model and the NobBS approach across all Lower-Tier Local Authorities by lag. The box plot whiskers cover the 5th and 95th percentiles. Model run on the 14th of December reports

Figure 15 shows the distribution of absolute errors for the imputed 7-day windowed average across all LTLAs alongside the performance of the current strategy of ignoring under-reports. It is clear that using the expected smoothing distribution as a nowcast improves the mean absolute error at early stages of reporting, and further that there is a significant reduction in the variance of this estimator compared to the baseline.

8.2 | Nowcasting by Bayesian Smoothing (NobBS) baseline

We also compare nowcast accuracy at each lag against NobBS (McGough et al., 2020).¹⁵ With our model, we estimate the number of positive tests as the mean of the x_T smoothing distribution $\mathbb{E}_{p(x_T | y_{0:T})} [x_T]$, whilst NobBS uses $\mathbb{E}_{p(\lambda_T | y_{0:T})} [\lambda_T]$ as the nowcast estimate. In both cases we measure the absolute error between these estimates and the truth. Figure 16 shows the performance of both models for lags 1 to 4 marginally across all LTLAs; the models exhibit similar median performance. Figure E4 in Appendix E shows the same comparison but against Continuous Ranked Probability Score (CRPS) rather than absolute error.

¹⁵We adapt an implementation of NobBS in Stan (Carpenter et al., 2017) written by Hawryluk et al. (2021), which can be accessed here: https://github.com/ihawryluk/GP_nowcasting/blob/main/stan_models/nobbs.stan.

As mentioned in Section 1, the window length for inference in NobBS forces a trade off between reliable estimation of the random walk scale and the under-reporting process. We avoid this contention by choosing to estimate the under-reporting process with empirical Bayes on the most recent 2 weeks of reports. When the reporting process is changing on a time scale substantially shorter than the entire reporting period, our method for online estimation should better handle the dynamics.

9 | DISCUSSION

We have presented a probabilistic model for time series count data with weekend effects and a lagged reporting process. We demonstrated that it is possible to use incomplete, lagged reports and impute the true count to obtain a seven-day moving average that extends to the present date. Conditional on the nowcast being accurate, this is preferable to waiting for the most recent counts to converge before considering their use in monitoring the progress of the COVID-19 epidemic as it permits faster response to changes at the LTLA level. We also directly model the underlying intensity of the disease which lends itself naturally to probabilistic alerting strategies when incidence rates are increasing or have passed above some threshold value.

Our approach relates closely to a number of concurrent works. In Fronterre et al. (2020) a negative binomial model on swab tests is proposed, which is similar to that specified in Section 5.2 when considering the marginal posterior on x_{it} with λ_{it} integrated out. They further posit a spatio-temporal dependence on the Poisson rates which extends the work described here. A key difference between Fronterre et al. (2020) and here is the inference mechanism; Fronterre et al. (2020) employs restricted maximum-likelihood parameter estimation, whilst we pursue a fully Bayesian treatment such that all parameter uncertainties are propagated through to quantities of interest. The model described in Lee and Robertson (2020) shares the same binomial likelihood as proposed here. They couple the 442 Scottish post-code districts spatially with a (GMRF). A number of internal notes from PHE have also studied outliers under binomial likelihood models with some spatio-temporal dependence (Matt Keeling and the Warwick COVID modelling team, 2020; PHE Joint Modelling Cell and PHE COVID Outbreak Surveillance Team, 2020). Generally, LTLAs which are spatially close may exhibit similar trends in the number of positive tests, and so we may expect a spatial coupling between neighbouring LTLAs to yield performance improvements. The most common models for dealing with spatial dependence are typically based on GMRFs; the Besag York Mollié (BYM) model (Besag et al., 1991) posits spatial dependence through such a field, and although we do not employ this spatial structure here, a number of algorithms for posterior inference under this construction (Morris et al., 2019; Rue et al., 2009) have been used to study COVID-19 (D'Angelo et al., 2021). We note however that in this work we have observed mixed reliability in the spatial dependence of reporting rates.

Recent works have also modelled a lagged reporting process. For example, each of McGough et al. (2020), Seaman et al. (2020) and Stoner and Economou (2020) consider a temporal structure where the lagged counts are modelled as a generalised Dirichlet-multinomial distribution. This may deal more naturally with the reporting lag than our use of beta-binomial distributions on the latest available cumulative reports, though neither offers a posterior distribution on a latent intensity and would therefore have to rely on different strategies for probabilistic alerting than those justified in Section 7.2 here. Günther et al. (2021) propose a novel mechanism for capturing changes in the reporting delay distribution by using a linear spline with 2 week break-points and note its importance in improving the accuracy of their nowcast; in this work, we deal with these

changes in the delay distribution empirically by building a prior on the delay from the most recent 2 weeks of stable reporting data as discussed in Section 4.3.

The accuracy of our model in nowcasting the number of positive tests relies on stability of the reporting prior. As discussed in Section 4, over the course of the pandemic in the UK, there have been periods when the reporting delay was not stable, either due to technical issues in reporting or special circumstances such as Christmas holidays. Any approach that models the reporting behaviour to make a nowcast will not perform well in such cases, unless it can accurately estimate the dynamics of the reporting process. We suggest that building a reporting prior against testing laboratories rather than LTLAs may make available tighter prior estimates for the reporting rate—since reports for an LTLA constitute an aggregation of reports from testing sites serving that authority, variance in the rates may arise from differences in reporting behaviour across these testing sites. In fact, since the end of October, each lab has provided capacity estimates in publicly available data.¹⁶ It may be possible therefore to estimate lag at the lab level, conditional on the testing load faced by each lab.

The alerting framework proposed here directly extends monitoring strategies employed globally throughout the pandemic. A clear limitation is that Pillar 2 test counts suffer from ascertainment bias due to preferential testing of certain groups such as symptomatics or front-line workers. The counts are generally indicative of changing trends in cases but are also sensitive to changes in testing strategies and capacity. A robust alerting system could benefit from combining nowcasting with approaches that estimate true prevalence in the population by de-biasing test data such as proposed in Nicholson et al. (2022).

ACKNOWLEDGEMENTS

The authors thank the Joint Biosecurity Centre (JBC) for the proposal of and input to this problem. In particular we acknowledge Dmitry Semashkov for significant engineering effort in the deployment of this model. We also thank Theo Sanderson for providing a consistent and up-to-date repository of snapshots on open data¹⁷ and Angus Williams for his re-implementation of our model in Stan. This work was funded by The Department for Health and Social Care with in-kind support from The Alan Turing Institute and The Royal Statistical Society. BL was supported by the UK Engineering and Physical Sciences Research Council through the Bayes4Health programme [grant number EP/R018561/1] and gratefully acknowledges funding from Jesus College, Oxford.

DATA AVAILABILITY STATEMENT

All data and code used in the manuscript are publicly available at [https://urldefense.com/v3/https://github.com/alan-turing-institute/jbc-turing-rss-nowcasting_!!N11eV2iwtfs!pjiAVX-KON9EHvzIL4t8QH_cCAUKL0rNZX8n4FgAiHKecqoY6fanAne4goaPqOyXLQUA_3L_Ix52sw5PyMg\\$](https://urldefense.com/v3/https://github.com/alan-turing-institute/jbc-turing-rss-nowcasting_!!N11eV2iwtfs!pjiAVX-KON9EHvzIL4t8QH_cCAUKL0rNZX8n4FgAiHKecqoY6fanAne4goaPqOyXLQUA_3L_Ix52sw5PyMg$).

ORCID

Radka Jersakova  <https://orcid.org/0000-0001-6846-7158>

REFERENCES

Banbura, M., Giannone, D. & Reichlin, L. (2010) *Nowcasting*. ECB Working Paper No. 1275, 16. Available from: <https://ssrn.com/abstract=1717887>

¹⁶See <https://coronavirus.data.gov.uk/details/about-data>

¹⁷https://github.com/theosanderson/covid_uk_data_timestamped

- Besag, J., York, J. & Mollié, A. (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1–20. Available from: <https://doi.org/10.1007/BF00116466>
- Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M. et al. (2017) Stan: a probabilistic programming language. *Journal of Statistical Software*, 76(1).
- D'Angelo, N., Abbruzzo, A. & Adelfio, G. (2021) Spatio-temporal spread pattern of COVID-19 in Italy. *Mathematics*, 9(19), 2454. Available from: <https://doi.org/10.3390/math9192454>
- Doucet, A. & Johansen, A.M. (2012) A tutorial on particle filtering and smoothing: fifteen years later.
- Fong, E. & Holmes, C. (2020) On the marginal likelihood and cross-validation. *Biometrika*, 107, 489–496. Available from: <https://doi.org/10.1093/biomet/asz077>
- Frontierre, C., Read, J.M., Rowlingson, B., Alderton, S., Bridgen, J., Diggle, P.J. et al. (2020) COVID-19 in England: spatial patterns and regional outbreaks. [Accessed 9 July 2020]. Available from: <https://www.medrxiv.org/content/10.1101/2020.05.15.20102715v1.full.pdf>
- Günther, F., Bender, A., Katz, K., Küchenhoff, H. & Höhle, M. (2021) Nowcasting the COVID-19 pandemic in Bavaria. *Biometrical Journal*, 63(3), 490–502. Available from: <https://doi.org/10.1002/bimj.202000112>
- Hawryluk, I., Hoeltgebaum, H., Mishra, S., Miscouridou, X., Schnekenberg, R.P., Whittaker, C. et al. (2021) Gaussian process nowcasting: application to COVID-19 mortality reporting.
- Jung, R.C. & Tremayne, A.R. (2006) Binomial thinning models for integer time series. *Statistical Modelling*, 6(2), 81–96. Available from: <https://doi.org/10.1191/1471082X06st114oa>
- Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D), 35–45.
- Lee, D. & Robertson, C. (2020) A spatio-temporal Covid-19 surveillance tool for Scotland using telehealth data. Matt Keeling and the Warwick COVID modelling team. (2020) Preliminary exceedance methods.
- McGough, S.F., Johansson, M.A., Lipsitch, M. & Menzies, N.A. (2020) Nowcasting by Bayesian smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLoS Computational Biology*, 16(4), 1–20. Available from: <https://doi.org/10.1371/journal.pcbi.1007735>
- Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S.J., Gelman, A. & DiMaggio, C. (2019) Bayesian hierarchical spatial models: implementing the Besag York Mollié Model in Stan. *Spatial and Spatiotemporal Epidemiology*, 31, 100301. Available from: <https://doi.org/10.1016/j.sste.2019.100301>
- Nicholson, G., Lehmann, B., Padellini, T., Pouwels, K.B., Jersakova, R., Lomax, J. et al. (2022) Improving local prevalence estimates of SARS-CoV-2 infections using a causal debiasing framework. *Nature Microbiology*, 7(1), 97–107. Available from: <https://doi.org/10.1038/s41564-021-01029-0>
- Pearl, J. (2009) *Causality: models, reasoning and inference*, 2nd edition, Cambridge: Cambridge University Press.
- PHE Joint Modelling Cell and PHE COVID Outbreak Surveillance Team. (2020) COVID-19 local exceedance reporting.
- Rue, H., Martino, S. & Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2), 319–392. Available from: <https://doi.org/10.1111/j.1467-9868.2008.00700.x>
- Schneble, M., Nicola, G.D., Kauermann, G. & Berger, U. (2021) Nowcasting fatal COVID-19 infections on a regional level in Germany. *Biometrical Journal*, 63, 471–489.
- Seaman, S., Samartsidis, P., Kall, M. & De Angelis, D. (2020) Nowcasting COVID-19 deaths in England by age and region. *medRxiv*. Available from: <https://doi.org/10.1101/2020.09.15.20194209>
- Stoner, O. & Economou, T. (2020) Multivariate hierarchical frameworks for modeling delayed reporting in count data. *Biometrics*, 76, 789–798.

How to cite this article: Jersakova, R., Lomax, J., Hetherington, J., Lehmann, B., Nicholson, G., Briers, M. et al. (2022) Bayesian imputation of COVID-19 positive test counts for nowcasting under reporting lag. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1–27. Available from: <https://doi.org/10.1111/rssc.12557>

APPENDIX A. MARGINAL FILTERING AND SMOOTHING DERIVATIONS

In what follows we specify filtering and smoothing distributions for each time step t , from which we can draw samples but cannot compute in closed form. Denoting these samples by $\Lambda_t^{(i)}, K_t^{(i)} \sim p(\lambda_t, \kappa_t | y_t, \dots, y_0)$ and $L_t^{(j)}, \zeta_t^{(j)} \sim p(\lambda_t, \kappa_t | y_T, \dots, y_0)$, we build the atomic measures

$$p(\lambda_t, \kappa_t | \mathbf{y}_{0:t}) \approx \frac{1}{N} \sum_{i=1}^N \delta(\lambda_t - \Lambda_t^{(i)}) \delta(\kappa_t - K_t^{(i)}) \quad (\text{A1})$$

$$p(\lambda_t, \kappa_t | \mathbf{y}_{0:T}) \approx \frac{1}{M} \sum_{j=1}^M \delta(\lambda_t - L_t^{(j)}) \delta(\kappa_t - \zeta_t^{(j)}) \quad (\text{A2})$$

as particle approximations whose accuracy is determined by the number of samples N and M and the nature of the sampling scheme employed to draw these samples.

A.1. Filtering

For the forward pass we build a sequence of conditional distributions by

$$p(\lambda_t, \kappa_t, \mathbf{y}_t | \mathbf{y}_{0:t-1}) = \int p(\lambda_t, \lambda_{t-1}, \kappa_t, \kappa_{t-1}, \mathbf{y}_t | \mathbf{y}_{0:t-1}) d\lambda_{t-1} d\kappa_{t-1} \quad (\text{A3})$$

$$= p(y_t | \lambda_t) \int p(\lambda_t, \kappa_t | \lambda_{t-1}, \kappa_{t-1}) p(\lambda_{t-1}, \kappa_{t-1} | y_{0:t-1}) d\kappa_{t-1} d\lambda_{t-1} \quad (\text{A4})$$

$$= p(y_t | \lambda_t) \int p(\lambda_t | \kappa_t, \lambda_{t-1}) p(\kappa_t | \kappa_{t-1}) p(\lambda_{t-1}, \kappa_{t-1} | y_{0:t-1}) d\kappa_{t-1} d\lambda_{t-1} \quad (\text{A5})$$

$$\approx \frac{p(y_t | \lambda_t)}{N} \sum_{i=1}^N \delta(\lambda_t - (\kappa_t + \Lambda_{t-1}^{(i)})) \mathcal{N}(\kappa_t; K_{t-1}^{(i)}, \sigma^2) \quad (\text{A6})$$

from which we can draw samples with MH by first proposing a step in κ_t from a symmetric distribution, and then drawing λ_t uniformly from $\{\kappa_t + \Lambda_{t-1}^{(1)}, \dots, \kappa_t + \Lambda_{t-1}^{(N)}\}$.

A.2. Smoothing

We may use the conditional dependency structure of the model to write

$$p(\lambda_t, \kappa_t | \mathbf{y}_{0:T}) = \int p(\lambda_t, \kappa_t, \lambda_{t+1}, \kappa_{t+1} | \mathbf{y}_{0:T}) d\lambda_{t+1} d\kappa_{t+1} \quad (\text{A7})$$

$$= \int p(\lambda_t, \kappa_t | \lambda_{t+1}, \kappa_{t+1} \mathbf{y}_{0:t}) p(\lambda_{t+1}, \kappa_{t+1} | \mathbf{y}_{0:T}) d\lambda_{t+1} d\kappa_{t+1} \quad (\text{A8})$$

$$= p(\lambda_t, \kappa_t | \mathbf{y}_{0:t}) \int \frac{p(\lambda_{t+1}, \kappa_{t+1} | \lambda_t, \kappa_t)}{p(\lambda_{t+1}, \kappa_{t+1} | \mathbf{y}_{0:t})} p(\lambda_{t+1}, \kappa_{t+1} | \mathbf{y}_{0:T}) d\lambda_{t+1} d\kappa_{t+1} \quad (\text{A9})$$

$$= p(\lambda_t, \kappa_t | \mathbf{y}_{0:t}) \int \frac{p(\lambda_{t+1} | \kappa_{t+1}, \lambda_t) p(\kappa_{t+1} | \kappa_t)}{p(\lambda_{t+1}, \kappa_{t+1} | \mathbf{y}_{0:t})} p(\lambda_{t+1}, \kappa_{t+1} | \mathbf{y}_{0:T}) d\lambda_{t+1} d\kappa_{t+1} \quad (\text{A10})$$

$$\approx \frac{p(\lambda_t, \kappa_t | \mathbf{y}_{0:t})}{M} \sum_{j=1}^M \frac{p(L_{t+1}^{(j)} | \zeta_{t+1}^{(j)}, \lambda_t) p(\zeta_{t+1}^{(j)} | \kappa_t)}{p(L_{t+1}^{(j)}, \zeta_{t+1}^{(j)} | \mathbf{y}_{0:t})}. \quad (\text{A11})$$

Let $w_{ij} = \delta(L_{t+1}^{(j)} - (\zeta_{t+1}^{(j)} + \Lambda_t^{(i)})) \mathcal{N}(\zeta_{t+1}^{(j)}; K_t^{(i)}, \sigma^2)$, $w_{*j} = \sum_{i=1}^N w_{ij}$. Replacing the filtering distribution in Equation (A11) by its atomic measure we obtain

$$p(\lambda_t = \Lambda_t^{(i)}, \kappa_t = K_t^{(i)} | \mathbf{y}_{0:T}) \approx \frac{1}{M} \sum_{j=1}^M \frac{w_{ij}}{w_{*j}} \quad (\text{A12})$$

which indicates that each smoothing step is a multinomial resampling of the corresponding filtering atoms with weights (A12). For smoothing the counts x_t we begin with the joint

$$p(\lambda_t, x_t | \mathbf{y}_{0:T}) = p(x_t | \lambda_t, \mathbf{y}_{0:T})p(\lambda_t | \mathbf{y}_{0:T}) \quad (\text{A13})$$

$$= p(x_t | \lambda_t, \mathbf{y}_t)p(\lambda_t | \mathbf{y}_{0:T}) \quad (\text{A14})$$

and so

$$p(x_t | \mathbf{y}_{0:T}) = \int p(x_t, \lambda_t | \mathbf{y}_{0:T}) d\lambda_t \quad (\text{A15})$$

$$\approx \frac{1}{M} \sum_{j=1}^M p(x_t | L_t^{(j)}, \mathbf{y}_t) \quad (\text{A16})$$

$$\approx \frac{p(\mathbf{y}_t | x_t)}{M} \sum_{j=1}^M \frac{p(x_t | L_t^{(j)})}{p(\mathbf{y}_t | L_t^{(j)})} \quad (\text{A17})$$

so that if we have the smoothing atoms $\{L_t^{(j)}\}_{t=1, j=1}^{T, M}$ then we can evaluate (A17). With the addition of weekend effects we must alter the smoothing procedure for each x_t by considering

$$p(x_t, \lambda_t, z_t | \mathbf{y}_{0:T}) = p(x_t | \lambda_t, z_t, \mathbf{y}_{0:T})p(\lambda_t, z_t | \mathbf{y}_{0:T}) \quad (\text{A18})$$

$$= p(x_t | \lambda_t, z_t, \mathbf{y}_t)p(\lambda_t, z_t | \mathbf{y}_{0:T}) \quad (\text{A19})$$

$$= \frac{p(\mathbf{y}_t | x_t)p(x_t | \lambda_t, z_t)}{p(\mathbf{y}_t | \lambda_t, z_t)}p(\lambda_t, z_t | \mathbf{y}_{0:T}). \quad (\text{A20})$$

Now for the joint smoothing distribution on λ_t, z_t we have

$$p(\lambda_t, z_t | \mathbf{y}_{0:T}) = p(z_t | \lambda_t, \mathbf{y}_{0:T})p(\lambda_t | \mathbf{y}_{0:T}) \quad (\text{A21})$$

$$= p(z_t | \lambda_t, \mathbf{y}_t)p(\lambda_t | \mathbf{y}_{0:T}) \quad (\text{A22})$$

$$= \frac{p(\mathbf{y}_t, z_t | \lambda_t)}{p(\mathbf{y}_t | \lambda_t)}p(\lambda_t | \mathbf{y}_{0:T}) \quad (\text{A23})$$

$$= \frac{p(\mathbf{y}_t | \lambda_t, z_t)p(z_t)}{p(\mathbf{y}_t | \lambda_t)}p(\lambda_t | \mathbf{y}_{0:T}) \quad (\text{A24})$$

which we may employ to evaluate (A20).

APPENDIX B. EVIDENCE

By the sum and product rules of probability we may write

$$p(\mathbf{y}_{1:T}) = p(\mathbf{y}_0) \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{y}_{0:t-1}) \quad (\text{B25})$$

$$= p(\mathbf{y}_0) \prod_{t=1}^T \int p(\mathbf{y}_t, \lambda_t, \lambda_{t-1}, \kappa_t, \kappa_{t-1} | \mathbf{y}_{0:t-1}) d\lambda_t d\lambda_{t-1} d\kappa_t d\kappa_{t-1} \quad (\text{B26})$$

$$= p(\mathbf{y}_0) \prod_{t=1}^T \int p(\mathbf{y}_t | \lambda_t) p(\lambda_t | \lambda_{t-1}, \kappa_t) p(\lambda_{t-1}, \kappa_{t-1} | \mathbf{y}_{0:t-1}) p(\kappa_t | \kappa_{t-1}) d\lambda_t d\lambda_{t-1} \quad (\text{B27})$$

$$\approx p(\mathbf{y}_0) \prod_{t=1}^T \frac{1}{N} \int p(\mathbf{y}_t | \lambda_t) \sum_{i=1}^N \delta(\lambda_t - (\Lambda_{t-1}^{(i)} + \kappa_t)) \mathcal{N}(\kappa_t; K_{t-1}^{(i)}, \sigma^2) d\lambda_t \quad (\text{B28})$$

in (B28), we can also approximate the remaining integral by sampling from the Gaussian mixture built from transition distributions centered on the filtering particles at the previous time step. This yields an approximation to the evidence at each time step; sampling some $\kappa_t = K_t^{(i)}$ immediately implies which $\lambda_t = \Lambda_{t-1}^{(i)} + K_t^{(i)}$.

APPENDIX C. THE MOST RECENT COUNT $\mathbf{y}_{it}^{(T-t)}$ IS SUFFICIENT FOR \mathbf{x}_{it}

Defining $d_{it}^{(j)}$ and $\phi_{it}^{(j)}$ to be the first-order differences of $y_{it}^{(j)}$ and $\theta_{it}^{(j)}$ respectively,¹⁸ that is for any $J \in \{1, \dots, T-t\}$ we have

$$y_{it}^{(J)} \equiv \sum_{j=1}^J d_{it}^{(j)} \quad (\text{C29})$$

$$\theta_{it}^{(J)} \equiv \sum_{j=1}^J \phi_{it}^{(j)}, \quad (\text{C30})$$

and under a multinomial model for $d_{it}^{(1:J)}$,

$$d_{it}^{(1:J)} | x_{it}, \phi_{it}^{(1:J)} \sim \text{Multinomial}(x_{it}, \phi_{it}^{(1:J)}), \quad (\text{C31})$$

we demonstrate here that $y_{it}^{(J)}$ is a sufficient statistic for x_{it} , that is that

$$p(y_{it}^{(1:J)} | x_{it}, \theta_{it}^{(1:J)}) \stackrel{x_{it}}{\propto} p(y_{it}^{(J)} | x_{it}, \theta_{it}^{(J)})$$

where $\stackrel{x_{it}}{\propto}$ denotes proportionality with respect to x_{it} .

It is clear by definition (C29) that either of $d_{it}^{(1:J)}$ and $y_{it}^{(1:J)}$ can be reconstructed from the other; they therefore contain equivalent information on x_{it} . The analogous joint reconstruction property in (C30) allows us to condition interchangeably on $\theta_{it}^{(1:J)}$ and $\phi_{it}^{(1:J)}$.

Sufficiency of $y_{it}^{(J)}$ for x_{it} then becomes clear when viewed through the multinomial likelihood at (C31):

$$\begin{aligned} p(y_{it}^{(1:J)} | x_{it}, \theta_{it}^{(1:J)}) &\stackrel{x_{it}}{\propto} p(d_{it}^{(1:J)} | x_{it}, \phi_{it}^{(1:J)}) \\ &= \frac{x_{it}!}{(x_{it} - \sum_{j=1}^J d_{it}^{(j)})! \prod_{j=1}^J d_{it}^{(j)}!} \left(1 - \sum_{j=1}^J \phi_{it}^{(j)}\right)^{x_{it} - \sum_{j=1}^J d_{it}^{(j)}} \prod_{j=1}^J (\phi_{it}^{(j)})^{d_{it}^{(j)}} \end{aligned}$$

¹⁸The differences are non-negative as the reports are monotone increasing.

$$\begin{aligned} &\propto \frac{x_{it}!}{(x_{it} - y_{it}^{(J)})!} \left(1 - \theta_{it}^{(J)}\right)^{x_{it} - y_{it}^{(J)}} \\ &\propto \text{Binomial}(y_{it}^{(J)} | x_{it}, \theta_{it}^{(J)}). \end{aligned}$$

APPENDIX D. KALMAN GAIN

Let $v_t = \text{Var}_{p(x_t | y_{0:t})}(x_t)$ be the variance of the filtering posterior at time-step t . The Kalman Gain is given by

$$K_t = \frac{\sigma^2 + v_{t-1}}{\sigma^2 + v_{t-1} + \sigma_y^2} \quad (\text{D32})$$

which in effect controls the weight given to the most recent data point in computing the filtering posterior; the relative scale of v_{t-1} and σ_y determines this weighting.

APPENDIX E. SUPPORTING PLOTS

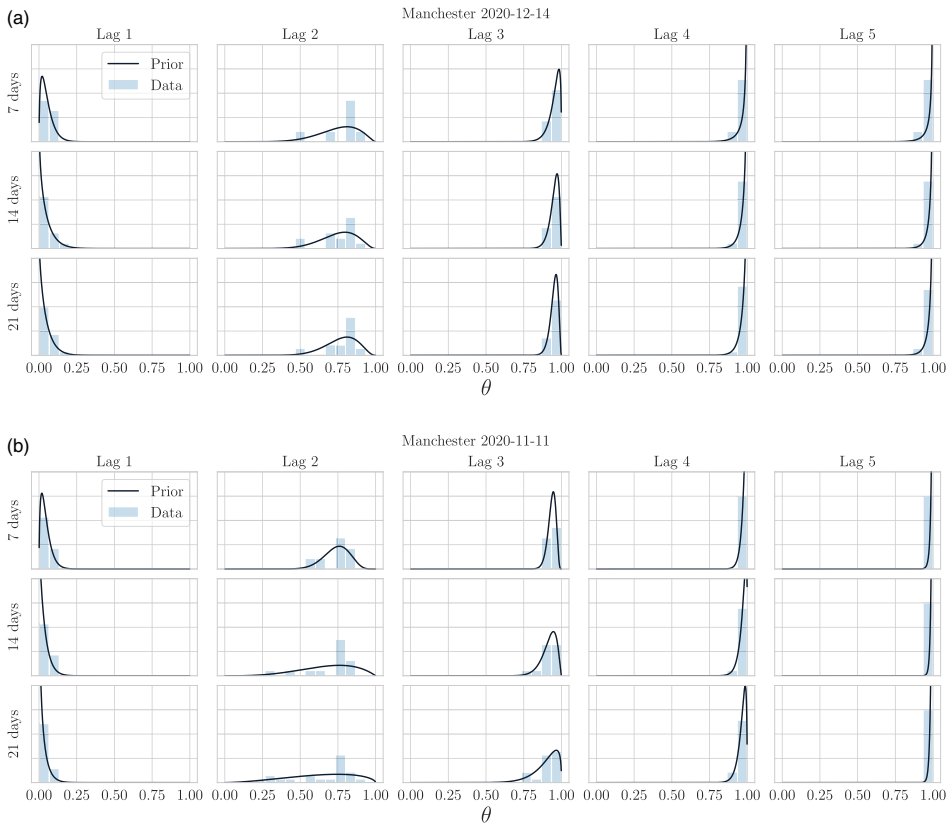


FIGURE E1 Empirical distributions and moment matched Beta priors $\theta_i^{(j)}$ for Manchester in a temporally stable period (14th December 2020, top) and temporally unstable period (11th November 2020, bottom). The priors are matched to either 7, 14 or 21 days of converged data. In a temporally stable period, the constructed priors are fairly insensitive to the number of data points considered. In a temporally unstable period we observe the priors broadening as more historical data is used, reflecting the instability in the reporting process. In order to improve this prior estimation, it would be necessary to pose a dynamic model on the reporting process

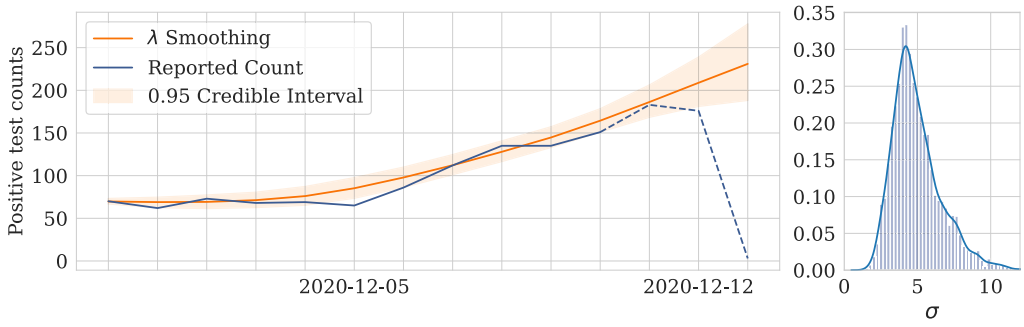


FIGURE E2 Smoothing distribution on the rate λ and σ for Thurrock using a model that places a half-normal prior on σ . The model was fit using Stan (Carpenter et al., 2017)

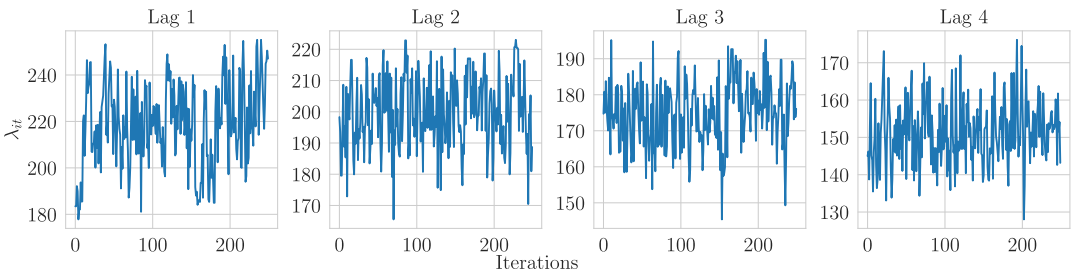


FIGURE E3 Trace plots for the filtering distributions λ_{it} for Thurrock by lag

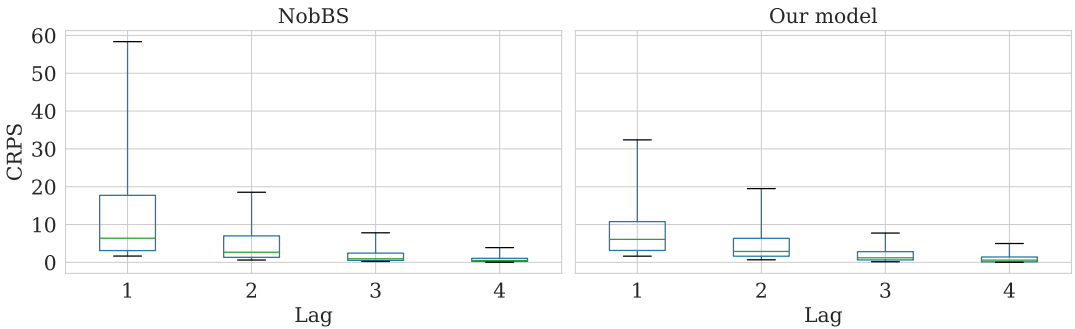


FIGURE E4 Distribution of Continuous Ranked Probability Score (CRPS) for daily nowcasts made by our model and the NobBS approach across all Lower-Tier Local Authorities by lag. The box plot whiskers cover the 5th and 95th percentiles. Model run on the 14th of December reports