

Efficient Bayesian Inference of Instantaneous Reproduction Numbers at Fine Spatial Scales, with an Application to Mapping and Nowcasting the Covid-19 Epidemic in British Local Authorities

Yee Whye Teh⁽¹⁾, Avishkar Bhoopchand⁽²⁾, Peter Diggle⁽³⁾, Bryn Elesedy⁽¹⁾, Bobby He⁽¹⁾, Michael Hutchinson⁽¹⁾, Ulrich Paquet⁽²⁾, Jonathan Read⁽³⁾, Nenad Tomasev⁽²⁾, Sheheryar Zaidi⁽¹⁾

⁽¹⁾Dept Statistics, University of Oxford

⁽²⁾Dept Statistics, University of Oxford, seconded from DeepMind

⁽³⁾Lancaster Medical School, University of Lancaster

1. Introduction

The spatio-temporal pattern of Covid-19 infections, as for most infectious disease epidemics, is highly heterogeneous as a consequence of local variations in risk factors and exposures. Consequently, the widely quoted national-level estimates of reproduction numbers are of limited value in guiding local interventions and monitoring their effectiveness. It is crucial for national and local policy makers as well as health protection teams that accurate, well-calibrated and timely predictions of Covid-19 incidences and transmission rates are available at fine spatial scales. Obtaining such estimates is challenging, not least due to the prevalence of asymptomatic Covid-19 transmissions, as well as difficulties of obtaining high resolution and high frequency data. In addition, low case counts at a local level further confounds the inference for Covid-19 transmission rates, adding unwelcome uncertainty.

In this paper we develop a hierarchical Bayesian method for inference of transmission rates at fine spatial scales. Our model incorporates both temporal and spatial dependencies of local transmission rates in order to share statistical strength and reduce uncertainty. It also incorporates information about population flows to model potential transmissions across local areas. A simple approach to posterior simulation quickly becomes computationally infeasible, which is problematic if the system is required to provide timely predictions. We describe how to make posterior simulation for the model efficient, so that we are able to provide daily updates on epidemic developments.

The results can be found at our website <https://localcovid.info> which is updated daily to display estimated instantaneous reproduction numbers and predicted case counts for the next weeks, across local authorities in Great Britain. We hope that our methodology and website will be of interest to researchers, policy makers and the public alike, to help identify upcoming local outbreaks and to aid in the containment of Covid-19 through both public health measures as well as personal decisions taken by the general public.

2. Data

Our model is applied to publicly available daily counts of positive test results reported under the combined Pillars 1 (NHS and PHE) and 2 (commercial partners) of the UK’s Covid-19 testing strategy¹. The data is available for 312 lower-tier local authorities (LTLAs) in England, 14 NHS Health Boards in Scotland (each covering multiple local authorities), and 22 unitary local authorities in Wales, for a total of $n = 348$ local areas. The data are daily counts of lab-confirmed (PCR swab) cases presented by specimen date, starting from January 30, 2020. The original data are from the respective national public health authorities of England², Scotland³ and Wales⁴ and we access them through the DELVE Global Covid-19 Dataset⁵ (Bhoopchand et al., 2020). Due to delays in processing tests, we ignore the last 7 days of case counts.

3. Method

Our method is based on an approach to infectious disease modelling using discrete renewal processes. These have a long history, and have served as the basis for a number of recent studies estimating instantaneous reproduction numbers, most notably (Wallinga and Teunis, 2004; Fraser, 2007; Cori et al., 2013; Flaxman et al., 2020). See Bhatt et al. (2020) and references therein for historical and mathematical background, as well as Gostic et al. (2020) for important practical considerations.

Following Flaxman et al. (2020), we model latent time series of incidence rates via renewal processes, and separate observations of reported cases using negative binomial distributions, to account for uncertainties in case reporting, outliers in case counts, and delays between infection and testing. We introduce a number of extensions and differences addressing issues that arise for applications to modelling epidemics at local authority level rather than regional or national levels. Firstly, we

introduce dependencies between reproduction numbers across neighbouring localities, in order to smooth estimates of reproduction numbers and share statistical strength across localities and time. We do this using a spatiotemporal Gaussian process (GP) prior for the log-transformed reproduction numbers. Secondly, we model transmissions across localities using a spatial metapopulation model. Our metapopulation model incorporates commuter flow data from the UK 2011 Census in order to capture heterogenous cross-infection rates among local authorities. Human mobility patterns may also reflect the introduction of non-pharmaceutical interventions (NPIs), though our model does not explicitly use real-time mobility data so cannot estimate the direct or indirect effects of NPIs.

The model is implemented in the Stan probabilistic programming language (Carpenter et al., 2017), which uses the No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014) for posterior simulation. A number of modelling design choices as well as inference approximations are made to improving mixing and computational efficiency. These are described in Appendix B.

3.1. Model Overview

In this section we give an overview of our model, which we refer to as EpiMap. The model consists of three layers: a latent Gaussian process over the log reproduction numbers, a metapopulation model for the epidemics across local areas, and an observation model relating the size of the epidemic with the observed number of positive tests in each day and area.

We first introduce some notations. We are interested in estimating the instantaneous reproduction numbers, $R_{i,t}$, across local areas in the UK (indexed by i) and across time (indexed by t). For each local area i and day t , the observed daily Pillars 1+2 case counts are denoted $C_{i,t}$. Let the unobserved daily infection (incidence) counts be $X_{i,t}$.

Starting with the observation model, we model the number of reported cases using a delay distribution and an over-dispersed negative binomial observation model:

$$C_{i,t}|X_{i,1:t}, \phi_i \sim \text{NegBin}(V_{\text{day_of_week}(t)}E_{i,t}, \phi_i), \quad E_{i,t} = \sum_{s=1}^t X_{i,t-s}D_s, \quad (1)$$

where D_s is the probability that an infected person gets tested and tests positive s days after infection and $E_{i,t}$ is the expected number of positive test cases on day t in area i . $\text{NegBin}(\mu, \phi)$ is the negative binomial

distribution with mean μ and dispersion parameter ϕ , while $V_{\text{day.of.week}(t)}$ models day-of-week variations in reported cases. Section 3.1.2 gives more details.

Assuming a homogeneously mixing population in each area, and interactions across areas modelled using a cross-coupled metapopulation model, we model the number of new infections in each area as follows. Conditioned on the history of infections, let

$$Z_{i,t} = \sum_{s=1}^t X_{i,t-s} W_s \quad (2)$$

be the infection load on day t caused by previous infections in area i , if each primary case produces one secondary case. W_s describes the generation distribution, and is the probability that a secondary infection occurs s days after the primary infection. See Section 3.1.2 for more details on how we parameterise W_s . These secondary infections can occur in area i , or in another area, e.g. due to individuals working in an area different from where they live. We model this with a time dependent flux matrix $F_{ji}^{(t)}$, which is interpreted as the probability that a primary case living in area j infects a secondary case living in area i on day t . The resulting cross-coupled infection load in area i is:

$$\tilde{Z}_{i,t} = \sum_{j=1}^n F_{ji}^{(t)} Z_{j,t}. \quad (3)$$

We describe the metapopulation model in further detail in Section 3.1.3, including how the flux matrices are parameterised. We model the number of new infections on day t as,

$$X_{i,t} | R_{i,t}, X_{1:n,1:t-1} \sim \text{NegBin}(R_{i,t} \tilde{Z}_{i,t}, \psi) \quad (4)$$

where $R_{i,t} \tilde{Z}_{i,t}$ is the force of infection in area i and day t , and ψ is a dispersion parameter which allows for over-dispersion. We expect this to be a better model for Covid-19 than using a Poisson distribution in (4) due to super-spreading events. Note that if we used a Poisson then the secondary infections resulting from a primary infection would have been modelled as conditionally iid given the primary infection. The use of a negative binomial distribution instead introduces a positive correlation among the secondary infections.

In order to make the posterior simulation computationally efficient using Stan, we approximated this with a positivised Gaussian distribution; see Appendix B.2.

3.1.1. Latent Gaussian Process

With low case counts, inferring $R_{i,t}$ over small local areas can lead to high uncertainty. A standard Bayesian hierarchical modelling approach is to share statistical strengths across the inferences of different local areas and across different time points. We use Gaussian processes (GPs) to do so; namely, for area i and time t we model:

$$R_{i,t} = \exp(S_{i,t} + U_{i,t}) \quad (5)$$

where $S_{:,}$ is a GP with a separable Matern(1/2) kernel:

$$\text{Cov}(S_{i,s}, S_{j,t}) = (\sigma^{\text{spatial}})^2 \exp(-\|y_i - y_j\|/\rho^{\text{spatial}} - \|s - t\|/\rho^{\text{time}}) \quad (6)$$

and $U_{i,:}$ are independent copies of a GP with Matern(1/2) kernels:

$$\text{Cov}(U_{i,s}, U_{i,t}) = (\sigma^{\text{local}})^2 \exp(-\|s - t\|/\rho^{\text{time}}) \quad (7)$$

Here, y_i and y_j are the geographical centres of areas i and j respectively, s and t are weekly time indices, as we assume that the instantaneous reproduction numbers are constant within each week. Note that our prior covariances in Equations 6 & 7 enjoy a Kronecker structure across the space and time dimensions, which allows for efficient computations (see Appendix B.1). In the temporal case, which is one-dimensional, the GP prior with the Matern(1/2) kernel is equivalent to an AR(1) process with 0 mean. We also considered Matern(3/2), Matern(5/2) and squared-exponential covariance kernels, which produced similar inferences.

The hyperparameters of the spatiotemporal GP are: scale parameters σ^{spatial} and σ^{local} and length scale parameters ρ^{spatial} and ρ^{time} . We place independent truncated normal priors $\mathcal{N}_+(0, 0.5)$ over the scale parameters. For the length scale parameters, we have found that if we inferred these along with the rest of the random variables in the model, the posterior distribution places mass on large spatial length scales and short temporal length scales. This has an undesirable over-generalisation effect, and we believe this behaviour is due to model misspecification with respect to the length scale parameters. Instead we selected these using an initial cross validation run optimising for performance of forecasted case counts three weeks into the future, and selected $\rho^{\text{spatial}} = 10\text{km}$ and $\rho^{\text{temporal}} = 200$ days.

3.1.2. Observation and Infection Model

Weekly variations are modelled using multiplicative factors in (1), with a uniform prior over positive vectors of length 7 and sums to 7. Following Flaxman et al. (2020) we use an over-dispersed negative binomial

observation model (1), with a broad half normal prior for the dispersion parameters, $\phi_i \sim \mathcal{N}_+(0, 5)$ iid. However, we use a different parameterisation of the negative binomial, rather than the `neg.binomial.2` parameterisation in Stan, with a mean parameter μ , an inverse-dispersion parameter c , and variance $\mu + \mu^2/c$. Instead we use a different parameterisation, and set $c = \mu/\phi$, where ϕ is a dispersion parameter. This gives a variance of $(1 + \phi)\mu$ and probability mass function:

$$p(x|\mu, \phi) = \binom{x + \mu/\phi - 1}{x} \left(\frac{\phi}{1 + \phi} \right)^x \left(\frac{1}{1 + \phi} \right)^{\mu/\phi} \quad (8)$$

This parameterisation is the natural one which emphasises the infinite divisibility of the negative binomial, i.e. if Y_1, \dots, Y_m are independent negative binomial random variables with means μ_1, \dots, μ_m and the same dispersion parameter ϕ , then $\sum_{i=1}^m Y_i$ is also negative binomially distributed with mean $\sum_{i=1}^m \mu_i$ and dispersion ϕ , a sensible choice in cases where we believe counts are sums of independent random events.

The infection-to-test delay distribution D_s is a convolution of two delay distributions: an incubation period distribution, and a symptom-onset-to-test distribution. Following Bi et al. (2020), we use a LogNormal(μ, σ^2) distribution for the incubation period, where μ has a 95% confidence interval (CI) of (1.44, 1.69) and mode 1.57, and σ^2 has 95% CI of (0.56, 0.75) with mode 0.65. This results in a median of 4.8 days and a 90% confidence interval of (1.64, 14.04) days for the incubation period, and we assume an additional two day delay to get tested.

Similarly, we parameterise the generation distribution W_s as a Gamma distribution where shape parameter has mode 2.29 with (1.77, 3.34) 95% CI, and rate parameter has mode 0.36 with (0.26, 0.57) 95% CI. This corresponds to the serial interval parameter distributions from Bi et al. (2020); we note that the serial interval is often used as an accessible proxy for the unobserved generation distribution (Cori et al., 2013). For both D_s and W_s , we aggregate predictions and inferences from 10 bootstrapped runs of our model, each with independently sampled LogNormal and Gamma parameters respectively. This is equivalent to a nested Monte Carlo approximation to a cut or modular model (Plummer, 2015; Jacob et al., 2017; Carmona and Nicholls, 2020). We found this to be crucial to avoiding overconfident predictions for R_t estimates.

For the dispersion parameter ψ , we use a weakly informative prior $\psi \sim \mathcal{N}_+(0, 2.5)$.

3.1.3. Metapopulation Model

Our final extension relaxes the assumption in many infectious disease models, that the epidemic is evolving in a homogeneously mixing population in an area, with no significant transmissions from other areas. While this might be sensible in large regions or countries, it is not a sensible assumption for modelling multiple small areas with likely a significant number of cross-area transmissions. To address these transmissions, we describe a simple cross-coupled metapopulation extension, given by equations (3)-(4).

In the following we describe how to parameterise the flux F_{ji} , which describes the chance that a primary case living in area j , *if they infect a secondary case*, that the secondary case lives in area i . One sensible choice, if the data were available, would be to use real-time data on the actual volume of travel between each pair of areas. Such data is unfortunately not publicly available, and in any case the relationship between the volume of travel and the number of transmissions is not straightforward due to heterogeneity in the population.

We use commuting flow data from the 2011 Census⁶. The data gives, after some preprocessing, a matrix M such that for each pair of areas i and j the number of individuals who live in area j and commute to work in area i is M_{ji} . Let P_j be the population of area j . We take M_{jj} to be the population who commute within their own area or who do not commute, so $\sum_i M_{ji} = P_j$. We consider three types of transmissions: an individual living in area j infecting another individual in area j (e.g. household transmissions), an individual living in area j working in area i infecting one living in area i , and an individual living in area i being infected while working in area j . These three types of transmissions can be described using three flux matrices:

$$F_{ji}^{\text{id}} = \delta_{ji} \quad F_{ji}^{\text{fwd}} = \frac{M_{ji}}{\sum_k M_{jk}} \quad F_{ji}^{\text{rev}} = \frac{M_{ij}}{\sum_k M_{kj}} \quad (9)$$

where $\delta_{ji} = 1$ if $j = i$ and 0 otherwise. We parameterise the overall flux matrix during week t using a convex combination of F^{id} , F^{fwd} and F^{rev} ,

$$F^{(t)} = \alpha_t F^{\text{id}} + (1 - \alpha_t)(\beta F^{\text{fwd}} + (1 - \beta)F^{\text{rev}}) \quad (10)$$

with $\alpha_t \in (0, 1)$ governing the amount of mixing across areas on week t (roughly the proportion of the population working from home), and $\beta \in (0, 1)$ governing the amount of home-to-work versus work-to-home transmissions. We use a uniform prior over β and a weekly AR(1) prior for the log-odds, specifically $\alpha_t = 1/(1 + \exp(-\mu_\alpha + \sigma_\alpha A_t))$ where the

AR(1) process is given by $A_1 \sim \mathcal{N}(0, 1)$, $A_t | A_{t-1} \sim \mathcal{N}(\delta_\alpha A_{t-1}, \sqrt{1 - \delta_\alpha^2})$, with weakly informative hyperpriors $\mu_\alpha \sim \mathcal{N}(0, 0.5)$, $\sigma_\alpha \sim \mathcal{N}_+(0, 0.5)$, while the hyperprior $\delta_\alpha \sim \mathcal{N}_{[0,1]}(1, 1 - e^{-0.25})$ is a weakly informative prior on the time scale of the AR(1) process centred around 4 weeks.

4. Empirical Evaluations

In this section, we report some empirical evaluations of our model, which we call EpiMap. We compared two variants of EpiMap: one which models each local area separately from the rest (hence no metapopulation model nor spatial component of GP), and one the full model. For the full model we have found that the inferences are sensitive to the length scale of the spatial GP, and so we compared the full model with varying spatial length scales and with no spatial GP component. We also compared against EpiEstim (Cori et al., 2013) and EpiNow2 (Abbott et al., 2020). We compared these methods on simulated data and on predicting future case counts in British local authorities. We also report estimates of R_t at regional and national levels.

4.1. Simulation data

One sanity check of our method is to fit the models to simulated data for which we know the underlying R_t , and check how well our models can recover this. In this section we do just this, and compare the results with a number of other common methods.

The simulation model we use is exactly the generative model we described. The data is simulated by taking initial real cases data from Oxford and the 4 surrounding LTLAs up to 2020-03-14, and from that point simulating new cases using the model. The main unspecified parameter is the R_t in each region over time. An R_t curve was manually designed in order to give a double peak epidemic similar in nature to the pattern seen accross the UK, with case numbers in the regions roughly similar. The same R_t curve was shared accross the LTLAs. Additionally we use 50:50 flux proportions of the forward and reverse commuter flow data, with a constant α_t of 0.45. These choices of parameters are somewhat arbitrary and were chosen to give qualitatively sensible looking epidemic curves. To these simulated data we fit the two variations of our model, with the full model using a temporal length scale of 200 days and a range of spatial length scales between 1 km and 100 km. The results can be seen in Figure 1. Plots showing the full sweep of spatial length scales for EpiMap can be found in Appendix Appendix C.1.

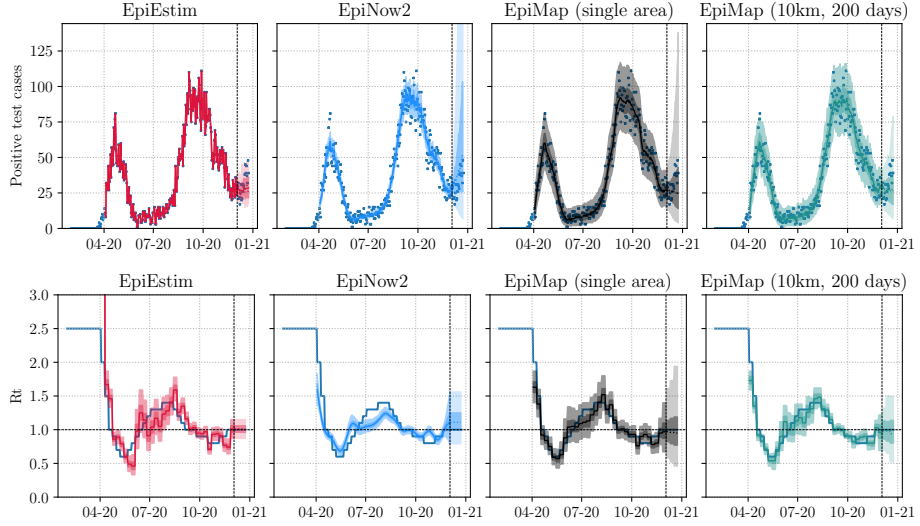


Fig. 1. Upper: Estimated median, 50% (inner) and 95% (outer) credible intervals of the posterior predictive distributions for $C_{i,t}$, along with observed case counts. Credible intervals to the right of the vertical line are future predictions. For EpiMap, we include the weekly variation in expected cases in the model, but plot the predictive distributions without this variation for clarity. EpiNow2 returns the distribution of the mean number of cases so under-estimates the uncertainty. EpiEstim does not model dependence of R_t over time, so we used the last inferred R_t distribution for future predictions. Lower: Estimated median, 50% (inner) and 95% (outer) credible intervals of R_t for the methods, along with the true R_t used to create the simulated epidemic. Plots shown only for Oxford, those for the four surrounding local authorities are given in the Appendix. EpiEstim does not model dependence of R_t over time, so we used the last inferred R_t distribution for future predictions.

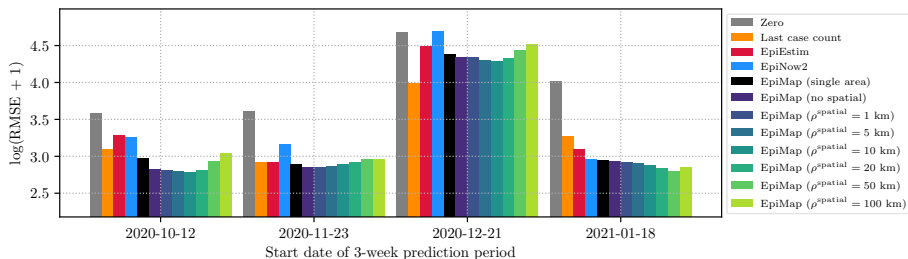


Fig. 2. A comparison of models for predicting future case counts over a 3-week unseen period. Each model is fitted on 15 weeks of data and makes predictions for the following 3 weeks.

The three methods produce qualitatively different estimates of R_t . EpiEstim struggles to estimate R_t well when case numbers are low, producing high variance estimates with poorly calibrated uncertainty. This is likely because EpiEstim does not account for noise in the observation of cases, or the weekly variation in reporting. Additionally these R_t lag by about a week as EpiEstim does not account for delays in the reporting of cases. EpiNow2 produces overconfident predictions which appear over-smoothed. Single area EpiMap recovers R_t reasonably well, with good uncertainty calibration. The full regional EpiMap model provides a small gain on top of the single area model, with slightly tighter credible intervals and reasonable median accuracy. Section 4.2 produces more quantitative comparisons of the methods.

4.2. Predicting future case counts

Next, we evaluate the methods' predictions of future case counts by comparing them to true case counts. In addition to measuring predictive performance, we also assess the model's uncertainty calibration by comparing the coverage probability of its prediction intervals with the actual, achieved (empirical) coverage. We first picked four spread out dates: 2020-10-12, 2020-11-23, 2020-12-21 and 2021-01-18. For each date, we use the 15 preceding weeks of data for inference and evaluated predictions of case counts for the following 3 weeks. Note that since the methods do not model drastic changes arising from NPIs changing, we expect them to perform poorly during such periods. Therefore, we include 2020-12-21 as an example of such a period, because there are drastic changes to the NPIs, whereas the remaining three dates involve periods without substantial changes to the spread dynamics. In addition to the variants of

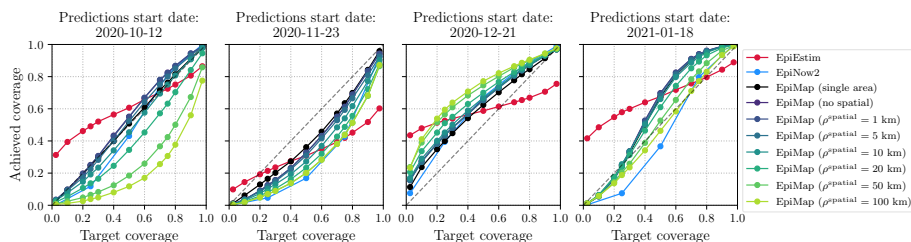


Fig. 3. Reliability curves assessing the uncertainty estimates produced by models. Each model yields percentiles of its case count posterior predictive distribution. The curves show the portion \hat{p} of predictions (across dates and LTLAs) for which the true case count is less than the p -th percentile \hat{c}_p of the model (y -axis) vs. p (x -axis).

EpiMap, EpiEstim and EpiNow2, we also included two simple baselines: “zero” which predicts zero cases for all dates and LTLAs, and “last case count” which predicts using the case count on the last day of the 15-week inference period for each LTLA.

Figure 2 shows $\log(\text{RMSE}+1)$ between predicted and true case counts. More precisely, the RMSE is separately computed for each LTLA’s predictions over the test period, then we average the resulting $\log(\text{RMSE}+1)$ across LTLAs. The log transformation is so that results are not dominated by areas with much higher case counts. EpiMap variants typically perform best at predicting the true case counts. The positive impact of modelling cross-area dependencies is observed, since EpiMap (single area) tends to slightly underperform the other variants of EpiMap. Moreover, the predictive performance of EpiMap is dependent on, though not very sensitive to, the choice of ρ^{spatial} . Note that for the start date 2020-12-21, all models perform worse relative to other dates. This is because of significant changes in the dynamics of Covid-19 spread due to changing NPIs over the Christmas period, information that is not incorporated into any of these models.

Figure 3 assesses the quality of the uncertainty estimates produced by the models using reliability curves. Each model outputs percentiles of the posterior predictive distribution of case counts. Let \hat{c}_p be the p -th percentile produced by a model for a given date and LTLA. Ideally, we expect that the percentage of dates and LTLAs for which the true case count c is less than or equal to \hat{c}_p , is approximately p . In other words, the actual, empirical coverage of the p -th percentile (y -axis of Figure 3) will ideally be equal to the target coverage p (x -axis of Figure 3), yielding

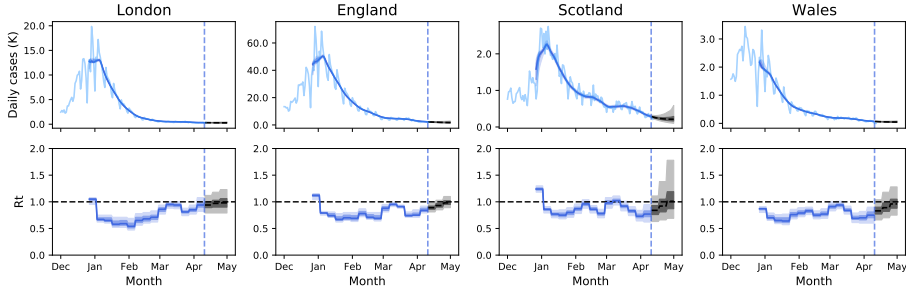


Fig. 4. Regional estimates of cases and R_t . Our model inferences are plotted in dark blue on both cases and R_t plots, and additionally for case plots the true cases are plotted in light blue. Our model projections for cases and R_t , as well as 50% & 95% credible intervals, are plotted in grey.

a reliability curve close to $y = x$. We observe that EpiMap’s uncertainty estimates generally capture the underlying case counts distribution well, though with some variation across start dates and model configurations. EpiEstim’s uncertainty estimates are overconfident as indicated by the flatter shaped curves. For the first three start dates, EpiMap (single area) and models with small ρ^{spatial} yield better uncertainty estimates. For 2020-12-21, the concave shape of the reliability curves indicates models are overestimating case counts, which is consistent with the fact that stricter NPIs curbed case counts while the models predicted case counts would increase assuming no changes in spread dynamics. For 2021-01-18, larger ρ^{spatial} perform best, likely because the prevailing *national* lockdown in that period meant that spread dynamics were more uniform across areas. Additional results are in Appendix C.2, including loss and reliability curves stratified by week during the 3-week prediction period and individual LTLA losses.

4.3. Regional estimates

While our model operates at the level of local authorities, we can estimate R_t ’s at coarser spatial scales by aggregating inferences across multiple local areas. Figure 4 shows the inferences produced by the full EpiMap model with spatial length scale of 20km for London, England, Scotland and Wales, using data available on 10th April 2021. Corresponding plots for other English NHS regions can be found in Appendix C.3.

Figure 4 shows sensible credible intervals both during the modelled 15-week time period and subsequent 3-week forecasts. In this example,

we see that our model projects an increasingly uncertain size of epidemic in Scotland in the near future, with a non-negligible probability of R_t being above 1 in Scotland and London on 10th April 2021, whereas other regions are projected to have stable or shrinking epidemics.

5. Discussion

We have proposed a hierarchical Bayesian approach to model epidemics at fine spatial scales, which incorporates movement of populations across local areas as well as spatiotemporal sharing of statistical strength. Empirical results suggest that our model, along with others, can be a useful tool for policy makers to locate future epidemic hotspots early, in order to direct resources such as surge testing as well as targeted local transmission reduction measures.

As with other methods that infer the extent of epidemics through identified cases alone, the main limitations of this work are due to the provenance of the Pillars 1+2 case data. Firstly, there can be substantial selection bias in the population who get tested, leading to discrepancies between reported cases and the true size of the epidemic. In addition, the amount of testing may change over time, e.g. due to localized testing or limited supplies of testing kits, potentially leading to spurious temporal patterns (Omori et al., 2020). Finally, case data are only reported for the combined Pillars 1 and 2 of the UK’s testing regime. These correspond to different sectors of society at different points of an infection, with different delay distributions between infection and getting tested. Moreover, the proportion of tests under each pillar has been changing systematically since Pillar 2 testing began.

Our model is the result of a number of modelling choices, and can be improved in a number of ways. Firstly, our aim is to track local reproduction numbers and provide nowcasting of epidemic development in local areas, rather than understanding how non-pharmaceutical interventions (NPIs) affect transmission rates. This lead to our choice of a nonparametric Gaussian process prior for the reproduction numbers, rather than a generalised linear model relating transmission rates to NPIs. It is possible to extend our model to model effect of NPIs as in Flaxman et al. (2020). It also lead to our choice not to explicitly model the susceptible population, since it impacts the model just via lowered transmission rates.

Secondly, our model uses only Pillars 1+2 case data, which as noted above have biases that are not well understood and affects our confidence

in the inferred local transmission rates and forecasts. Further, in our model we assumed that positive test cases correspond 1-1 to infections, which in fact does not hold due to asymptomatic infections. We can correct for these biases by incorporating less biased data like hospitalisation and death counts, as well as less granular but better understood estimates of prevalence data obtained from randomised surveys such as REACT (Riley et al., 2020) and the ONS infection survey (Pouwels et al., 2020).

Finally, with the increasing importance of the roles of vaccines and variants, it is interesting to consider how these can be incorporated into our model. This will require a number of extensions, including separating the population into age bands and modelling the susceptible population. These extensions will incur significantly higher computational costs, and additional work will have to be performed on the side of software and implementational efficiency.

Our hierarchical Bayesian model is sensitive to a number of hyperparameters, particularly those specifying the generation interval and incubation period distributions, and the spatial and temporal length scales of the latent Gaussian process. These are hard to specify in a fully Bayesian manner. For example, the posterior strongly prefers spatial length scales that are too long due to model misspecification. Until there are good fully Bayesian approaches to dealing with such situations, we have kept to a more pragmatic approach of using cut models and cross validation.

Our hierarchical model introduces stochasticity at all three layers of the model to capture different aspects of the unfolding epidemic. As a reviewer noted, there can be complex interplays between these layers, for example resulting in non-identifiable parameters. The various components of the model have been chosen to avoid the worse of these, but we have not performed a systematic study of the impacts of these choices. This will be an illuminating piece of future research.

Acknowledgements

This project was incubated at the Royal Society DELVE Initiative and has benefited greatly from feedback from many DELVE contributors, including: Sylvia Richardson, Frank Kelly, Aziz Shiekh, Bryan Grenfell, Guy Harling, Nigel Field and Neil Lawrence. We also gratefully acknowledge Barry Rowlingson and Chris Jewell for their help in data wrangling for the UK Census 2011 commuter flow data, as well as the reviewers for the insightful and helpful feedback.

Web Links

¹<https://www.gov.uk/government/publications/coronavirus-covid-19-scaling-up-testing-programmes>

²<https://coronavirus.data.gov.uk>

³<https://publichealthscotland.scot/our-areas-of-work/sharing-our-data-and-intelligence/coronavirus-covid-19-data-and-guidance/>

⁴<https://phw.nhs.wales/topics/latest-information-on-novel-coronavirus-covid-19/>

⁵https://github.com/rs-delve/covid19_datasets

⁶<https://census.ukdataservice.ac.uk/use-data/guides/flow-data.aspx>

References

Abbott, S. et al. (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research*, 5(112):112.

Bhatt, S., Ferguson, N., Flaxman, S., Gandy, A., Mishra, S., and Scott, J. A. (2020). Semi-mechanistic Bayesian modeling of COVID-19 with renewal processes. *arXiv preprint arXiv:2012.00394*.

Bhoopchand, A., Paleyes, A., Donkers, K., Tomasev, N., and Paquet, U. (2020). DELVE global COVID-19 dataset. https://github.com/rs-delve/covid19_datasets.

Bi, Q. et al. (2020). Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases*, 20:911–919.

Carmona, C. and Nicholls, G. (2020). Semi-modular inference: enhanced learning in multi-modular models by tempering the influence of components. In Chiappa, S. and Calandra, R., editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 4226–4235. PMLR.

Carpenter, B. et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32.

Cori, A., Ferguson, N. M., Fraser, C., and Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology*, 178(9):1505–1512.

- Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, pages 1–8.
- Flaxman, S., Wilson, A., Neill, D., Nickisch, H., and Smola, A. (2015). Fast Kronecker inference in Gaussian processes with non-Gaussian likelihoods. In *Proceedings of ICML*. PMLR.
- Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic. *PLOS ONE*, 2(8):1–12.
- Gostic, K. M. et al. (2020). Practical considerations for measuring the effective reproductive number, R_t . *medRxiv*.
- Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(47):1593–1623.
- Jacob, P. E., Murray, L. M., Holmes, C. C., and Robert, C. P. (2017). Better together? statistical learning in models made of modules.
- Kingma, D. and Welling, M. (2014). Auto-encoding variational Bayes. In *ICLR 2014*.
- Omori, R., Mizumoto, K., and Chowell, G. (2020). Changes in testing rates could mask the novel Coronavirus disease (COVID-19) growth rate. *International Journal of Infectious Diseases*, 94.
- Plummer, M. (2015). Cuts in Bayesian graphical models. *Statistics and Computing*, 25:37–43.
- Pouwels, K. B. et al. (2020). Community prevalence of SARS-CoV-2 in England from April to November, 2020: results from the ONS Coronavirus infection survey. *The Lancet Public Health*.
- Riley, S. et al. (2020). Transient dynamics of SARS-CoV-2 as England exited national lockdown. *medRxiv*.
- Saatçi, Y. (2012). *Scalable inference for structured Gaussian process models*. PhD thesis, University of Cambridge.
- Wallinga, J. and Teunis, P. (2004). Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *American Journal of Epidemiology*, 160(6):509–516.

A. Additional Model Variations

In addition to the final model described in the main paper, we have also considered a number of model variations which did not result in improved performance so did not include them.

A.1. Global effects term in GP prior

We have also explored adding an additional global effects term to the spatial part of the kernel in (6):

$$K_{ij}^{\text{space}} = (\sigma^{\text{spatial}})^2 \exp(-\|y_i - y_j\|/\rho^{\text{spatial}}) + (\sigma^{\text{local}})^2 \delta_{ij} + (\sigma^{\text{global}})^2 \quad (11)$$

This has the effect of adding another GP term $f_i^{\text{global}} \sim \text{GP}(0, \sigma^{\text{global}} K^{\text{time}})$ to (5) that is shared across all areas $i = 1, \dots, n$. However this has an effect of over-generalising estimates of $R_{i,t}$ from the high incidence areas (for which the likelihoods constrain inference of $R_{i,t}$ sufficiently) to the low incidence areas (for which they do not).

A.2. Modelling infectiousness and susceptibility separately

We have also explored a somewhat more elaborate metapopulation model. Note that in (3)-(4) the number of transmissions occurring in an area i depends only on $R_{i,t}$ and not on $R_{j,t}$ of the areas j that are “sending” infections to area i . We can extend this to a model where the predicted mean count depends on properties of both the area that “receives” an infection and the area that “sends” it:

$$C_{i,t} | C_{i,1:t-1} \sim \text{NegBin}(\mu_{i,t}, \psi) \quad \mu_{i,t} = \sum_{j=1}^n R_{i,t} F_{ji} R'_{j,t} Z_{j,t} \quad (12)$$

where $R'_{j,t}$ can be interpreted as an infectiousness level of area j , and $R_{i,t}$ a susceptibility of area i , with the overall transmission rate being a function of both, as well as of the fluxes. While this extension is more complex and flexible, it is not clear whether both the infectiousness and susceptibilities are well-identified from case count data. Empirically, we have not found it to perform differently from the simpler metapopulation model (3)-(4). We used the same GP prior for both the infectivities $R'_{i,t}$ and susceptibilities $R_{i,t}$ in these experiments. As a result of the lack of statistical gains and of computational costs, we decided to use the simpler model (3)-(4).

B. Computational efficiency considerations

B.1. Kronecker structured GP kernel

The Kronecker structure of the GP kernel allows for efficient computations (Saatçi, 2012; Flaxman et al., 2015). In particular, we never have to explicitly form or factorise $K^{\text{space}} \otimes K^{\text{time}}$, which would have computational cost of $O((nm)^3)$. Instead, if $f_{:,i}$ is represented as a $n \times m$ matrix, a draw from its GP prior can be expressed as:

$$f_{:,i} = L^{\text{space}} E(L^{\text{time}})^\top \quad (13)$$

where L^{space} and L^{time} are Cholesky factors of K^{space} and K^{time} respectively and E is an $n \times m$ matrix with iid standard normal distributed entries. The computational cost of this procedure is $O((n^2 + m^2)(n + m))$, which represents significant computational savings over $O((nm)^3)$.

B.2. Positivised Gaussian approximation for infection model

We used a negative binomial distribution for the number of new infections on each day given infections in past days (4). This is a discrete distribution and makes posterior simulation, particular with the Stan probabilistic programming system, challenging. Instead we considered a simple approximation of the negative binomial distribution using a positivised Gaussian distribution with matched mean and variance. Specifically, if $Y \sim \text{NegBin}(\mu, \phi)$, we approximate $Y \approx |\tilde{Y}|$, where $\tilde{Y} \sim \mathcal{N}(\mu, \sqrt{(1 + \phi)\mu})$. Note that $|\tilde{Y}|$ has mean higher than μ and variance lower than $(1 + \phi)\mu$, but for $\mu \geq 5$ the difference is practically negligible. However in cases where $\mu \leq 10$ this can lead to under-estimation of R , but we believe this is not a serious concern since the epidemic would then be of very small size anyway.

We chose this approximation as the computation for the infection model can be “reparameterised” (Kingma and Welling, 2014) using the so-called “non-centred” parameterisation and lead to a better mixing MCMC sampler. Specifically, and assuming no metapopulation model for simplicity, we can write the sampling statements for the positivised Gaussian approximation of (4) as:

$$\tilde{X}_{i,t} = \left| R_{i,t} \tilde{Z}_{i,t} + \eta_{i,t} \sqrt{(1 + \psi) R_{i,t} \tilde{Z}_{i,t}} \right|, \quad \eta_{i,t} \sim \mathcal{N}(0, 1) \text{ iid} \quad (14)$$

Note that the modelled epidemic sizes $\tilde{X}_{i,:}$ can be written as a differentiable and efficiently computed function of a sequence of iid standard normal random variables $\eta_{i,:}$ (and the reproduction numbers). The gradients

can be automatically computed by Stan, and the No-U-Turns Sampler mixes more effectively since while $\tilde{X}_{i,:}$ are highly correlated (which can lead to slow mixing if not reparameterised), the reparameterised random variables $\eta_{i,:}$ are independent a priori (hence faster mixing).

B.3. Regional inference

In order to track the daily evolution of the epidemic in real time it is preferable for the posterior simulations to run overnight. However, the model described in 3 is quite complex, and full posterior simulation for the whole of Great Britain using Markov chain Monte Carlo (MCMC) has significant computational costs. In this section we describe a two stage procedure to reduce the computational costs to a manageable level.

During the first stage the epidemic time courses of individual local areas are approximately inferred first by ignoring cross-area dependencies in both the metapopulation infection model and the GP prior. This first stage can be easily parallelised across the 348 areas and completed quickly.

In the second stage, we split Great Britain into 9 regions (7 NHS regions in England, plus Wales and Scotland), and modelled each region independently using the model described in Section 3. In order to account for transmissions to and from other regions, we fix the latent epidemic process for areas in other regions to the posterior median inferred during the first stage. To reduce the approximation error due to only modelling each region rather than the whole of Great Britain, we include in each region model a number of areas outside the region, such that for all areas within that region at least 80% of the off-diagonal flux probabilities (corresponding rows in F^{fwd} and F^{rev}) are included in the model.

C. Additional Figures

C.1. Simulation data

Figures 5 and 6 show case count and R_t predictions for all models and variants of EpiMap.

C.2. Predicting future case counts

Figures 7 and 8 append the results in Figures 2 and 3 respectively, showing losses and uncertainty calibration stratified by week during the 3-week prediction period. Figure 9 shows the $\log(\text{RMSE} + 1)$ for individual LTLAs which are stratified by week and compared between models.

C.3. Regional estimates

Figure 10 shows the regional estimates for cases and R_t on remaining NHS regions in England, in the same setting as Figure 4.

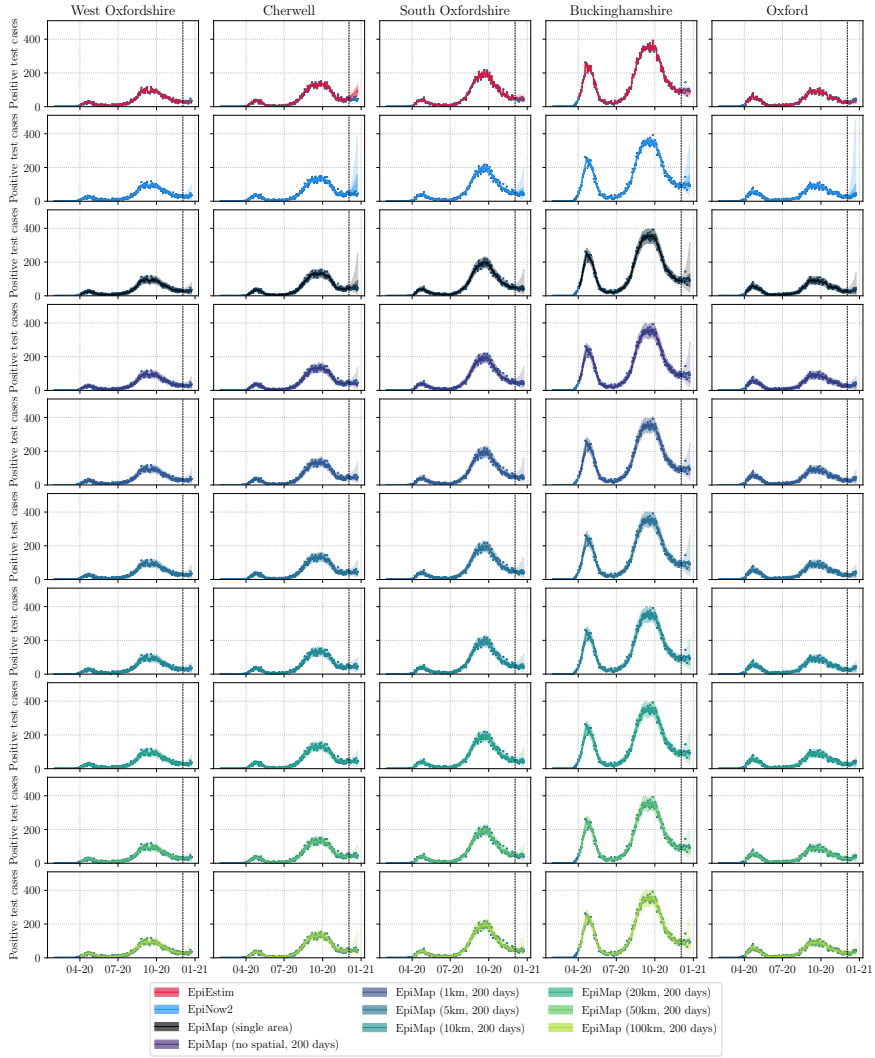


Fig. 5. Estimated median, 50% (inner) and 95% (outer) credible intervals of the posterior predictive distributions for daily case counts plotted against the observed case counts used to infer R_t . Values to the left of the vertical line are inferred from data, and those to the right are future predictions. For EpiMap, we include the weekly variation in expected cases in the model, but plot the distribution without this variation for clarity. The EpiMap methods report the full distribution of expected cases. EpiNow2 returns the distribution of the mean number of cases. EpiEstim does not provide estimated case distributions, so future predictions are stochastic rollouts of the epidemic based on the last inferred R_t distribution.

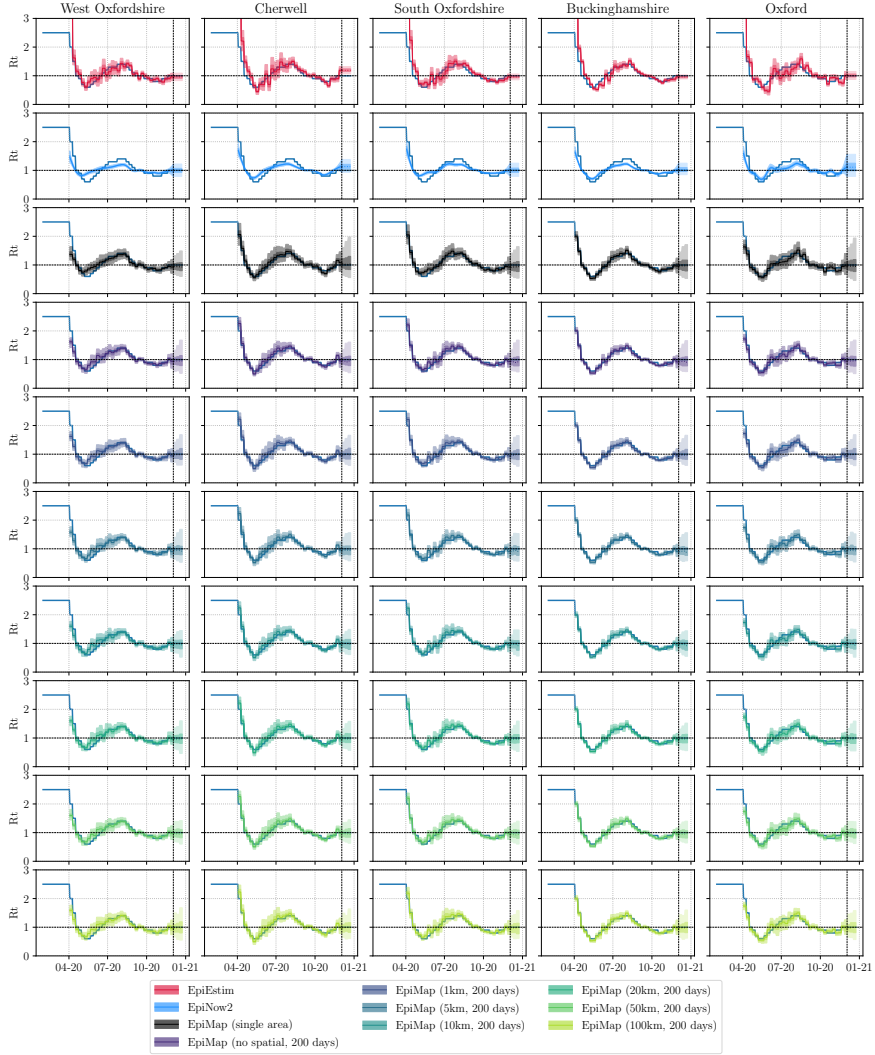


Fig. 6. Estimated median, 50% (inner) and 95% (outer) credible intervals of R_t for the methods plotted against the R_t used to create the simulated epidemic. EpiEstim does not provide future estimates of R_t , and so the final R_t posterior is used as a prediction.

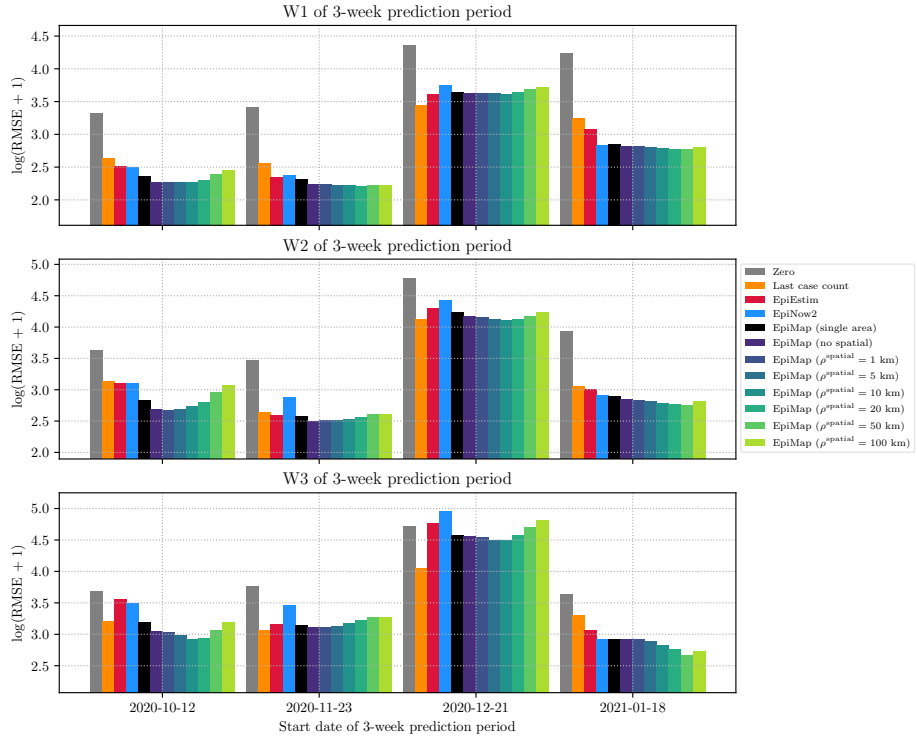


Fig. 7. A comparison of predictive performance similar to Figure 2 but stratified by week over the 3-week prediction period. As expected, the predictions made for later weeks, such as W3, are worse than those made for earlier weeks, such as W1, across models. The relative ordering of EpiMap variants typically remains unchanged for different weeks.

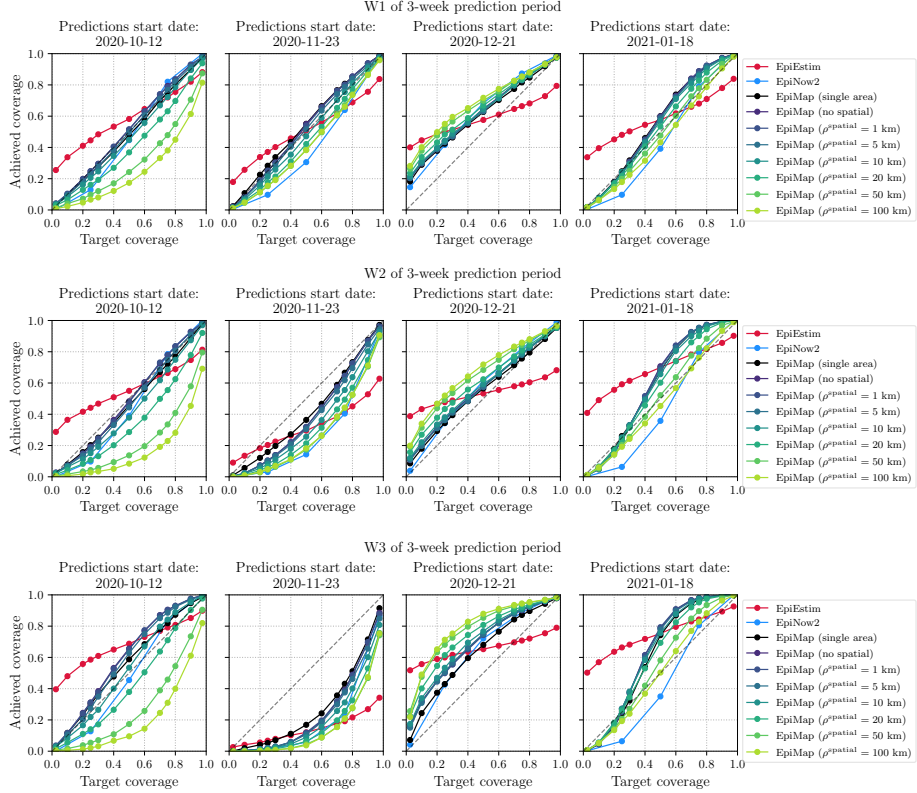


Fig. 8. An evaluation of the uncertainty estimates produced by methods similar to Figure 2 but stratified by week over the 3-week prediction period. As expected, the quality of uncertainty estimates degrades in the later weeks compared to earlier weeks, as indicated by reliability curves that are further from the ideal diagonal. Once again, the relative ordering of EpiMap variants typically remains unchanged for different weeks, however differences in uncertainty calibration between models tend to exacerbate.

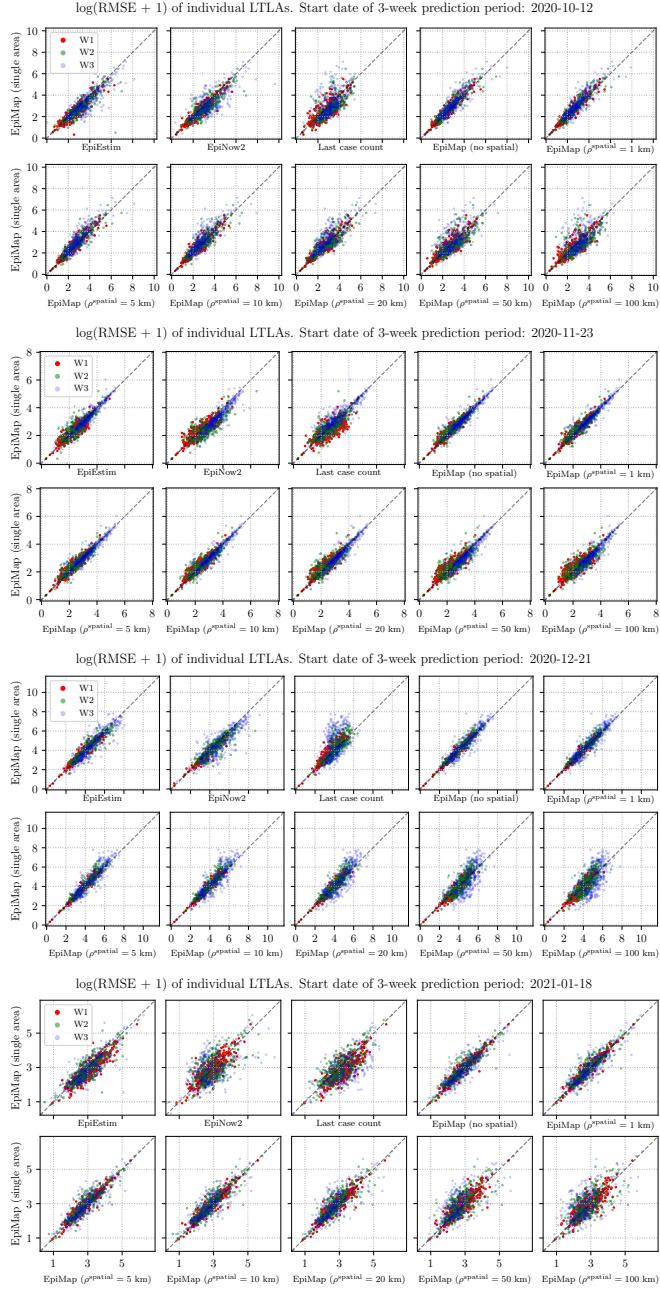


Fig. 9. We plot the $\log(\text{RMSE} + 1)$ for individual LTLAs (each dot is an LTLA) stratified by week. We observe variation in predictive performance for different LTLAs for all models, with large correlation in LTLA losses between methods.

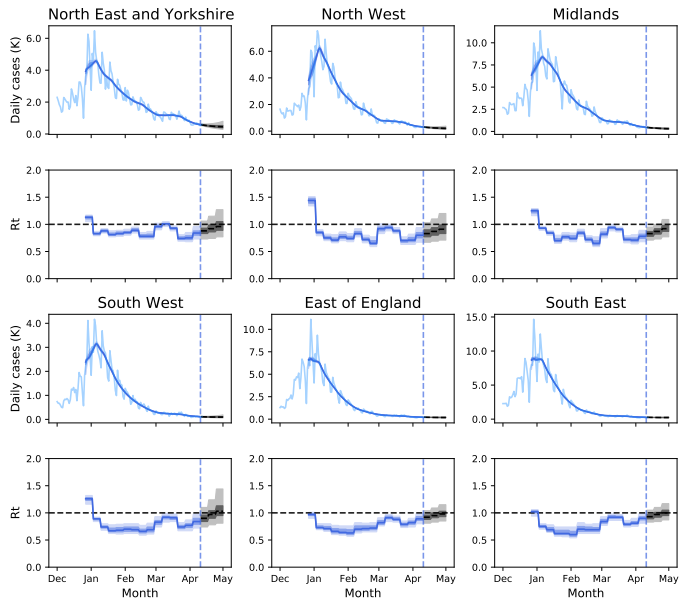


Fig. 10. Additional regional estimates of cases and R_t for NHS regions in England, using case data available by the 10th April 2021.