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, Avishkar Bhoopchand, Peter Diggle, Bryn Elesedy, Bobby He, Michael Hutchinson, Ulrich Paquet, Jonathan Read, Nenad Tomasev, Sheheryar Zaidi

April 19 2021

1 Introduction

The spatio-temporal pattern of Covid-19 infections, as for most infections disease epidemics, is highly heterogenous 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 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 incidence and 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 of our model 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 Method

Our method builds on work by Cori et al. (2013) and Flaxman et al. (2020) that estimate instantaneous reproduction numbers using a renewal process approach. Following the approach by 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 also introduce a number of extensions 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. We choose a Kronecker factored GP kernel which can be used to enable efficient computation for posterior simulation. 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 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.

2.1 Data

We applied our model 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 strategy[1]. 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 England[2], Scotland[3] and Wales[4] and we access them through the DELVE Global Covid-19 Dataset[5] (Bhoopchand et al., 2020). Due to delays in processing tests, we ignore the last 7 days of case counts.

2.2 Model

We estimate the instantaneous reproduction numbers, $R_{i,t}$, across local areas in the UK (indexed by i) and across time (indexed by t) using a Bayesian semi-mechanistic approach, so as to capture the uncertainties inherent in the estimation procedure, and incorporate prior knowledge about the process through a mechanistic description of epidemic spread via the renewal equation and through prior distributions for the model parameters. That is, we model the available data using a generative model, with observed daily case counts $C_{i,t}$, unobserved daily infection counts $X_{i,t}$ and unobserved reproduction numbers $R_{i,t}$, for each local area i and day t. All quantities are modelled as random variables in the generative model, and the posterior distribution over the unobserved random variables ($R_{i,t}$, $X_{i,t}$ and other random variables) are estimated using Markov chain Monte Carlo simu.

2.2.1 The Renewal Process Approach

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 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 when estimating effective reproduction numbers.

[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/

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

In this section we will briefly describe the approaches of Cori et al. (2013) and Flaxman et al. (2020) to serve as bases on which we build our model. We will assume knowledge of the generation distribuion W, with W_s being the probability that a secondary infection happens s days after the primary infection.

Cori et al. (2013) assumes that case counts correspond exactly to infection counts, $C_{i,t} = X_{i,t}$, that is, all infections along with the day of infection are ascertained. For each day t, the observed *infection* count $C_{i,t}$ is modelled as Poisson distributed, conditional on previous counts and the local $R_{i,t}$,

$$C_{i,t}|C_{i,1:t-1} \sim \text{Poisson}(R_{i,t}Z_{i,t}) \qquad Z_{i,t} = \sum_{s=1}^{t} C_{i,t-s}W_s \tag{1}$$

where $R_{i,t}$ controls the overall expected growth/reduction rate of the epidemic in local area i and day t, and $Z_{i,t}$ is the infection load due to previous infections in the area.

Cori et al. (2013) assume that $R_{i,t}$ has a Gamma distributed prior, which is conjugate to the Poisson likelihood and leads to a Gamma distributed posterior that can be computed directly and efficiently. However the resulting posterior uncertainty can be quite large, particularly with low counts expected in small local areas. This was addressed by assuming that $R_{i,t}$ stays constant over a number of days and using the resulting combined likelihood, which contains multiple terms (one for each day) to reduce the posterior uncertainty. In our model we do the same and assume $R_{i,t}$ is constant in each weekly period.

The generation distribution (time between primary and secondary infections) is typically hard to obtain and Cori et al. (2013) suggest approximating this with the serial interval distribution instead (time between symptom onsets for primary and secondary cases), and showed that if infectiousness only starts after symptom onset, the serial interval and generation interval are the same. Following Flaxman et al. (2020), we use the estimate for Covid-19 from Bi et al. (2020), as a Gamma distribution with a mean of 6.3 days. Cori et al. (2013) also considered uncertainty in the generation interval distribution by bootstrapping/ensembling over parameters of this distribution, but this is not considered in Flaxman et al. (2020) nor in our model, and is likely a source of under-estimatation for our uncertainty in $R_{i,t}$.

A significant limitation of Cori et al. (2013) is that they assume that known cases represent all the infections in the epidemic. While this might be reasonably true in some other epidemics, it is definitely is not true for Covid-19, with asymptomatic and presymptomatic cases accounting for a substantial proportion of all infections. The estimates of the proportion of asymptomatic patients across studies vary greatly, for example 19% Wells et al. (2020) in SE England and 42.5% Lavezzo et al. (2020) in Italy. Additionally, there may be substantial biases in the testing process, with respect to subpopulations more likely to be tested, biasing the estimates of symptomatic Covid-19 cases. Cori et al. (2013) noted that uniform under-reporting of infections does not lead to significant bias in estimates of $R_{i,t}$, though it increases variance and uncertainty, which is to be expected.

Flaxman et al. (2020) separate out the (unobserved) evolution of the epidemic (counts of infections $X_{i,t}$) from the (observed but more noisy) data consisting of death counts, to model the impact of various NPIs in European countries. The latent epidemic process is modelled using a deterministic renewal equation:

$$X_{i,t} = R_{i,t}Z_{i,t} \qquad Z_{i,t} = \sum_{s=1}^{t} X_{i,t-s}W_s \tag{2}$$

while observed counts of deaths are modelled using a case fatality rate, a delay distribution and an over-dispersed negative binomial observation model. The separation of the modelling into a latent epidemic process and an observation process addresses one of the major concerns of Cori et al. (2013), and is advantageous as it is less sensitive to observation noise and gives more flexibility in modelling different forms of data. For example, the R software package epidemia[6] allows for jointly modelling of both cases and deaths. Bhatt et al. (2020) describes a broader framework of models which allows for stochasticity in the epidemic process as well as fusion case counts and hospitalisations observations. Posterior inference requires MCMC sampling which is slower than the conjugate posterior of Cori et al. (2013).

[6] `https://github.com/ImperialCollegeLondon/epidemia`

2.2.2 Model Overview

In this section we describe in detail our model, which we refer to as EpiMap. Our model extends works by Cori et al. (2013); Flaxman et al. (2020) in two directions: a metapopulation model accounting for travel and infections across local areas, and a spatiotemporal prior for the reproduction numbers to implement sharing of statistical strength across the estimation of the reproduction numbers $R_{i,t}$. Recall that $X_{i,t}$ and $C_{i,t}$ are respectively the unobserved number of infections and observed number of reported cases with positive PCR swab tests on day t in local area i, and $R_{i,t}$ is the corresponding instantaneous reproduction number.

Following Flaxman et al. (2020); Bhatt et al. (2020) we model the number of reported cases using a delay distribution and an over-dispersed negative binomial observation model:

$$C_{i,t}|X_{i,1:t},\phi_i \sim \text{NegBin}(V_{\text{day\_of\_week}(t)}E_{i,t},\phi_i) \qquad E_{i,t} = \sum_{s=1}^{t} X_{i,t-s}D_s \tag{3}$$

where D_s is the probability that an infected person 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. Section 2.2.4 gives more details. We model weekly variations in reported cases using $V_{\text{day\_of\_week}(t)}$.

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

$$Z_{i,t} = \sum_{s=1}^{t} X_{i,t-s}W_s \tag{4}$$

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. 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:

$$\tilde{Z}_{i,t} = \sum_{j=1}^{n} F_{ji}^{(t)} Z_{j,t} \tag{5}$$

We describe the metapopulation model in further detail in Section 2.2.5, including how the flux matrices are parameterised. We model the number of new infections on day t as,

$$X_{i,t}|R_{i,t},X_{1:n,1:t-1} \sim \text{NegBin}(R_{i,t}\tilde{Z}_{i,t},\psi) \tag{6}$$

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

Finally, in order to share statistical strength in the estimation of $R_{i,t}$ across local areas and across time, we model

$$R_{i,t} = \exp(f_{i,t}) \tag{7}$$

where the log reproduction numbers $f_{i,t}$ across local areas and days are modelled jointly as a Gaussian process (described in Section 2.2.3).

The overall model consists of three layers: a GP prior 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.

2.2.3 Gaussian Process Prior

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; specifically, we model $R_{i,t} = \exp(f_{i,t})$, with $f_{i,t}$ modelled using a GP. We assume that the instantaneous reproduction numbers are constant within each week, so the time indices t here correspond to weeks. We assume there are m weeks being modelled.

The prior covariance structure of the GP is given by a kernel; we use a spatiotemporal kernel with a Kronecker product structure:

$$\begin{aligned} f_{:,:} &\sim \mathcal{N}(0, K^{\text{space}} \otimes K^{\text{time}}) & K_{ij}^{\text{space}} &= K_{ij}^{\text{spatial}} + K_{ij}^{\text{local}} \\ \text{Cov}(f_{i,s}, f_{j,t}) &= K_{ij}^{\text{space}} \times K_{st}^{\text{time}} & K_{ij}^{\text{spatial}} &= (\sigma^{\text{spatial}})^2 \exp(-\|y_i - y_j\|/\rho^{\text{spatial}}) \\ K_{st}^{\text{time}} &= \exp(-\|s-t\|/\rho^{\text{time}}) & K_{ij}^{\text{local}} &= (\sigma^{\text{local}})^2 \delta_{ij} \end{aligned} \tag{8}$$

where y_i and y_j are the geographical centres of areas i and j respectively, s and t are weekly time indices, and ρ^{spatial} and ρ^{time} are spatial and temporal length scales for the respective Matern(1/2) kernels. 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 kernel K^{space} also contains a local term, $(\sigma^{\text{local}})^2\delta_{ij}$, which gives additional independent random effects for each local area. The resulting GP can be described as the sum of $n+1$ independent GPs, $f_{:,:} = f_{:,:}^{\text{spatial}} + f_{:,:}^{\text{local}}$, with

$$f_{:,:}^{\text{spatial}} \sim \text{GP}(0, K^{\text{spatial}} \otimes K^{\text{time}}), \qquad f_{i,:}^{\text{local}} \sim \text{GP}(0, (\sigma^{\text{local}})^2 K^{\text{time}}) \text{ iid across } i = 1, \ldots, n \tag{9}$$

The Kronecker structure of the GP kernel allows for efficient computations (see Appendix B.1).

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 and sampled over them. 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 the same undesirable over-generalisation effect as for the use of a global term. We believe this behaviour is due to model misspecification with respect to the length scale parameters. Instead we selected the length scale parameters using cross validation performance of forecasted case counts three weeks into the future.

2.2.4 Observation Model

Following Flaxman et al. (2020) we used an over-dispersed negative binomial observation model (3). We use broad half normal prior for the dispersion parameters, $\phi_i \sim \mathcal{N}_+(0, 5)$ iid. In Stan, the parameterisation as neg_binomial_2, given a mean parameter μ and an inverse-dispersion parameter c, has 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 $(1+\phi)\mu$. The probability mass function is given as:

$$p(x|\mu, \phi) = \binom{x + \mu/\phi - 1}{x} \left(\frac{\phi}{1+\phi}\right)^x \left(\frac{1}{1+\phi}\right)^{\mu/\phi} \tag{10}$$

This parameterisation is the natural one which emphasises the infinite divisibility of the negative binomial, i.e. if $Y_1, \ldots, Y_m$ are independent negative binomial random variables with means $\mu_1, \ldots, \mu_m$ 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 ψ, a sensible choice in cases where we believe counts are sums of independent random events.

The infection-to-test delay distribution is a convolution of two delay distributions: an incubation period distribution, and a symptom-onset-to-test distribution. We used a LogNormal(1.57, 0.65) distribution with

a median of 4.8 days and a 90% confidence interval of (1.64,14.04) days for the incubation period (Bi et al., 2020), and assume an additional two day delay to get tested. Weekly variations are modelled using multiplicative factors in (3), with prior $(V_{\text{Mon}}, \ldots, V_{\text{Sun}})/7 \sim \text{Dirichlet}(1, \ldots, 1)$.

2.2.5 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 (5)-(6).

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 Census[7]. 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:

$$F_{ji}^{\text{id}} = \delta_{ji} \qquad F_{ji}^{\text{fwd}} = \frac{M_{ji}}{\sum_k M_{jk}} \qquad F_{ji}^{\text{rev}} = \frac{M_{ij}}{\sum_k M_{kj}} \tag{11}$$

where $\delta_{ji} = 1$ if $j = i$ and 0 otherwise. The denominator for F_{ji}^{rev} is the daytime population in area j consisting of those that do not commute elsewhere and those that commute in to area j. For both the second and third types the transmission is from area j to area i, and for all three matrices the probabilities sum to one when we sum over each row. We parameterised the overall flux matrix during week t using a convex combination of F^{id}, F^{fwd} and F^{rev},

$$F^{(t)} = \alpha_t F^{\text{id}} + (1 - \alpha_t)(\beta F^{\text{fwd}} + (1 - \beta)F^{\text{rev}}) \tag{12}$$

with $\alpha_t \in (0, 1)$ governing the amount of mixing across areas on week t, and $\beta \in (0, 1)$ governing the amount of home-to-work vs work-to-home transmissions. We use a uniform prior over β and a weekly AR(1) prior for α_t.

3 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.

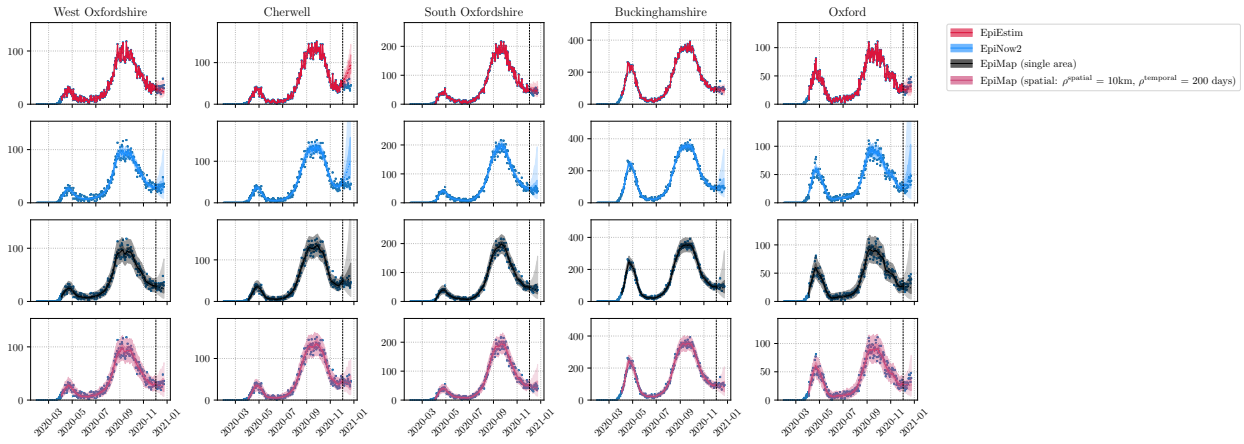

Figure 1: Estimated median, 50% (inner) and 95% (outer) credible intervals of the daily case counts plotted against the observed case counts used to infer R_t. Values to the left of the vertical line are inferred from data, and those to the right are future predictions. For EpiMap, we include the weekly 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.

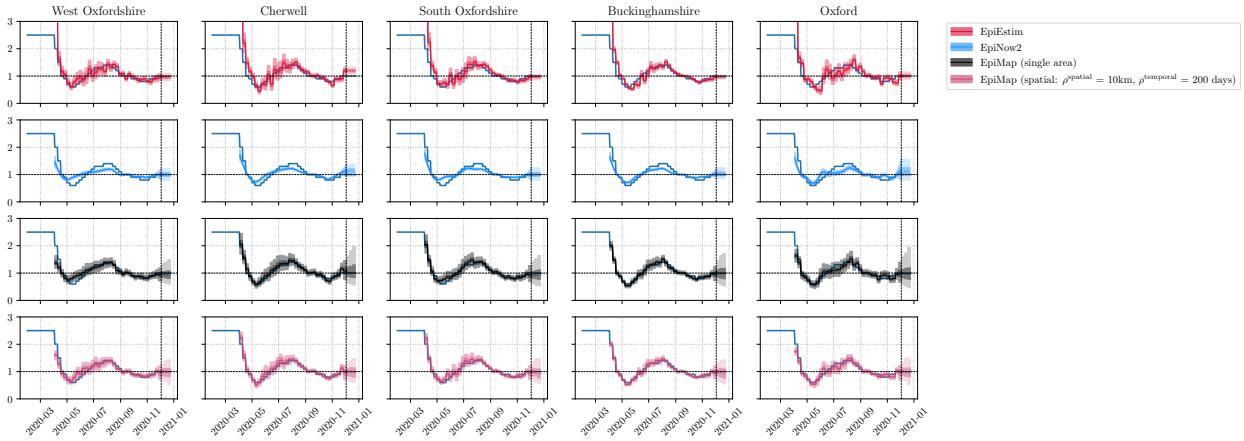

Figure 2: Estimated median, 50% (inner) and 95% (outer) credible intervals of R_t for the methods plotted against the R_t used to create the simulated epidemic. EpiEstim does not provide future estimates of R_t, and so the final R_t posterior is used as a prediction.

3.1 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 intial 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

[7] https://census.ukdataservice.ac.uk/use-data/guides/flow-data.aspx

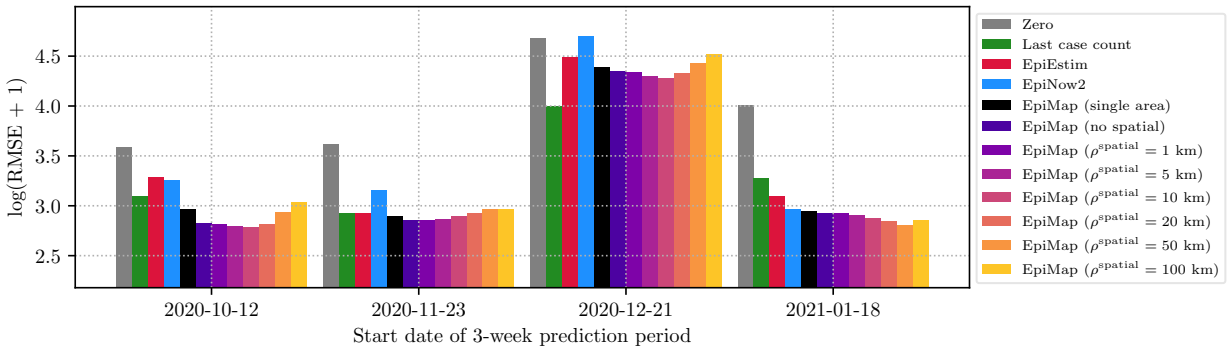

Figure 3: A comparison of models for predicting future case counts over a 3-week unseen period. Each model is fitted on 15 weeks of data and makes predictions for the following 3 weeks.

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 Figures 1 and 2. Plots showing the full sweep of length scales for EpiMap can be found in Appendix Appendix C.1.

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 predctions which appear oversmoothed. 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 3.2 produces more quantitative comparisons of the methods.

3.2 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 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 3 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

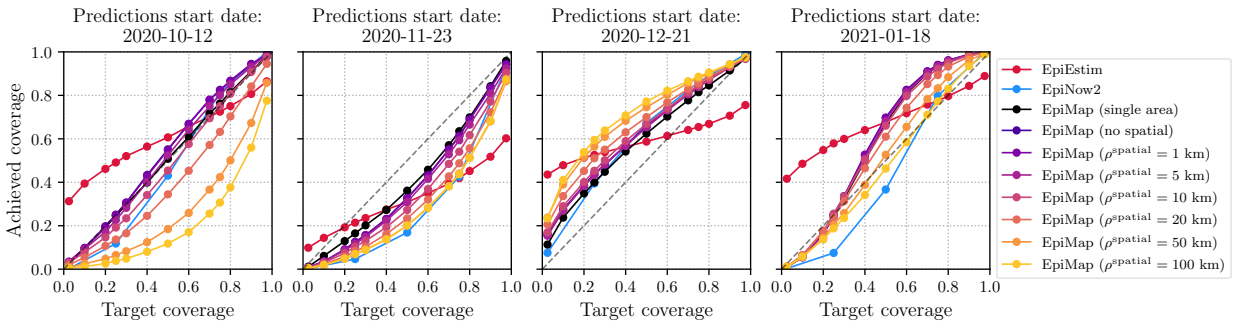

Figure 4: Reliability curves assessing the uncertainty estimates produced by models. Each model yields percentiles of its case count posterior predictive distribution. The curves show the portion $\hat{p}$ of predictions (across dates and LTLAs) for which the true case count is less than the p-th percentile $\hat{c}_p$ of the model (y-axis) vs. p (x-axis).

positive impact of modelling cross-area dependencies is observed, since EpiMap (single area) tends to slightly underperform the other variants of EpiMap. Morever, 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 4 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 4) will ideally be equal to the target coverage p (x-axis of Figure 4), yielding 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.

3.3 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 5 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. In order to account for uncertainty in the generation interval and testing delay distribution, we bootstrap over 10 independent runs, where each run has an independent generation interval and delay distribution, parameterised by a gamma and lognormal distribution respectively. We sample each set of bootstrap parameters using the serial interval and incubation period parameter distributions in Bi et al. (2020). We found bootstrapping to be crucial to avoiding overconfidet predictions for R_t estimates, whilst adding little runtime overhead due to parallelisation.

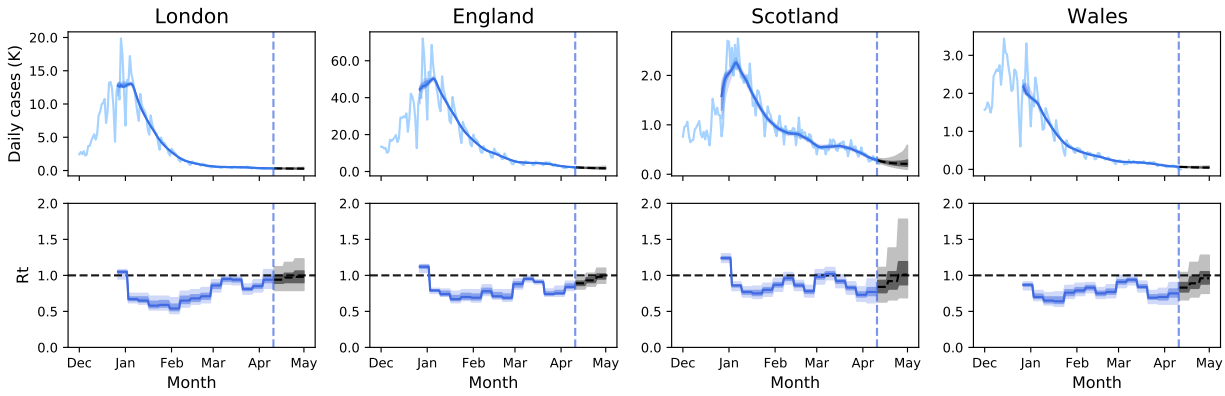

Figure 5: Regional estimates of cases and R_t. Our model inferences are plotted in dark blue on both cases and R_t plots, and additionally for case plots the true cases are plotted in light blue. Our model projections for cases and R_t, as well as 50% & 95% credible intervals, are plotted in grey.

Figure 5 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.

4 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 data. We outline some of the issues here. 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). This can be exacerbated by the impact of false positives under different regimes (Surkova 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 could benefit from incorporating additional data, including hospitalisations, death counts, test positivity rates, whether the tests are carried out in the community or in outbreak investigations, and individual-level demographic data (e.g. age, sex, ethnicity). In future, it would be important to ground the inferences produced by our method by calibrating against unbiased, albeit potentially less granular, estimates of prevalence obtained from randomised surveys such as the REACT study (Riley et al., 2020) and the ONS seroprevalence survey (Pouwels et al., 2020).

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.

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, Rt. *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.

Kingma, D. and Welling, M. (2014). Auto-encoding variational Bayes. In *ICLR 2014*.

Lavezzo, E. et al. (2020). Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature*, 584(7821):425–429.

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.

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.

Surkova, E., Nikolayevskyy, V., and Drobniewski, F. (2020). False-positive COVID-19 results: hidden problems and costs. *The Lancet Respiratory Medicine*.

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.

Wells, P. M. et al. (2020). Estimates of the rate of infection and asymptomatic covid-19 disease in a population sample from se england. *Journal of Infection*.

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.

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 space kernel in (8):

$$K_{ij}^{\text{space}} = (\sigma^{\text{spatial}})^2 \exp(-\|y_i - y_j\|/\rho^{\text{spatial}}) + (\sigma^{\text{local}})^2 \delta_{ij} + (\sigma^{\text{global}})^2 \tag{13}$$

This has the effect of adding another GP term $f_{:}^{\text{global}} \sim \text{GP}(0, \sigma^{\text{global}} K^{\text{time}})$ to (9) that is shared across all areas $i = 1, \ldots, n$. However this has an effect of over-generalising estimates of $R_{i,t}$ from the high incidence areas (for which the likelihooods 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 (5)-(6) 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:

$$C_{i,t}|C_{i,1:t-1} \sim \text{NegBin}(\mu_{i,t}, \psi) \qquad \mu_{i,t} = \sum_{j=1}^{n} R_{i,t} F_{ji} R'_{j,t} Z_{j,t} \tag{14}$$

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 (5)-(6). 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 (5)-(6).

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 $f_{:,:}$ is represented as a $n \times m$ matrix, a draw from its GP prior can be expressed as:

$$f_{:,:} = L^{\text{space}} E (L^{\text{time}})^\top \tag{15}$$

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 (6). 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 (6) as:

$$\tilde{X}_{i,t} = \left| R_{i,t}\tilde{Z}_{i,t} + \eta_{i,t}\sqrt{(1+\psi)R_{i,t}\tilde{Z}_{i,t}} \right|, \qquad \eta_{i,t} \sim \mathcal{N}(0,1) \text{ iid} \tag{16}$$

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 preferrable for the posterior simulations to run overnight. However, the model described in 2.2 is quite complex, and full posterior simulation for the whole of Great Britian 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 2.2. 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 Britian, 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-diagonl flux probabilities (corresponding rows in F^{fwd} and F^{rev}) are included in the model.

C Additional Figures

C.1 Simulation data

Figures 6 and 7 show case count and R_t predictions for all models and variants of EpiMap.

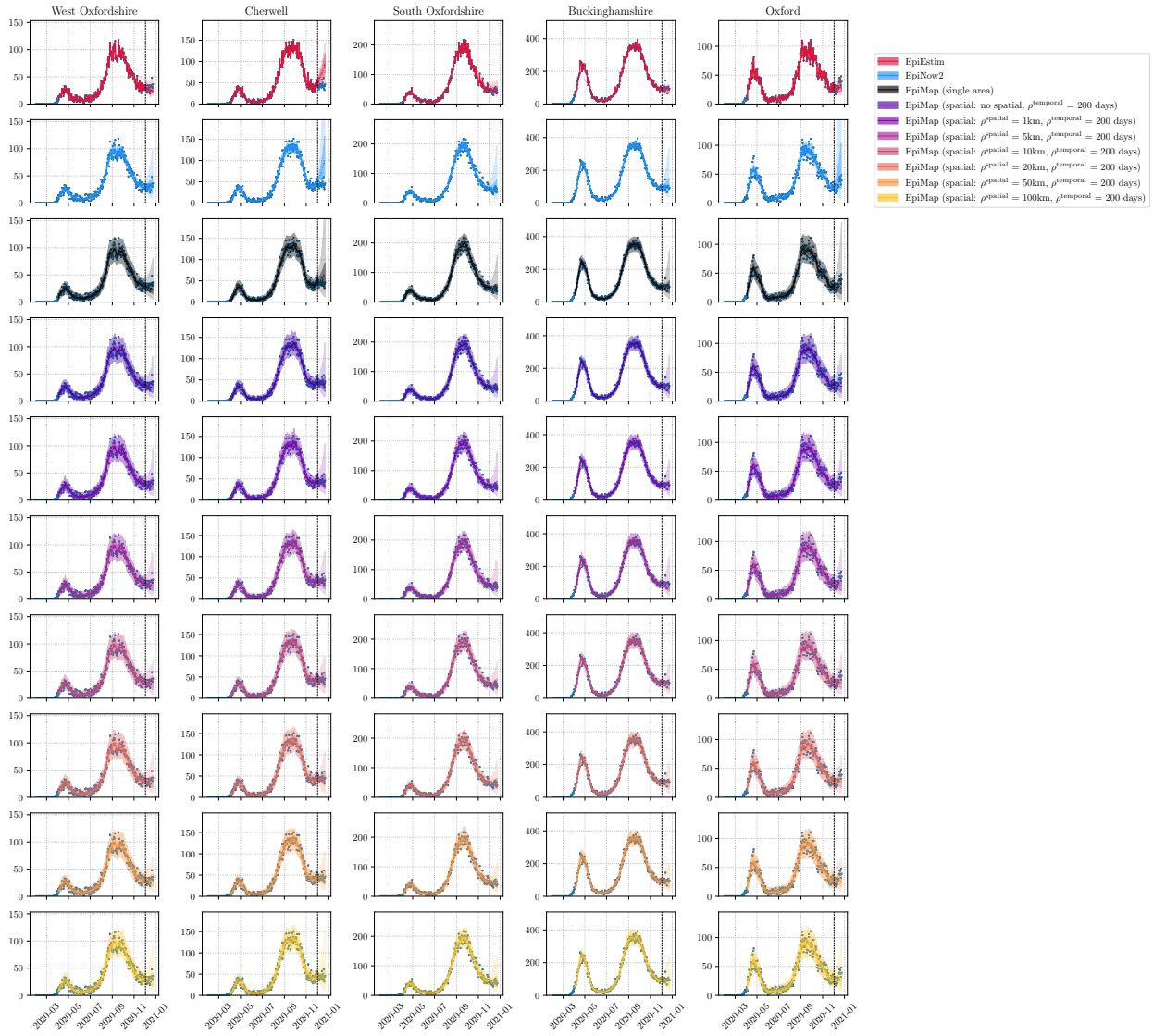

Figure 6: Estimated median, 50% (inner) and 95% (outer) credible intervals of the daily case counts plotted against the observed case counts used to infer R_t. Values to the left of the vertical line are inferred from data, and those to the right are future predictions. For EpiMap, we include the weekly 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.

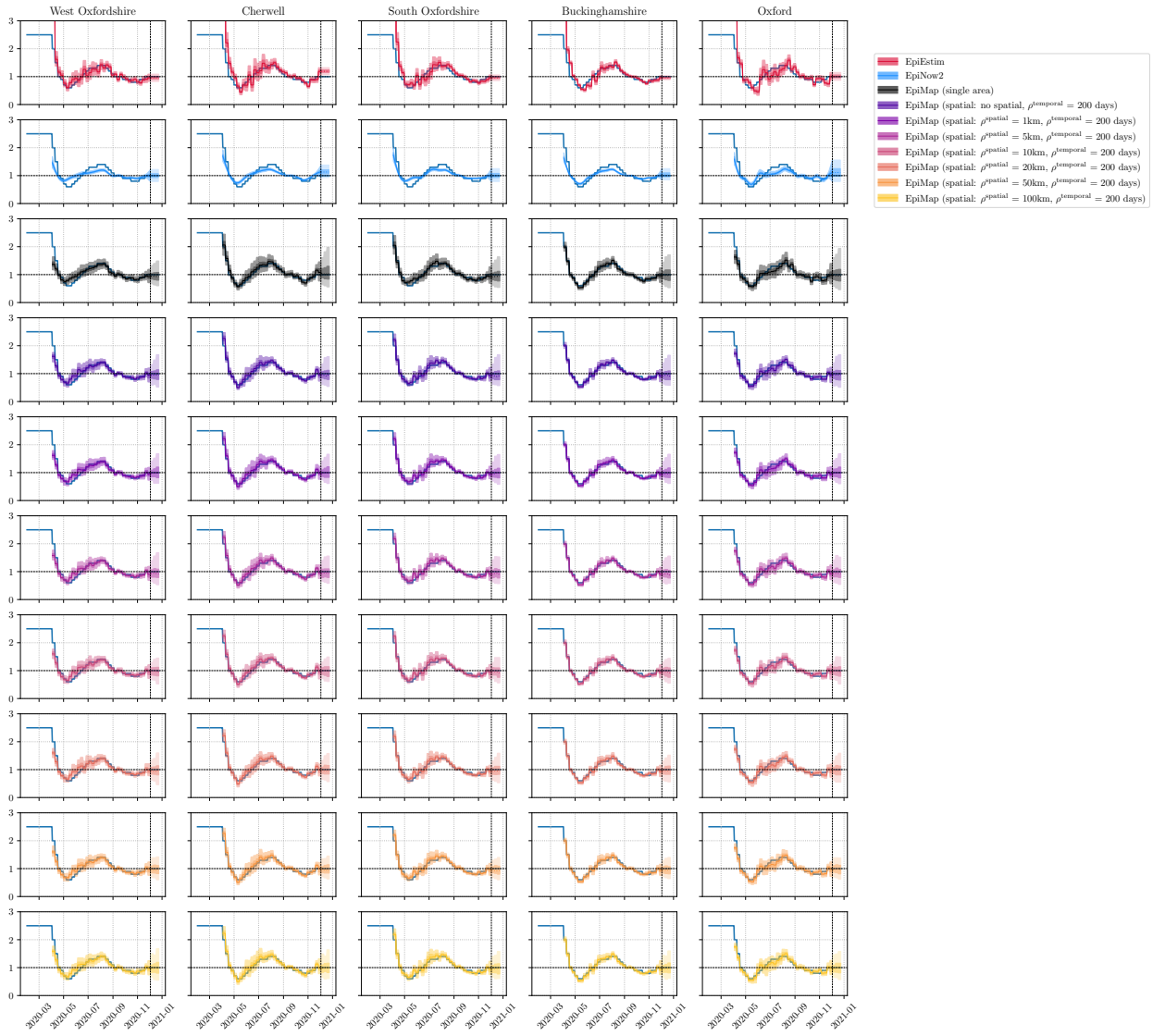

Figure 7: Estimated median, 50% (inner) and 95% (outer) credible intervals of R_t for the methods plotted against the R_t used to create the simulated epidemic. EpiEstim does not provide future estimates of R_t, and so the final R_t posterior is used as a prediction.

C.2 Predicting future case counts

Figures 8 and 9 append the results in Figures 3 and 4 respectively, showing losses and uncertainty calibration stratified by week during the 3-week prediction period. Figure 10 shows the log(RMSE + 1) for individual LTLAs which are stratified by week and compared between models.

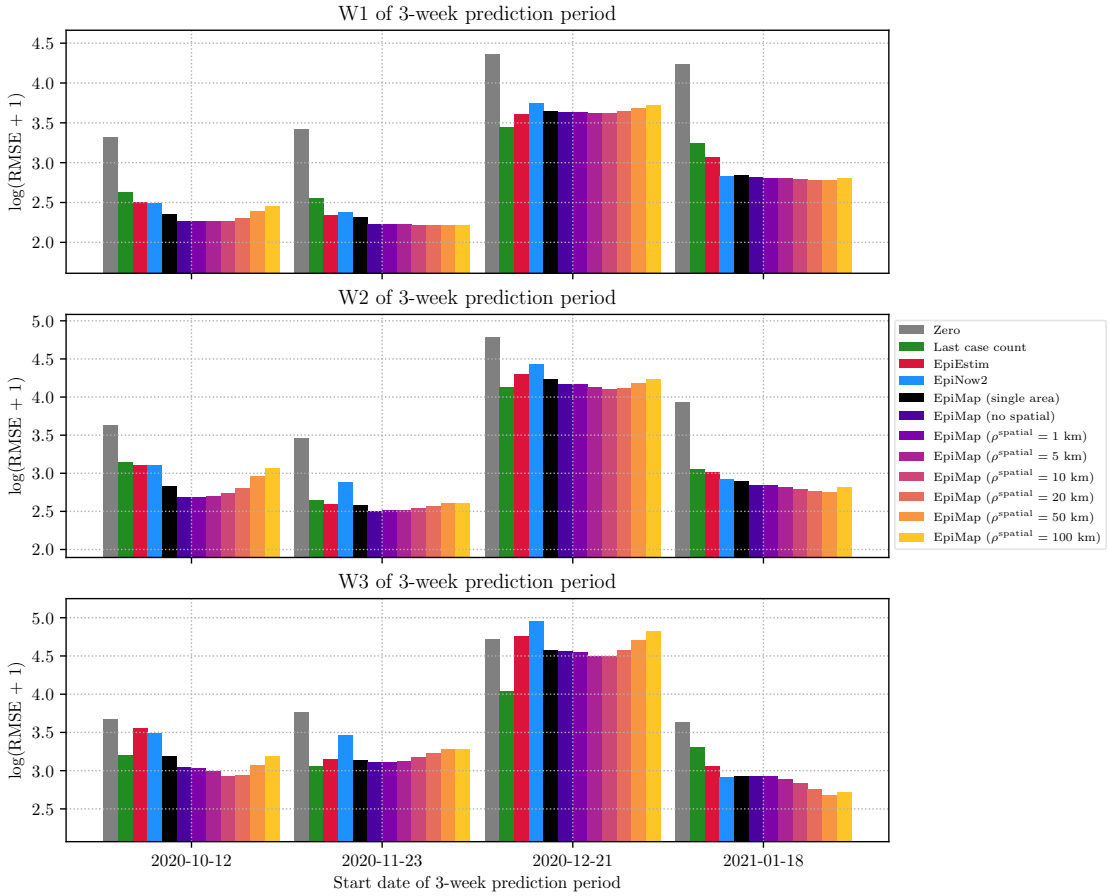

Figure 8: A comparison of predictive performance similar to Figure 3 but stratified by week over the 3-week prediction period. As expected, the predictions made for later weeks, such as W3, are worse than those made for earlier weeks, such as W1, across models. The relative ordering of EpiMap variants typically remains unchanged for different weeks.

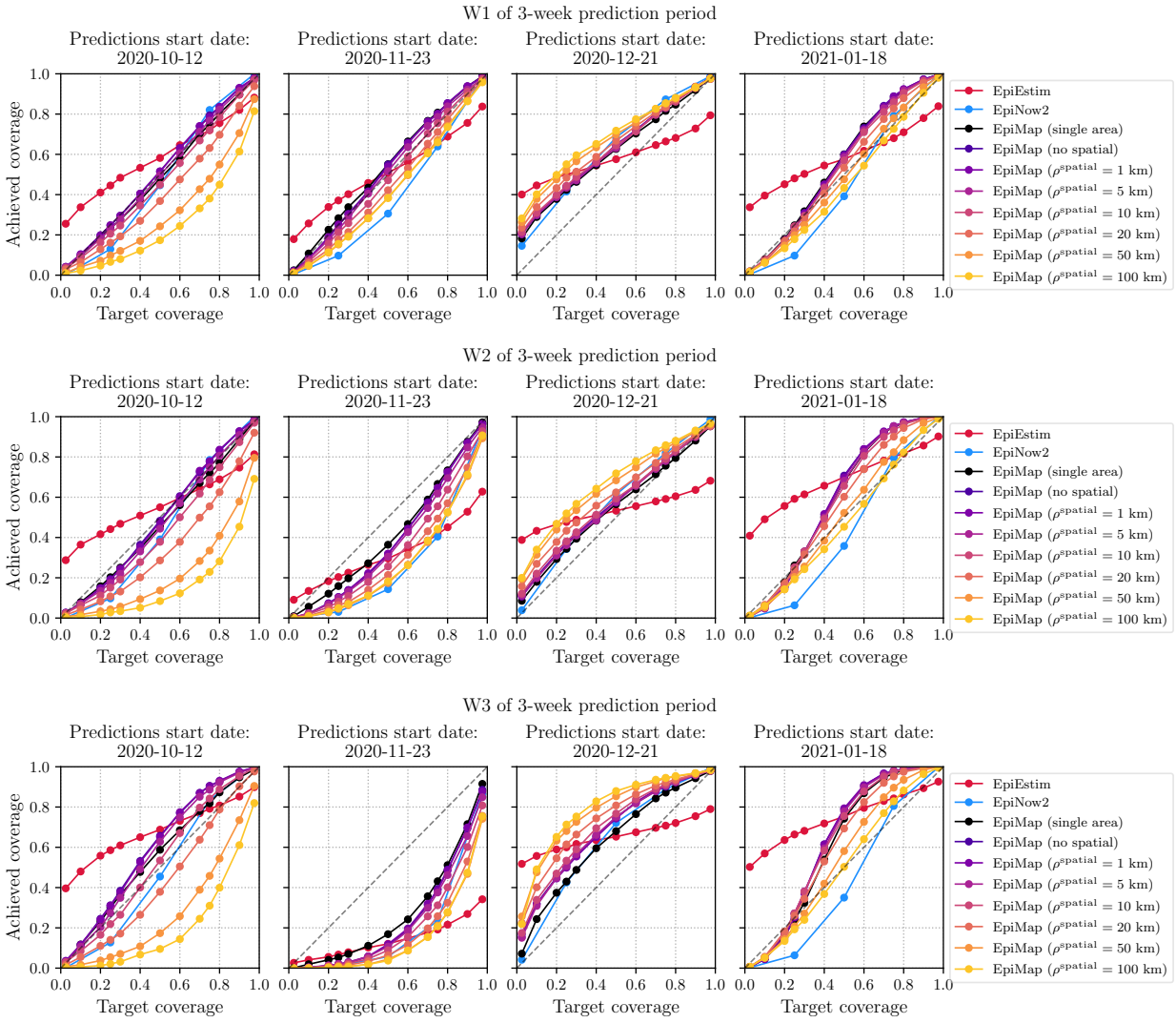

Figure 9: An evaluation of the uncertainty estimates produced by methods similar to Figure 3 but stratified by week over the 3-week prediction period. As expected, the quality of uncertainty estimates degrades in the later weeks compared to earlier weeks, as indicated by reliability curves that are further 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.

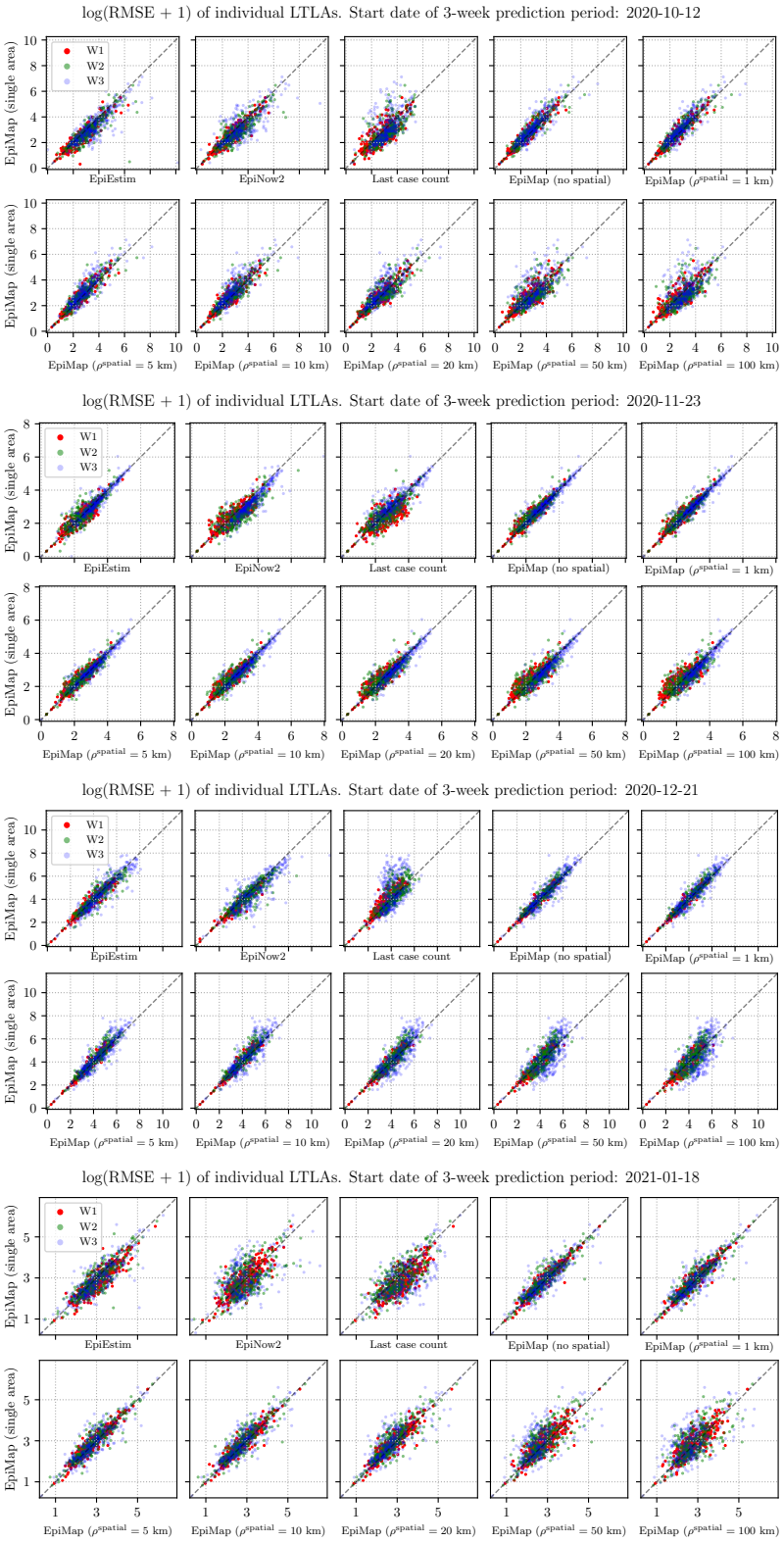

Figure 10: We plot the log(RMSE + 1) for individual LTLAs (each dot is an LTLA) stratified by week. We observe variation in predictive performance for different LTLAs for all models, with large correlation in LTLA losses between methods.

C.3 Regional estimates

Figure 11 shows the regional estimates for cases and R_t on remaining NHS regions in England, in the same setting as Figure 5.

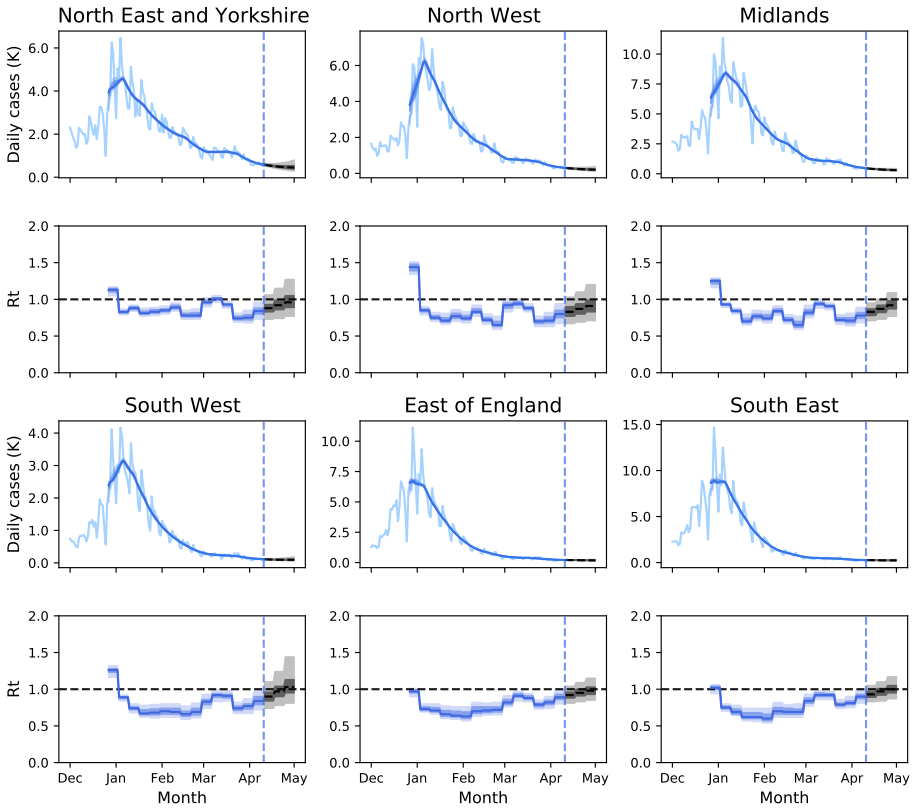

Figure 11: Additional regional estimates of cases and R_t for NHS regions in England, using case data available by the 10th April 2021.