Bayesian spatio-temporal joint disease mapping of Covid-19 cases and deaths in local authorities of England

Sujit K. Sahu* and Dankmar Böhning

Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, SO17 1BJ, UK

Abstract

The overwhelming spatio-temporal nature of the spread of the ongoing Covid-19 pandemic demands urgent attention of data analysts and model developers. Modelling results obtained from analytical tool development are essential to understand the ongoing pandemic dynamics with a view to helping the public and policy makers. The pandemic has generated data on a huge number of interesting statistics such as the number of new cases, hospitalisations and deaths in many spatio-temporal resolutions for the analysts to investigate. The multivariate nature of these data sets, along with the inherent spatio-temporal dependencies, poses new challenges for modellers. This article proposes a two-stage hierarchical Bayesian model as a joint bivariate model for the number of cases and deaths observed weekly for the different local authority administrative regions in England. An adaptive model is proposed for the weekly Covid-19 death rates as part of the joint bivariate model. The adaptive model is able to detect possible step changes in death rates in neighbouring areas. The joint model is also used to evaluate the effects of several socio-economic and environmental covariates on the rates of cases and deaths. Inclusion of these covariates points to the presence of a north-south divide in both the case and death rates. Nitrogen dioxide, the only air pollution measure used in the model, is seen to be significantly positively associated with the number cases, even in the presence of the spatio-temporal random effects taking care of spatio-temporal dependencies present in the data. The proposed models provide excellent fits to the observed data and are seen to perform well for predicting the location specific number of deaths a week in advance. The structure of the models is very general and the same framework can be used for modelling other areally aggregated temporal statistics of the pandemics, e.g. the rate of hospitalisation.

Keywords: Bayesian space-time modelling, CAR, Covid-19 case rate, Covid-19 death rate, ecological analysis, temporal disease mapping

*Corresponding author

Email addresses: S.K.Sahu@soton.ac.uk (Sujit K. Sahu), D.A.Bohning@soton.ac.uk (Dankmar Böhning)

1. Introduction

The ongoing global coronavirus pandemic has so far caused an unprecedented world-wide death toll and severe economic damage. The pandemic has taken many national governments by surprise and as a result the government response has varied widely across countries leading to different dynamics of the infections and deaths. The course of the pandemic in individual countries, and within different administrative regions, has been very varied and changed rapidly as a result of the government interventions. Of immense statistical interest is to study the spatio-temporal spread of the pandemic by jointly modelling the case and the death rates as a function of various socio-economic and environmental factors.

By now a large volume of articles in both popular press and scientific journals has been written up describing the course of the pandemic from various different view points, see e.g. Polo et al. (2020). Viewed from a statistical modelling and analysis perspective, the pandemic, however, has provided a rich collection of international and country specific data sets. Statistical methodologies of various flavours e.g. time series data analysis and spatial and spatio-temporal data analysis and comparison techniques, are at the forefront of understanding the dynamics of the pandemic. The large number of research papers published in this issue is a simple statistic that bears testament to this fact. Methodologically sound modelling of real time large scale spatial and fast dynamic of the data is a daunting task because of various reasons including the time and due diligence needed to develop robust methodologies.

The primary data set that we gather for our modelling endeavour and data analysis is the one for the different local authorities and districts in England. The nation of England, within the United Kingdom (UK), is one of the worst affected parts of the world and the administration here is run differently than the other constituent nations of the UK. The urgency of the need to develop and test modelling methodologies limits the time window, as well as the spatial domain, of the data set we use to illustrate the methods.

The main methodological focus of this article is to develop a joint descriptive model for the number of positive Covid-19 (henceforth covid) cases and deaths that can be used to explain the effect of a wide range of socio-economic risk factors, environmental pollutants and the timing of different governmental policy interventions. In so doing we aim to shed light on spatial effect of socio-economic deprivation and population density on the rates of cases and deaths. There is a growing number of studies linking the effect of pollution on the case and death rates of the pandemic, see. e.g. Konstantinoudis et al. (2020); Ramirez-Aldana et al. (2020). Our model includes the levels of atmospheric nitrogen dioxide, NO_2, which has been found to have the largest significant effect on hospitalisation due to respiratory diseases by Lee et al. (2017). The model also includes a temporal factor covariate charting the timing of the UK government's major policy changes.

Covid-19 testing capacity in the UK was hugely improved during the course of the pandemic. Hence the number of reported cases are not strictly comparable for the different weeks since increased testing is already likely to increase the number of reported cases. There are many other ways to model and analyse number of cases with the help of other data sets. For example, Lee et al.

(2021) consider modelling the volume of calls to the National Health Service in Scotland. The huge uncertainty in the case numbers motivates us to study the pandemic dynamics mainly by modelling the number of deaths. The number of cases are modelled nevertheless to build a joint model for number of cases and deaths.

A well fitted Bayesian model can potentially be used for a variety of purposes. This paper focuses on demonstrating two main uses of model fitting:

(i) understanding the spatial pattern in the death rates,

(ii) the use of the models for short-term forecasting of the number of deaths.

The proposed adaptive Bayesian model, described in Section 4, allows us to discover and highlight, if there is any, step changes of large differences in death rates between neighbouring spatial units. Identification of such step changes may help the authorities in detecting appropriate risk factors responsible for such statistically significant large differences and this in turn has the potential to lead to better policies in the future. The second objective of short-term forecasting of number of deaths is also likely to help the authorities prepare in advance for such future number of deaths in their local areas.

The topic of this paper borders an area in medical statistics commonly referred to as *disease mapping*. Other names used are spatial epidemiology, environmental epidemiology or small area health studies and *spatial biostatistics* as suggested by Lawson (2018). All of these names refer to the same essential issue: the use of spatial information to detect any form of systematic spatial disease occurrence. In fact, the analysis and recognition of disease clustering in space and its representation on a map is one of the oldest problems in epidemiology. One of the earliest applications of the method are traced back to John Snow who studied the residential distribution of cholera victims in London to lead him to identify the water pump in Broad Street as the source of contamination (Snow, 1854). In the last decades of the last century the appearance of various cancer atlases (and other disease atlases) such as the European cancer atlas (Boyle, 2008) led to a boost of methodological investigations on constructing profound maps of aggregated disease occurrence in administratively constructed spatial units (Lawson et al., 1999; Lawson, 2000).

Another bordering field is *medical surveillance* which seeks to monitor human populations for changes in disease occurrence. Disease mapping serves the monitoring purpose well in space, but medical surveillance also always has its focus on temporal change. Hence, as the next step, spatial-temporal disease mapping has occurred. In the special issue on disease mapping (Lawson, 2000), several papers addressed already the joint modelling of disease occurrence in space and time (Pickle, 2000; Böhning et al., 2000; Knorr-Held, 2000). A standard model in this area is the celebrated Besag, York and Mollié (BYM) model due to Besag et al. (1991) which has been generalised in several ways to incorporate the temporal dimension including a space-time interaction, see e.g. Anderson and Ryan (2017). Spatio-temporal disease mapping has recently received prime attention as can be seen in the books by Lawson (2018) or Finkenstädt et al. (2007). Recent review articles in this area include Wah et al. (2020).

The plan of the remainder of this article is as follows. Section 2 discusses the steps and methods used to collect the data sets to be modelled. Section 3 provides an exploratory analysis of different component data sets, viz. the death, case, and pollution data sets. This section also studies possible inter relationships between these data sets by using graphical displays and summary statistics. Section 4 discusses the modelling and validation methods using the `CARBayesST` package in the `R` language (https://cran.r-project.org/). Modelling results are discussed in Section 5. The paper ends with a brief discussion of the results in Section 6 with some suggestions for future research.

2. Data collection and pre-processing

2.1. Spatio-temporal domain of our data set

The study region selected in this paper is mainland England and the Isle of Wight, UK, partitioned into $k = 1, \ldots, K = 313$ Local Authority Districts, Counties and Unitary Authorities (LADCUA) according to the boundary data published by the UK Office for National Statistics (ONS)[1] in May 2020. Designated as the basic areal spatial units in our study, these LADCUAs include the 32 boroughs in Greater London but excludes the one for Isle of Scilly in the Atlantic ocean of the coast of Cornwall in the far South-West. Figure 1 provides a map of these LADCUAs where the boundaries are plotted using a red colour pen (thin lines if the map is read in black and white). The plotted map also shows the boundaries, using a thicker black pen, of the nine larger scale administrative regions in England. The nine regions are: East, East Midlands, London, North East, North West, South East, South West, West Midlands, and Yorkshire and Humber. Our study region excludes the other three UK nations: Scotland, Wales and Northern Ireland as those have devolved administrative powers adopting different health policies and interventions.

The time period of our study is from week 11 starting March 13 to week 46 which ended on November 20, 2020. We stopped our data collection at week 46 before moving on to starting our modelling effort for this special issue. We take week as the basic time unit of our study since a finer time unit such as daily is problematic because of the presence of excessive number of zeros in the daily covid, and also of all cause mortality data at the local authority levels. Moreover, the ONS publishes weekly covid and all cause mortality data for all the LADCUAs in the UK and these estimates are much more stable than the daily figures which vary largely due to time lags in reporting and also due to a weekday/weekend effect where the number of deaths reported in the weekend is systematically much smaller than that for the weekdays even at the peak of the pandemic.

The weekly number of coronavirus cases are obtained as aggregates from the UK Government data dashboard[2]. These weekly case numbers are completely aligned with the weekly death figures so that we are also able to study any lag-effects of cases on deaths. We, however, were unable to collect reliable data on the number of tests administered in each week. Hence the number of cases are normalised using overall population figures which we describe in the following sub-sections.

[1] https://geoportal.statistics.gov.uk/datasets/

[2] https://coronavirus.data.gov.uk/

This data selection and pre-processing stage also tackled a few issues regarding data availability for some particular LADCUAs. For example, the case numbers for the the two London boroughs of City of London and Hackney were only available as aggregated totals. We divided the weekly total case numbers proportionately between the two boroughs according to their population size in 2019.

2.2. Socio-economic variables

It is generally believed that socio-economic deprivation leads to worse health on average and this in turn may affect covid case and death rates. To investigate such issues, for each LADCUA we collected the following data from the source ONS website[3]):

(i) population data for 2019 (denoted *popn*),

(ii) population density (denoted *popden* on the log scale), defined as size of population divided by the area of the LADCUA,

(iii) percentage of the working age population receiving job-seekers allowance (denoted *jsa*) during January 2020,

(iv) median house price in March 2020 (denoted *price* on the log to the base 10 scale).

The first of these four variables, *popn*, will be used to calculate rates of cases and deaths in the subsequent sections. None of these variables vary over time (the weeks) but are spatially varying providing information regarding the long-term socio-economic nature of the spatial units. The raw values of population density have a large range from 25 to 16,427 which will cause problems in modelling. We apply the log transformation to stabilise the values of this covariate. Similarly we apply the log transformation, but using the base 10 for ease of interpretation, for the house price data.

2.3. Air pollution

To investigate the effect of air pollution we collected daily mean NO_2 concentration data reported by the functioning monitoring sites in the UK's Automatic Urban and Rural Network (AURN)[4]. We do not consider other pollutants such as particulate matter since those have less severe effect on hospitalisation due to respiratory diseases (Lee et al., 2017). The downloaded pollution data were aggregated to weekly averages by removing any daily missing observations where necessary. In total we have NO_2 data for 121 valid pollution monitoring sites and these sites are shown as blue dots in Figure 1. As this figure shows, there were many LADCUAs without any pollution monitoring site. For these LADCUAs, we estimated the weekly NO_2 by taking simple average of that week's data from the five nearest AURN monitoring sites. For all the LADCUAs where there were

[3]https://ons.gov.uk/

[4]https://uk-air.defra.gov.uk/

at least one AURN monitoring site we took the simple average of the NO_2 levels. These approximation methods have been used here for simplicity and speed instead of much better spatio-temporal modelling and aggregation methods for air pollution that are available in the literature, see e.g. Mukhopadhyay and Sahu (2018).

2.4. Spatial neighbourhood structure through the adjacency matrix

To explore spatial dependence we need to define what is known as the neighbourhood or adjacency matrix which defines the adjacent LADCUAs for each of the K LADCUAs. A common practice is to define the neighbourhood structure by using a binary $K \times K$ matrix, W where $W_{kj} = 1$ if the kth and jth LADCUA share a common border for any $k \neq j,\ k, j = 1, \ldots, K$, and 0 otherwise. Also, we assume $W_{kk} = 0$ for all $k = 1, \ldots, K$.

Note that if any of the spatial unit, say j, is an island then that row is a null vector. This is not allowed in many model fitting routines as the spatial domain must be a connected set with each unit having at least one neighbour. This is relevant here since the Isle of Wight at the bottom of the map of England (see Figure 1) is an island. To overcome this problem we assume Southampton, Portsmouth and Dorset as neighbours of the island since there are passenger ferry connection from these LADCUAs to the Isle of Wight. In so doing, we note that, Isle of Wight is also assumed to be a neighbour of these three spatial units in our study.

The resulting adjacency matrix W has 1602 non-zero entries distributed according to the following frequency table: The median number of neighbours is 5 and the standard deviation is 2.05

Table 1: Frequency table of number neighbours.

No of neighbours	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	38	57	56	45	52	17	10	3	2	1

indicating a rich spatial adjacency structure of the LADCUAs.

3. Exploratory data analysis

3.1. Calculating the expected numbers of cases and deaths

The data collection exercise described in Section 2 provides us with a data set consisting of 11,268 (313×36) rows of data from the $K = 313$ LADCUAs and $T = 36$ weeks from March to November. The main response variables for our spatio-temporal modelling are the number of covid deaths, denoted by Y_{kt}, and cases denoted by Z_{kt}, $k = 1, \ldots, K$ and $t = 1, \ldots, T$. Also let D_{kt} denote the number of deaths *from all causes*, including covid, in LADCUA k in week t. The magnitude of Y_{kt} depends on the size and demographics of the population at risk, which we adjust for by computing the expected number of covid deaths E_{kt}. There are many alternative ways to obtain the expected number of deaths, e.g. by considering the national age and sex specific death rates and then by performing internal or external indirect standardisation.

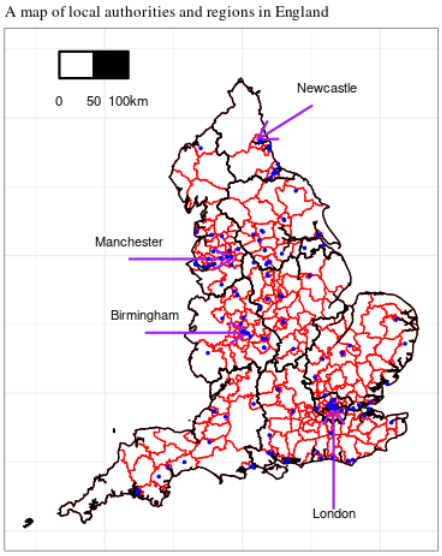

Figure 1: A map of the local authorities and nine administrative regions in England. Air pollution monitoring sites are shown as blue dots in the map.

The latter is an elementary and basic method in epidemiology to achieve comparability of mortality or morbidity across different sub-populations see e.g. Lilienfeld and Lilienfeld (1980) or Bland (2000). Indirect standardisation is often considered the preferred method when the study populations have only few events so that stratification, as required by the direct method, would contribute to instability of the stratified estimates (Miettinen, 1985). A review of standardisation is provided by Keiding (1987) who also discusses its historic roots. Indirect standardisation uses the concept of what number of events would be expected if the study population had the stratum distribution of a reference population. If this reference population is available the indirect standardisation turns out to be *external*. Often this external reference population is not available and needs to be constructed. This can be accomplished by merging the study populations over the different groups to be compared (here the LADCUAs). This process is called *internal* indirect standardisation. Applying the rates of the reference population to the number at risk in the study population gives the *expected* number of events E which then is compared with the observed number of events O, leading to the standardised event ratio, more popularly known as the standardised mortality (or morbidity) ratio, $SMR = O/E$. Often the *SMR* is presented as multiplied by 100 which suggests the interpretation as percentage which is misleading. The *SMR* is a rate ratio comparing the event rate in study population to the event rate in the reference population.

Statistical inference for the *SMR* is widely based on the assumption that O follows a Poisson distribution with the number of expected events E as the offset, see e.g. Keiding (1987) and Clayton and Hills (1993). One difference between internal and external indirect standardisation is that for the internal method the sum of the number of observed event over all study populations equals the

sum of expected number of events over all study populations. This is not necessarily the case for the external indirect method.

Estimation of mortality and morbidity requires a denominator (the number of persons at risk) which is not easily accessible in our application. However, the number of deaths from all causes is available which leads to the concept of *proportional mortality ratio*, ***PMR***, defined as the number of deaths from a specific cause (covid) divided by the number of total deaths, see e.g. Last (2000). A direct argument shows that the ***PMR*** is identical to the ratio of the specific mortality rate to the mortality rate from all causes. The ***PMR*** is frequently used if population denominators are not readily accessible.

We now combine the philosophies of internal indirect standardisation and ***PMR*** to an appropriate application in our case study. The study populations are given by the various regions and the various weeks. We begin by calculating the number of expected deaths per week t and region k.

The availability of the number of deaths from all causes, D_{kt}, in week t and region k allows us to estimate E_{kt}s by indirect standardisation as follows. Let

$$R = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T} Y_{kt}}{\sum_{k=1}^{K}\sum_{t=1}^{T} D_{kt}}$$

denote the overall proportion of covid deaths among deaths from all causes. We consider this as an estimate of the ratio of the Covid-19 mortality to the mortality due to all causes for the entire country England, i.e. collapsed over all weeks and regions. Then, the expected number of covid deaths is defined by:

$$E_{kt} = R \times D_{kt}, \;\; k = 1, \ldots, K, t = 1, \ldots, T.$$

Obviously, $\sum_{k=1}^{K}\sum_{t=1}^{T} Y_{kt} = \sum_{k=1}^{K}\sum_{t=1}^{T} E_{kt}$, so this can be viewed as a form of internal indirect standardisation. Having obtained the E_{kt} we obtain the standardised mortality ratio, $SMR_{kt} = Y_{kt}/E_{kt}$ for all values of k and t. The SMR_{kt} has a simple interpretation, e.g. a value of 1.2 corresponds to a 20% increased risk compared to E_{kt}. A zero value of D_{kt} for some particular combinations of k and t leads to $E_{kt} = 0$, which in turn causes problem in defining the ***SMR***. To overcome this problem we replace the zero D_{kt} values by 0.5 and adjust R and E_{kt} accordingly. This problem would have been much more severe were we to model daily instead of weekly data.

A similar expected number of positive covid cases is required to analyse the observed number of cases, previously denoted by Z_{kt}, at the tth week in the kth LADCUA. Corresponding to R, the overall proportion of covid deaths, we define the overall proportion of infections in the population as:

$$R_C = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T} Z_{kt}}{\sum_{k=1}^{K}\sum_{t=1}^{T} P_{kt}}$$

where P_{kt} denotes the population at risk at time t in the kth LADCUA. Hence, the expected number of cases, corresponding to Z_{kt}, denoted by C_{kt} and is defined as:

$$C_{kt} = R_C \times P_{kt}, \;\; k = 1, \ldots, K, t = 1, \ldots, T. \tag{1}$$

Clearly, $\sum_{k=1}^{K}\sum_{t=1}^{T} Z_{kt} = \sum_{k=1}^{K}\sum_{t=1}^{T} C_{kt}$. In our implementation we assume constant population over the weeks, i.e. $P_{kt} = P_k$ where the P_k is the ONS estimated population in 2019, *popn*, obtained previously at the data collection stage in Section 2.2. Note that C_{kt} is always positive since $P_{kt} > 0$ for all values of k and t. Again, this can be viewed as a form of internal indirect standardisation. Our focus is on the standardised Covid-19 morbidity ratio (*SCMR*), $\textit{SCMR}_{kt} = \frac{Z_{kt}}{C_{kt}}$.

3.2. Summary statistics and graphical displays

The number of covid deaths, Y_{kt}, varies over k and t with a mean value of 5.2 and a maximum value of 246. These values are not comparable across the different spatial units due to the differences in the at risk population size. Hence, for the purposes of exploratory analysis only, we transform Y_{kt} into the rate per 100,000 people to have comparable weekly rates for the K spatial units. The death rate, so obtained, ranges from 0 to 40 with a mean value of 2.9. This rate of death is plotted in side-by-side boxplots in Figure 2 for the 36 weeks in our study. In this figure, a simple 4-colour pen shows the four different phases of the course of the pandemic as seen through the timings of government interventions. The four phases are:

1. weeks 11 to 22 (March 13 to June 4) first phase ending at the end of the first lockdown period,
2. weeks 23 to 37 (June 5 to September 17) during the summer easing of the lockdown,
3. weeks 38 to 41 (September 18 to October 8) when some social mixing restrictions were re-introduced,
4. weeks 42 to 46 (October 9 to November 20) when local tiering system was introduced at the start of the period.

This factor variable, denoted *Course*, will be used in our modelling and analysis of the number of cases in Section 5. Figure 2 shows the first peak of the pandemic during weeks 15 and 16 (April 4-23). The death rate rises very steeply during weeks 12-15 but comes down very slowly during weeks 15 to 28. The death rate comes to the lowest phase of severity in week 28 (July 7th) and then starts to rise up again in week 42 (October 16th). The remaining colour black has been used to join the medians of the box-plots.

To associate the covid death rates with the temporally static but spatially varying covariates, *popden*, *price* and *jsa* we obtain an aggregated over time value of the *SMR* of covid deaths for each of the K LADCUAs. This aggregated *SMR* is obtained simply as the ratio of the sum of the observed and expected number of covid deaths over the 36 weeks. Note that this is the ratio of the sum over the weeks – not the average of the ratios for 36 weeks. Figure 3 plots a map of the *SMR* and the three important covariates in different panels. The map of the log-population density, *popden*, shows the highest level of correlation among the three covariates plotted in this figure. The covariate *j*sa shows up areas in the north of England having high levels of both *SMR* and *SCMR* and *j*sa. The median house price in the log-10 scale, *price*, also correlates (but negatively) well with the

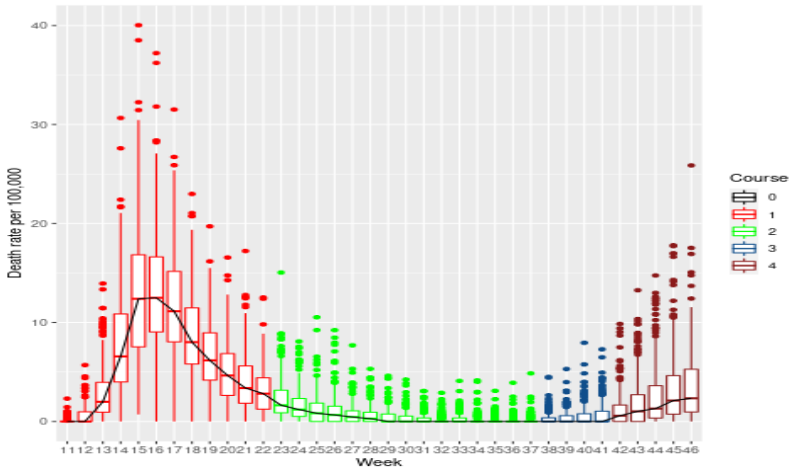

Figure 2: Boxplots of weekly death rates per 100,000 population. The boxplots are coloured according to the factor variable Course.

SMR values – showing on average higher house price values in the south where the *SMR* values are lower on average.

We now investigate the rate of reported number of cases, previously denoted as Z_{kt}. The number of cases, ranging from 0 to 4447, has a very high level of variability and this is not comparable across the different spatial units. Hence we apply the same transformation to convert the raw number of cases into the number of cases per 100,000 people. Figure 4 provides three important plots showing various aspects of the case numbers. The top left panel shows a clear north-south divide in the *SCMR*. The panel in the top-right, similar to Figure 2, shows the rise and fall and then rise again of the case dynamics. The colour indicator in this panel is the same as the one in Figure 2 showing the four phases of the course of the pandemic. Note that during the latter phases the number of cases were recorded higher as the testing capacity hugely improved during the course of the pandemic. The bottom panel in this figure plots the average weekly number of deaths and cases both per 100,000 people. This clearly shows the reduced death rates compared to case rates in the second peak starting in October. This is because of several reasons including the availability of more number of tests during the later weeks and also improvement in both understanding and treatment of the disease.

To explore the relationships between the *SCMR* values on the top-panel of Figure 4 for the cases and the *SMR* values on the top panel of Figure 3 for deaths, we provide a pair-wise scatter plot in Figure 5. The scales of the plotted variables have been described in the caption of this figure. Various relationships between the variables present in the data set can be read from Figure 5. The numerical values of the correlations shown in the plot reveal significantly moderate level of positive association between the log of the *SCMR* for cases and square-root NO_2, *jsa, popden* and the log of the *SMR* of covid deaths. The *SCMR* of the cases are higher in the lower house price areas - mostly in the north and the *SCMR* of the cases are lower in the areas where *jsa* is lower. Population density

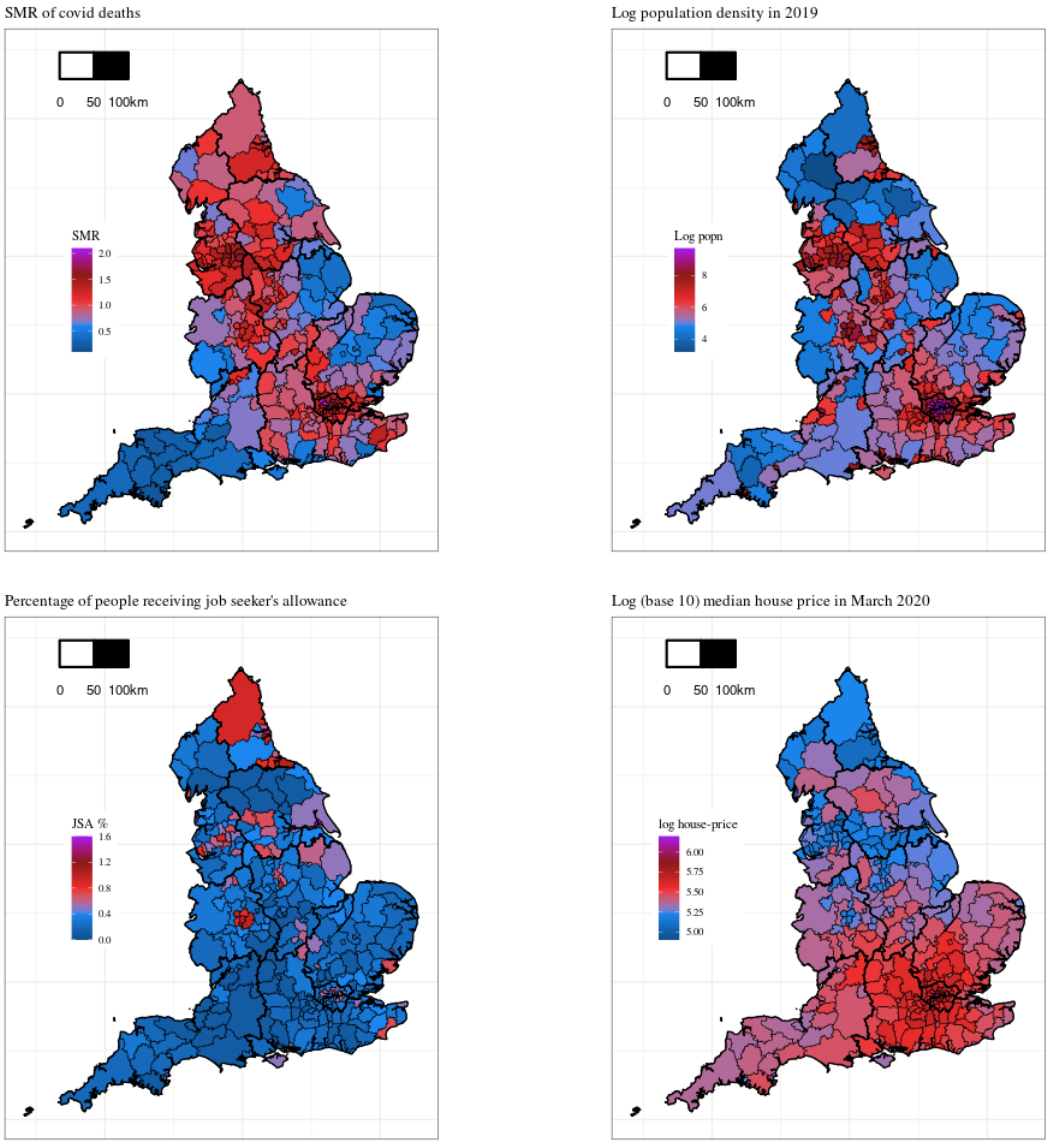

Figure 3: Average *SMR* of covid deaths (top-left) and three relevant variables capturing socio-economic information.

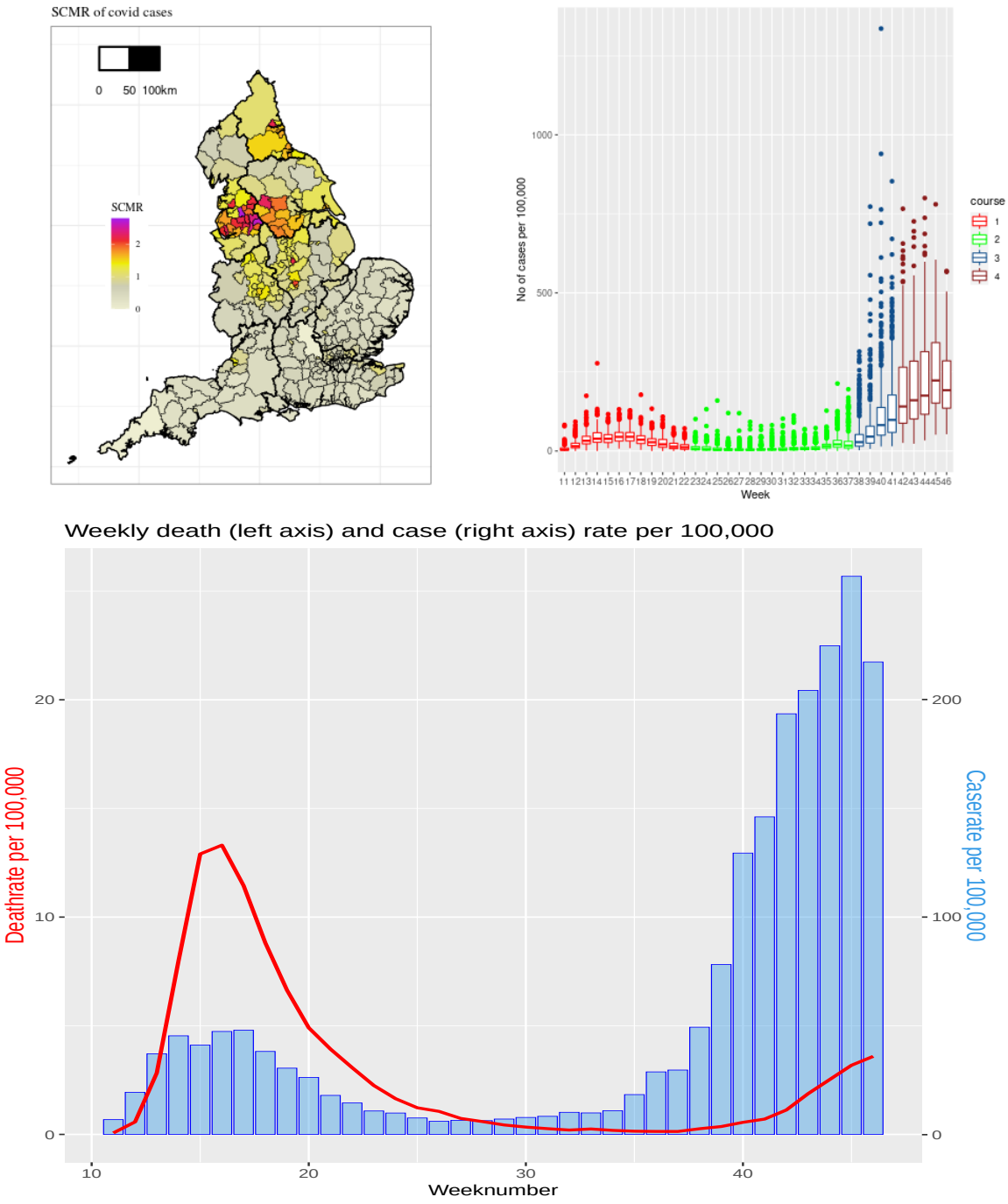

Figure 4: Graphical summaries of case rates

(on the log scale) correlates positively with the *SCMR* of cases and *SMR* of deaths. Obviously, higher levels of case rates leads to higher numbers of death.

Looking at the last column of the plot labelled `deathsmr` we see that the *SMR* of covid deaths is positively associated with *jsa* and *popden* but it shows a negative association with *price*. This is because of the north-south divide in house prices in England and also a similar divide in opposite direction for the death rates as seen in Figure 3. However, the higher house prices in London in the south and also the higher levels of deaths in London reduces the strength of this north-south divide.

The levels of the air pollutant NO_2 is the only spatio-temporally varying covariate in our study. Recall that Figure 5 shows a significant positive association (correlation value 0.314) of NO_2 with the log of *SCMR* for cases but not *SMR* for deaths. Figure 6 provides two further important exploratory plots for associating NO_2 with the case and death rates. The left panel of this figure shows higher levels of average (over weeks) NO_2 in the major urban areas including London. The right panel shows higher level of case rates are associated with higher levels of NO_2, confirming similar findings in Figure 5.

4. A two stage Bayesian hierarchical model

Recall that the primary aim of this article is to study the spatio-temporal dynamics of number of weekly covid deaths previously denoted by Y_{kt}, $k = 1, \ldots, K$ and $t = 1, \ldots, T$. Obviously, Y_{kt} is dependent on Z_{kt}, the number of positive cases in that week and before in the previous weeks as there is a time lag between the events diagnosis, possible hospitalisation and finally death. Hence we adopt a natural two stage Bayesian hierarchical modelling framework for joint modelling of Z_{kt} and Y_{kt}. Specifically, we model Z_{kt} marginally and then model Y_{kt} given Z_{kt} and its time lag values Z_{kj} for $j = t, t-1, \ldots, t-q$ where q is the number of lags to be determined.

To build explanatory models based on covariates, let $\mathbf{x}_{kt}$ denote the p-dimensional spatio-temporal covariates at space-time combination k and t which may explain either the number of cases, number of of deaths or both. That is, $\mathbf{x}_{kt}$ is the collection of all relevant covariates at location k and at time t.

The two stage model is specified by:

$$Z_{kt} \sim \text{Poisson}\left(C_{kt}\lambda^z_{kt}\right), \tag{2}$$

$$Y_{kt}|\mathbf{Z}_{k[t]} \sim \text{Poisson}\left(E_{kt}\lambda^y_{kt}\right), \tag{3}$$

where $\mathbf{Z}_{k[t]}$ denotes the vector of the number of cases at location k upto and including time t and λ^z_{kt} and λ^y_{kt} are the relative risks of cases and deaths that we model below based on the covariates, structured spatio-temporal random effects denoted by ψ^z_{kt} and ψ^y_{kt} below, and additionally the observed $\mathbf{z}_{k[t]}$ for λ^y_{kt}. The models are:

$$\log(\lambda^z_{kt}) = \mathbf{x}'_{kt}\boldsymbol{\beta}^z + \psi^z_{kt}, \tag{4}$$

$$\log(\lambda^y_{kt}) = \mathbf{x}'_{kt}\boldsymbol{\beta}^y + \sum_{j=0}^{q} \eta_j \log\left(\frac{z_{k(t-j)}}{C_{k(t-j)}}\right) + \psi^y_{kt}, \tag{5}$$

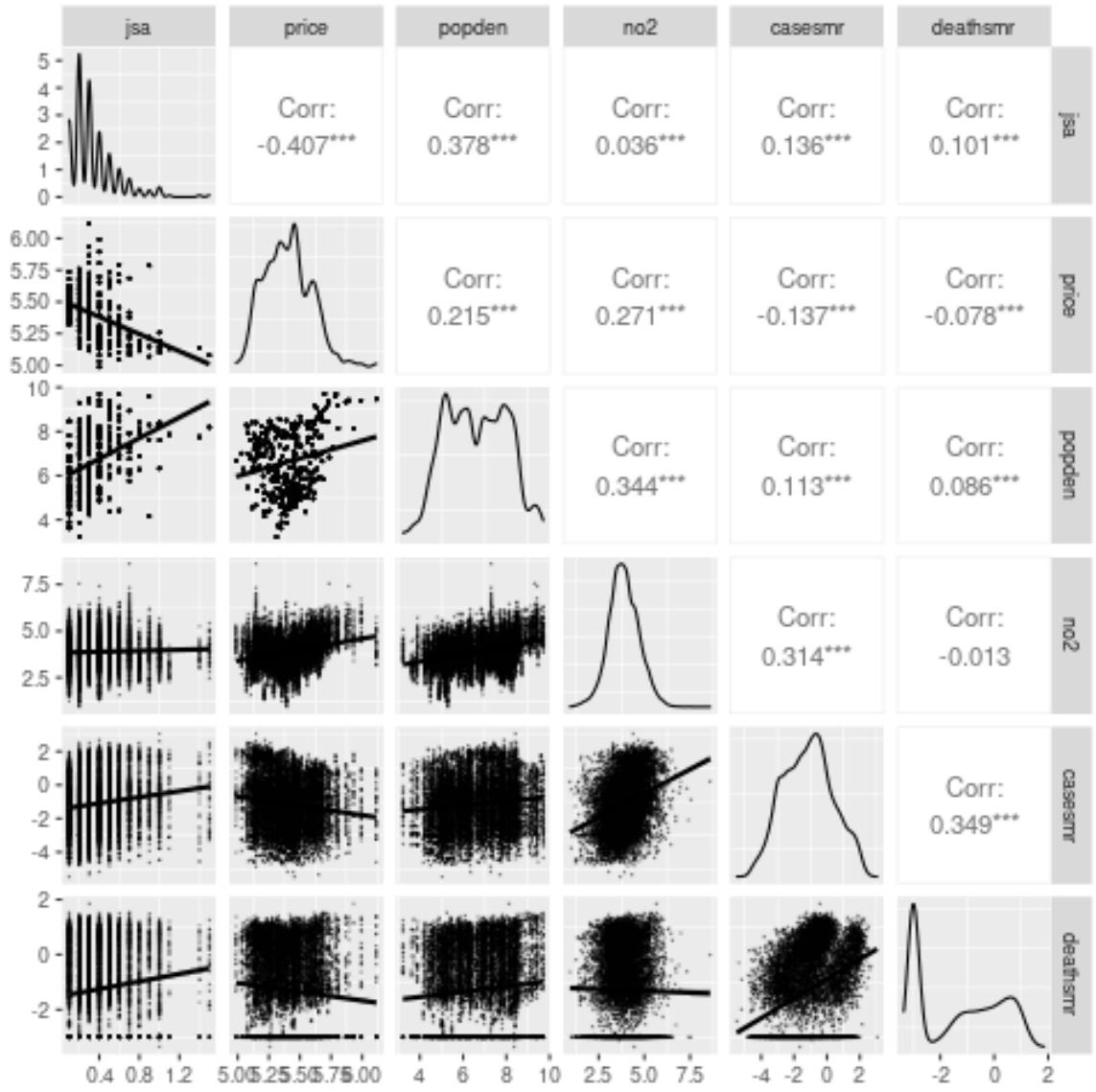

Figure 5: Pairwise scatter plots of the log *SCMR* values of covid cases and log *SMR* values for deaths along with various covariates. The variable *price* is on log to the base 10 scale, *popden* is on the log scale, NO_2 is on the square-root scale, *case smr* and *smr* for death are on the log-scale. To avoid having to take log of zero the zero values of the *SMR* have been replaced by 0.05.

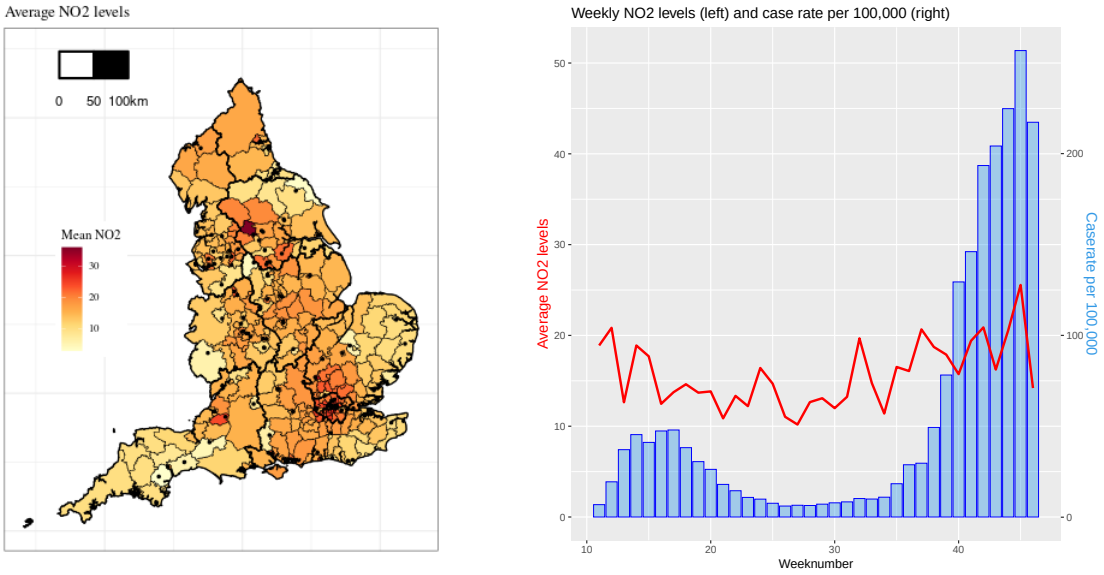

Figure 6: Left panel: average NO_2 levels in each LADCUA during the 36 weeks. Right panel: plot associating NO_2 with covid case rates.

for $k = 1, \ldots, K$ and $t = 1, \ldots, T$. The rationale used to specify the above two models and the notations used are explained below. Inclusion of a large number of lagged values as regressors may lead to the multicollinearity problem which is discussed further in Section 5.

- Here $\boldsymbol{\beta}^z$, $\boldsymbol{\beta}^y$ and $\boldsymbol{\eta} = (\eta_0, \ldots, \eta_q)$ are the regression coefficients and q is the number of past week's number of cases that we intend to include in the model for the number of deaths. The number of lags q can be chosen by usual methods of model choice. Note that $\boldsymbol{\beta}^z$ and $\boldsymbol{\beta}^y$ are allowed to be of different dimensions since the two sets of covariates for explaining cases and deaths can be different, although we abuse the notation $\mathbf{x}_{kt}$ by keeping it the same in both (4) and (5). All these regression co-efficients are provided default flat prior distributions in our Bayesian modelling.

- In model (5) we use the lagged values of $SCMR_{kt} = \frac{z_{kt}}{C_{kt}}$, defined in Section 3.1, to build the auto-regression in number of cases. This is justified since,

$$SCMR_{kt} = \frac{z_{kt}}{C_{kt}} = \frac{z_{kt}}{R_C P_{kt}} = \frac{1}{R_C P_k} z_{kt},$$

and as a result

$$\log\left(SCMR_{kt}\right) = \log \frac{z_{kt}}{C_{kt}} = -\log(R_C P_k) + \log(z_{kt}).$$

Thus the $\log(\lambda_{kt}^y)$ in (5) is truly built on the auto-regression in number of cases z_{kt} using the appropriate log scale.

- The random effects ψ^z_{kt} and ψ^y_{kt} are entertained to have the following rich variety of models based on the different assumptions for their structures. Bayesian model selection criteria such as the DIC (Spiegelhalter et al., 2002) and WAIC (Watanabe, 2010) will be used to choose among these models. For notational convenience, we omit the super scripts y and z in ψ^z_{kt} and ψ^y_{kt} since each of these two can be assumed to follow the models described below. We consider the following general spatio-temporal models.

$$\psi_{kt} = \begin{cases} \phi_k + \delta_t + \gamma_{kt}, & \text{Anova model} \\ \phi_{kt} + \delta_t, & \text{Separable or AR model} \\ \phi_{kt} + \lambda_{U_{kt}}, & \text{Localised model} \\ \phi_k + \delta_t, & \text{BYM-AR model} \end{cases} \tag{6}$$

for $k = 1, \ldots, K$ and $t = 1, \ldots, T$. All the terms in the right hand side of Equation (6) are random effects with different structures discussed in the sub-sections below.

To facilitate the discussion, we use the generic notation $\boldsymbol{\xi}_{-i}$ to denote the $n-1$ dimensional vector by deleting the $i$the element, i.e.,

$$\boldsymbol{\xi}_{-i} = (\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_n),$$

for $i = 2, \ldots, n-1$, and $\boldsymbol{\xi}_{-1}$, $\boldsymbol{\xi}_{-n}$ are defined by deleting the first and last entry of $\boldsymbol{\xi}$ respectively. This notation allows us to introduce the Leroux (Leroux et al., 2000) CAR prior distribution,

$$\xi_i | W, \rho, \tau^2, \boldsymbol{\xi}_{-i} \sim N\left(\frac{\rho \sum_{j=1}^n w_{ij} \xi_j}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{j=1}^n w_{ij} + 1 - \rho} \right), \tag{7}$$

for $i = 1, \ldots, n$. We denote this CAR prior distribution as NCAR $\left(\boldsymbol{\xi} | W, \rho, \tau^2\right)$ for future reference. The precision matrix for the random vector $\boldsymbol{\xi}$ implied by this prior specification is given by

$$Q(W, \rho, \tau^2) = \rho \left[\text{diag}(W\mathbf{1}) - W\right] + (1-\rho)\mathbf{I} \tag{8}$$

where $\mathbf{1}$ is the K dimensional vector of ones and $\mathbf{I}$ is the n dimensional identity matrix and diag$(W\mathbf{1})$ is a diagonal matrix, see e.g. Banerjee et al. (2015). The celebrated BYM CAR model (Besag et al., 1991) is a special case of this model when we take $\rho = 1$. We discuss these models in detail below.

4.1. Anova model

All three sets of parameters denoted by ϕ_k, δ_t and γ_{kt} in the Anova model (6) are random effects and let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_T)$. Following Knorr-Held (2000) we assume the models:

$$\begin{aligned} \boldsymbol{\phi} &\sim \text{NCAR}\left(\boldsymbol{\phi} | W, \rho_S, \tau_S^2\right) \\ \boldsymbol{\delta} &\sim \text{NCAR}\left(\boldsymbol{\delta} | D, \rho_T, \tau_T^2\right) \\ \gamma_{kt} &\sim N(0, \tau_I^2), k = 1, \ldots, K, \ t = 1, \ldots, T. \end{aligned}$$

Here $\rho_S, \rho_T, \tau_S^2, \tau_T^2$ are auto-regression and variance parameters for the spatial and temporal processes and D denotes the $T \times T$ temporal adjacency matrix with $d_{ij} = 1$ if $|i - j| = 1$ and 0 otherwise. The interaction effect, γ_{kt}, is assumed to be independent and normally distributed for all values of k and t with variance τ_I^2.

The parameters ρ_S and ρ_T are given independent uniform prior distributions in the unit interval $(0, 1)$ and the variance parameters τ_S^2, τ_T^2 and τ_I^2 are given the inverse gamma prior distribution. This model has been implemented as the `ST.CARanova()` model in the `CarBayesST` (Lee et al., 2018) package.

4.2. Separable model

A separable model is obtained as a special case of the second version of (6) (labelled as Separable or AR model) as follows. Here we first define $\boldsymbol{\phi}_t = (\phi_{1t}, \ldots, \phi_{Kt})$ for each $t = 1, \ldots, T$. Each of these vector random effects $\boldsymbol{\phi}_t$ is given an independent NCAR prior distribution with the the same spatial correlation ρ_S but with a different temporal variance τ_t^2.

$$\begin{aligned} \boldsymbol{\phi}_t &\sim \text{NCAR}\left(\boldsymbol{\phi}_t | W, \rho_S, \tau_t^2\right), t = 1, \ldots, T \\ \boldsymbol{\delta} &\sim \text{NCAR}\left(\boldsymbol{\delta} | D, \rho_T, \tau^2\right). \end{aligned}$$

Here ρ_S, ρ_T are auto-regression parameters which are the given independent uniform prior distribution as before. The variance parameters $\tau_1^2, \ldots, \tau_T^2$ and $\tau^2$are given the inverse gamma prior distribution. This model has been implemented as the `ST.CARsepspatial()` model in the `CarBayesST` (Lee et al., 2018) package.

4.3. AR model

A temporal auto-regressive model is assumed as a special case of the separable model in Section 4.2 where we assume $\delta_t = 0$ for all t and

$$\begin{aligned} \boldsymbol{\phi}_t - \rho_T \boldsymbol{\phi}_{t-1} &\sim \text{NCAR}\left(\boldsymbol{\phi}_t | W, \rho_S, \tau^2\right), \\ \boldsymbol{\phi}_1 &\sim \text{NCAR}\left(\boldsymbol{\phi}_1 | W, \rho_S, \tau^2\right). \end{aligned}$$

Here the temporal auto-correlation is induced by the autoregression term $\rho_T \boldsymbol{\phi}_{t-1}$. Prior distributions for the parameters are assumed as before. This model has been implemented as the `ST.CARar()` model in the `CarBayesST` (Lee et al., 2018) package.

4.4. Localised AR model

A localised model(Lee et al., 2014) is obtained as an extension of the AR models defined in Section 4.3 where we assume:

$$\begin{aligned} \psi_{kt} &= \phi_{kt} + \lambda_{U_{kt}}, k = 1, \ldots, K, \ t = 1, \ldots, T \\ \boldsymbol{\phi}_t - \rho_T \boldsymbol{\phi}_{t-1} &\sim \text{NCAR}\left(\boldsymbol{\phi}_t | W, 1, \tau^2\right) \\ \boldsymbol{\phi}_1 &\sim \text{NCAR}\left(\boldsymbol{\phi}_1 | W, 1, \tau^2\right), \end{aligned}$$

where note that the BYM CAR model is assumed for the autoregressive errors. The parameters $\lambda_{U_{kt}}$ determine the localised structure in the model. The random variables Y_{kt} has a piece wise constant clustering or intercept $\lambda_{U_{kt}}$ that depends on another random variable U_{kt} specified as below. Two data points Y_{kt} and Y_{js} will have similar values $\lambda_{U_{kt}} = \lambda_{U_{js}}$ but they will show a larger difference if $\lambda_{U_{kt}} \neq \lambda_{U_{js}}$. Suppose that piece wise constant intercept or clustering process, $\lambda_{U_{kt}}$, comprises of at most G distinct levels, making this component a piece wise constant intercept term. The G levels are ordered via the prior specification:

$$\lambda_j \sim U(\lambda_{j-1}, \lambda_{j+1}), \text{ for } j = 1, \ldots, G,$$

where $\lambda_0 = -\infty$ and $\lambda_{G+1} = \infty$. Here $U_{kt} \in \{1, \ldots, G\}$ and this controls the assignment of Y_{kt} to have one of the G intercept levels. A penalty based approach is used to model U_{kt} and G is chosen larger than necessary and a penalty prior is used to shrink it to the middle intercept level. More details regarding this specification are provided in Lee et al. (2018). This model has been implemented as the `ST.CARlocalised()` model in the `CarBayesST` (Lee et al., 2018) package. All the associated variance parameters are given the package default inverse gamma prior distributions.

4.5. Adaptive AR model

In this section we describe an adaptive AR model developed by Rushworth et al. (2017). In the AR model the random effects have a single level of spatial dependence that is controlled by ρ_S. The adaptive model of this section allows locally varying levels of spatial correlation. As in the localised model in Section 4.4 the adaptive model allows for localised spatial autocorrelation by allowing spatially varying random effects to be correlated (inducing smoothness) or conditionally independent (no smoothing), which is achieved by modelling the non-zero elements of the binary proximity matrix W instead of assuming those to be fixed at 1.

The collection of non-zero W_{ij}'s are denoted by $\mathbf{w}^+$ and those are transformed to the logit scale by writing

$$\mathbf{v}^+ = \log\left(\frac{\mathbf{w}^+}{1 - \mathbf{w}^+}\right).$$

Rushworth et al. (2017) assume a shrinkage prior

$$f(\mathbf{v}^+|\tau_w^2, \mu) \propto \exp\left[-\frac{1}{2\tau_w^2}\left(\sum_{v_{it} \in \mathbf{v}^+}(v_{it} - \mu)^2\right)\right],$$

and then τ_w^2 is assumed to have an inverse gamma prior distribution. The parameter μ is treated to be fixed as in Rushworth et al. (2017) to avoid having numerical problems in MCMC model fitting. The prior distribution for $\mathbf{v}^+$ assumes that the degree of smoothing between pairs of adjacent random effects is not spatially dependent. Rushworth et al. (2017) provide more theoretical discussion regarding this model. This model has been implemented as the `ST.CARadaptive()` model in the `CarBayesST` (Lee et al., 2018) package. All the associated variance parameters are given the package default inverse gamma prior distributions.

4.6. Intermediate models

The rich variety of models described in the earlier subsections naturally provides scope for experimenting with a battery of intermediate models. For example, one may assume an independent BYM CAR prior distribution for the spatial effects and a simple auto-regressive model for the temporal random effects. Thus we may assume,

$$\begin{aligned}
\psi_{kt} &= \phi_k + \delta_t, \;\; k = 1, \ldots, K, \; t = 1, \ldots, T, \\
\boldsymbol{\phi} &\sim \text{NCAR}\left(\boldsymbol{\phi}|W, 1, \tau^2\right), \\
\delta_t &\sim N(\rho_T \delta_{t-1}, \tau^2), \\
\delta_0 &\sim N(0, \tau^2).
\end{aligned}$$

This model can be easily fitted using the R-INLA package[5], see, e.g. Blangiardo and Cameletti (2015) and we label model as the BYM-AR1 model.

Note that this BYM-AR1 model, fitted by R-INLA, does not directly correspond to any of the previously described models. It does not correspond to the AR model in Section 4.3 since in this model the AR1 term for time is taken to be independent of the CAR model for the spatial effects.

We do not consider any further intermediate models. The broad class of models parameterised with the choice of fixed $\rho_S = 1$ and the flexibility in defining the dependence structures through the spatial adjacency matrix W and temporal adjacency matrix D may provide a very similar model which may be a good approximation to the new model. Hence we do not consider any further new model and instead proceed to fit the models already discussed.

5. Modelling results

5.1. Models for the weekly number of cases

Joint modelling of the number cases and deaths begins by modelling the number of cases marginally. We consider only three possible models for the number of cases that are implemented in the `CarBayesST` software package. We do not consider the localised and adaptive models since those models had MCMC convergence problems. The random effects part of the three models, ψ^z_{kt} in (4), are specified as in `ST.CARanova()`, `ST.CARsepspatial()` and `ST.CARar()` described in the vignette of the package, see Lee et al. (2018). These models, each with the regression part specified below, are labelled respectively as C3 to C5.

The regression part of the model $\mathbf{x}'_{kt}\boldsymbol{\beta}$ is specified by the model:

$$\textit{jsa} + \textit{price} + \textit{popden} + \text{NO}_2 + \textit{Course},$$

where *price* and *popden* are on the log scale as defined before, NO_2 is on the square-root scale, and *Course* is a factor variable with four levels defined in Section 3.2.

[5] https://www.r-inla.org/

Each of the three models, C3-C5, was run for 100,000 iterations after an initial burn-in of 20,000 iterations. The samples collected after a thinning of 100 iterations to reduce autocorrelation in the MCMC samples produced the following model choice statistics reported in Table 2. This table shows that C3 (Anova) is the best model according to the WAIC and the log-likelihood (last column) while C5 is the best according to the DIC. However, there is only negligible difference in the values of the DIC criteria for the two models C3 and C5 and further investigation (not shown here) do not show any large differences between the two models. Hence, we we adopt the simpler Anova model C3 for making inference.

Table 2: Model choice statistics for the six fitted models for the number of cases.

Model	DIC	p.d	WAIC	p.w
C3 (Anova)	80898.2	10125.5	**78577.5**	5664.9
C4 (Separable)	81376.2	10283.1	79393.3	5986.3
C5 (AR)	**80895.1**	10072.0	78632.9	5671.3

The parameter estimates from the CAR Anova model C3 are presented in Table 3. The table shows that the covariate *jsa* is not important in modelling the number of cases, but *price*, NO_2 and the factor variable for different phases of the pandemic management are all significant. More expensive house price areas in the south of England are seen to have a significantly less number of cases compared to the cheaper house price areas mostly in the north. Higher levels of NO_2 significantly increase the number of cases - establishing a possible significant link between air pollution and disease vulnerability. This is seen to be especially important since this significant effect is in addition to the pandemic intervention effect as measured by the factor `Course`. The estimates for the three higher levels of this factor, compared to the base line category of the first lockdown period, are all significant, which indicate that the number of cases during the summer easing period were significantly down but the number of cases were significantly higher after week 37. However, this effect must be interpreted with a bit of caution since the UK government also increased the number of tests over time, and also since our data set does not include the number of tests administered in each week as mentioned before in Section 2.

5.2. Models for the weekly number of deaths

We first perform model choice from among six possible models that are implemented in the `CarBayesST` software package. The random effects part of the model, ψ_{kt}^y in (5), are specified as in `ST.CARanova(), ST.CARsepspatial(), ST.CARar(), ST.CARlocalised()` and `ST.CARadaptive()`, described in the vignette of the package, see Lee et al. (2018). These models, each with the regression part specified in the below, are labelled respectively as M3 to M7.

The number of deaths clearly depends on the number of cases observed in the current week and previous weeks. Also, in the UK death due to Covid-19 is defined as death which occurs within 28-days of testing positive for Covid-19. Hence, and following the modelling developments in Section 4, we use the log *SCMR* values for the current and four previous weeks, viz. $\log(SCMR_{k(t-j)})$

Table 3: Parameter estimates for the CAR Anova model, C3 fitted to the number of cases.

	Median	2.5%	97.5%
Intercept	2.885	1.282	4.627
jsa	-0.078	-0.198	0.076
price	-0.993	-1.321	-0.682
popden	0.094	0.075	0.126
NO_2	0.214	0.184	0.243
Week(22,37]	-1.166	-1.201	-1.109
Week(37,41]	0.941	0.881	1.121
Week(41,46]	1.975	1.813	2.008
τ_S^2	0.005	0.002	0.009
τ_T^2	0.075	0.046	0.128
τ_I^2	0.648	0.628	0.668
ρ_S	0.854	0.436	0.986
ρ_T	0.442	0.072	0.814

for $j = 0, \ldots, 4$ as covariates. These five lag variables are expected to be highly correlated which may cause multicollinearity. In order to investigate this, we obtain the correlation matrix in Table 4. The maximum correlation, 0.92 is between lags 3 and 4 followed by 0.91 for lags 2 and 3. These correlation values are only moderate and these do not cause any problem in model fitting.

Table 4: Correlation matrix for the log of the five lag *SCMR* values.

	0	1	2	3	4
Lag 0	1.00	0.62	0.53	0.42	0.31
Lag 1	0.62	1.00	0.90	0.80	0.72
Lag 2	0.53	0.90	1.00	0.91	0.82
Lag 3	0.42	0.80	0.91	1.00	0.92
Lag 4	0.31	0.72	0.82	0.92	1.00

For the common regression part of the models we include the three socio-economic factors: *jsa*, *price*, and *popden*, and the log *SCMR* values of the cases for the current and four previous weeks. We exclude the environmental covariate NO_2 from this regression model since a significantly positive effect of NO_2 has already been identified in the model for the number of cases. Moreover, our modelling investigations including NO_2 in the model have not found a significantly positive effect of NO_2 in the model for death.

Each of the six models, M3-M7, was run for 100,000 iterations after an initial burn-in of 20,000 iterations. The samples collected after a thinning of 100 iterations to reduce autocorrelation in the MCMC samples produced the following model choice statistics reported in Table 5. This table shows that the adaptive model M7 is the best model according to both the DIC and WAIC.

Table 5: Model choice statistics for the five fitted models for the number of deaths.

	DIC	p.d	WAIC	p.w
M3 (Anova)	41125.0	6735.1	40067.7	4227.1
M4 (Separable)	40770.5	6111.7	40101.0	4071.6
M5 (AR)	40534.8	6092.1	39754.0	3980.4
M6 (localised)	40587.1	5929.8	40045.1	4036.4
M7 (Adaptive)	**38577.3**	4064.9	**38480.4**	3079.2

The chosen model M7 contains eight terms: *jsa, price, popden* and log *SCMR* with lags 0-4. It is of interest to see if any number of terms can be dropped for model simplification. Toward this end we fit a number of models using a backward elimination method to remove the lagged values of log *SCMR*. We have also considered various other models including the Base model with the three socio-economic covariates (*jsa, price* and *popden*) only. We have also fitted five models with each with one of the lagged log *SCMR* values to see if a single lag variable will be sufficient. Table 6 provides the model choice statistics for the different models. The table shows that the best model is the one with three socio-economic covariates and the *SCMR* with lags 0-3. This best model does not include the *SCMR* for the number of cases observed four weeks ago when the *SCMR* for the previous three weeks and the current week are included in the model.

Table 6: Model choice statistics for model selection of the Adaptive models for the number of deaths. The term 'Base' in the table refers to three covariates: *jsa*, *price*, *popden*.

M7 with terms	DIC	p.d	WAIC	p.w
Base	39904.7	5175.3	39665.1	3752.8
Base + Lag 4	39116.3	4791.3	38801.5	3421.3
Base + Lag 3	38899.7	4681.3	38579.6	3336.7
Base + Lag 2	38963.2	4595.2	38714.9	3334.6
Base + Lag 1	39023.6	4416.3	38875.7	3290.1
Base + Lag 0	38985.0	4369.3	38856.4	3274.3
Base + Lag 0, 1	38662.9	4040.8	38589.2	3082.8
Base + Lag 0, 1, 2	38593.6	4069.9	38504.5	3089.1
Base + Lag 0, 1, 2, 3	**38538.1**	4039.9	**38440.8**	3061.3
Base + Lag 0, 1, 2, 3, 4	38577.3	4064.9	38480.4	3079.2

The parameter estimates for the chosen version of the model M5 and M7 are presented in Table 7. The parameter estimates for the population density, lagged *SCMR* values for the current week and lagged weeks 2 and 3 are positively significant. However, *jsa*, *price* and the lagged *SCMR* value for the previous week are not significant as the 95% credible intervals include the zero values. The positive estimate of *jsa* captures the north-south divide in the death rates since higher observed death rates in the north were positively associated with the higher values of *jsa* seen in Figure 3. However,

the same conclusions cannot be drawn for the house price variable, *price*, since it correlated much less with covid death rate than *jsa*, see Figure 5. The table also shows that the spatial correlation (ρ_S) is estimated to be much higher than the temporal correlation (ρ_T) indicating stronger spatial than temporal association which makes sense for this highly contagious disease.

Table 7: Parameter estimates for the temporally auto-regressive CAR model, M5 and the adaptive model M7 fitted to the number of deaths.

	M5			M7		
	Median	2.5%	97.5%	Median	2.5%	97.5%
Intercept	-0.538	-2.143	0.909	-0.090	-1.395	1.016
jsa	0.019	-0.169	0.206	0.080	-0.046	0.208
price	-0.093	-0.362	0.207	-0.159	-0.368	0.081
popden	0.082	0.047	0.118	0.065	0.035	0.091
log(lag 0 of *SCMR*)	0.263	0.234	0.294	0.295	0.270	0.320
log(lag 1 of *SCMR*)	-0.028	-0.069	0.014	0.101	0.074	0.129
log(lag 2 of *SCMR*)	0.135	0.089	0.174	0.059	0.032	0.084
log(lag 3 of *SCMR*)	0.056	0.029	0.086	0.037	0.017	0.055
τ^2	3.222	3.053	3.401	0.153	0.126	0.186
ρ_S	0.701	0.650	0.750	0.993	0.991	0.995
ρ_T	0.001	0.000	0.003	0.001	0.000	0.007
τ_w^2	-	-	-	377.903	341.294	419.533

To compare the results of the Bayesian models we have also fitted non-Bayesian mixed models using the same covariates such as (*jsa, price, popden* and *SCMR* with lags 0-3) with region and week as unstructured random effects (week nested in region). These popular models are particular cases of Generalised Linear Mixed Models and are fitted using numerical methods such as Laplace or Gauss-Hermite quadrature approximations to deal with likelihoods involving bivariate, nested integrals, see e.g. Faraway (2006) and Wood (2016). One such approach taken here follows the modelling suggested by Rabe-Hesketh and Skrondal (2012) who looked at spatial data from the *Atlas of Cancer Mortality in the European Economic Community*. In our case, the likelihood based model fitting due to Rabe-Hesketh and Skrondal (2012) produced results which were quite similar and comparable to the size of the fixed effects reported in Table 7, although no structured spatial modelling was undertaken for this class of models and estimation methods. These models, however, performed poorly in model comparison using the AIC (Akaike, 1973) (51505.3) and BIC (Schwartz, 1978) (51571.2). These non-Bayesian model fitting methods do not produce the DIC and WAIC criteria values and hence are not included in our model comparison Table 5. However, the DIC criterion reported in Table 5 is comparable in magnitude to the AIC and BIC from non-Bayesian model fitting (Spiegelhalter et al., 2002).

5.3. Graphical exploration of the best fit model for cases and deaths

In this subsection we graphically explore the two best fit models. Figure 7 provides the observed and fitted average covid case and death rates. The fitted plots agree very closely with the corresponding observed plots both for the case and death rates. We do not include the maps of the standard deviations of the fitted values for brevity. Those standard deviations were smaller compared to those in the weekly temporal scale of the data because of the aggregation over the 36 weeks.

We explore the localised structure of the random effects implied by the chosen adaptive model in Figure 8. This figure plots the fitted values of the random effects

$$\hat{\psi}_k = \frac{1}{LT} \sum_{\ell=1}^{L} \sum_{t=}^{T} \psi_{kt}^{(\ell)},\ k = 1, \ldots, 313,$$

where $\psi_{kt}^{(\ell)}$ denote the ℓth MCMC sample of the random effect ψ_{kt} and L is the total number of saved MCMC samples. The highlighted borders are the borders where, according to the model, there are step changes in the *SMR* between the two neighbouring spatial units. Most of the step changes in the time averaged random effects surface are seen in the urban areas such as London, Birmingham, Manchester, Liverpool, and Nottingham. Note that the changes in the random effects changes the *SMR* values as per our model (5). Such a map with highlighted borders is potentially informative to the public health professionals who may search for potential contrasts in risk factors in the spatial units separated by the highlighted borders.

Notes for computing the highlighted borders: The adaptive `CARBayesST` model fitted object saves a summary of the localised structure in the `localised.structure` component. This component, see the documentation of the `CARBayesST` function `ST.CARadaptive`, is a list with two $K \times K$ matrices, `Wmedian` and `W99`, which summarises the estimated adjacency relationships. `Wmedian` contains the posterior median for each positive w_{ij} element estimated in the model for adjacent areal units, while `W99` contains the binary indicator variables for whether $P(w_{ij} < 0.5|\text{data}) > 0.99$. For both matrices, elements corresponding to non-adjacent pairs of areas have missing (NA) values.

To draw the highlighted borders we need to use the function `highlight.borders` from the `CARBayes` package. This function takes two arguments `border.locations` and `spdata`. The argument `border.locations` should be set as the `W99` matrix and the `spdata` argument should be set as the map polygon object used to draw the map. The map polygon object is the `sp` object as read from the shape files. The output of the `highlight.borders` is a spatial points data frame which has been plotted on the same plot as the original shape-file object which was passed to the `highlight.borders` function. For further details see the vignette of the `CARBayes` package, Lee (2021).

5.4. Forecasting the number of deaths

The chosen adaptive model M7 when fitted using `ST.CARadaptive` does not allow forecasting in future but the Anova and AR models do permit forecasting. Here we illustrate forecasting using

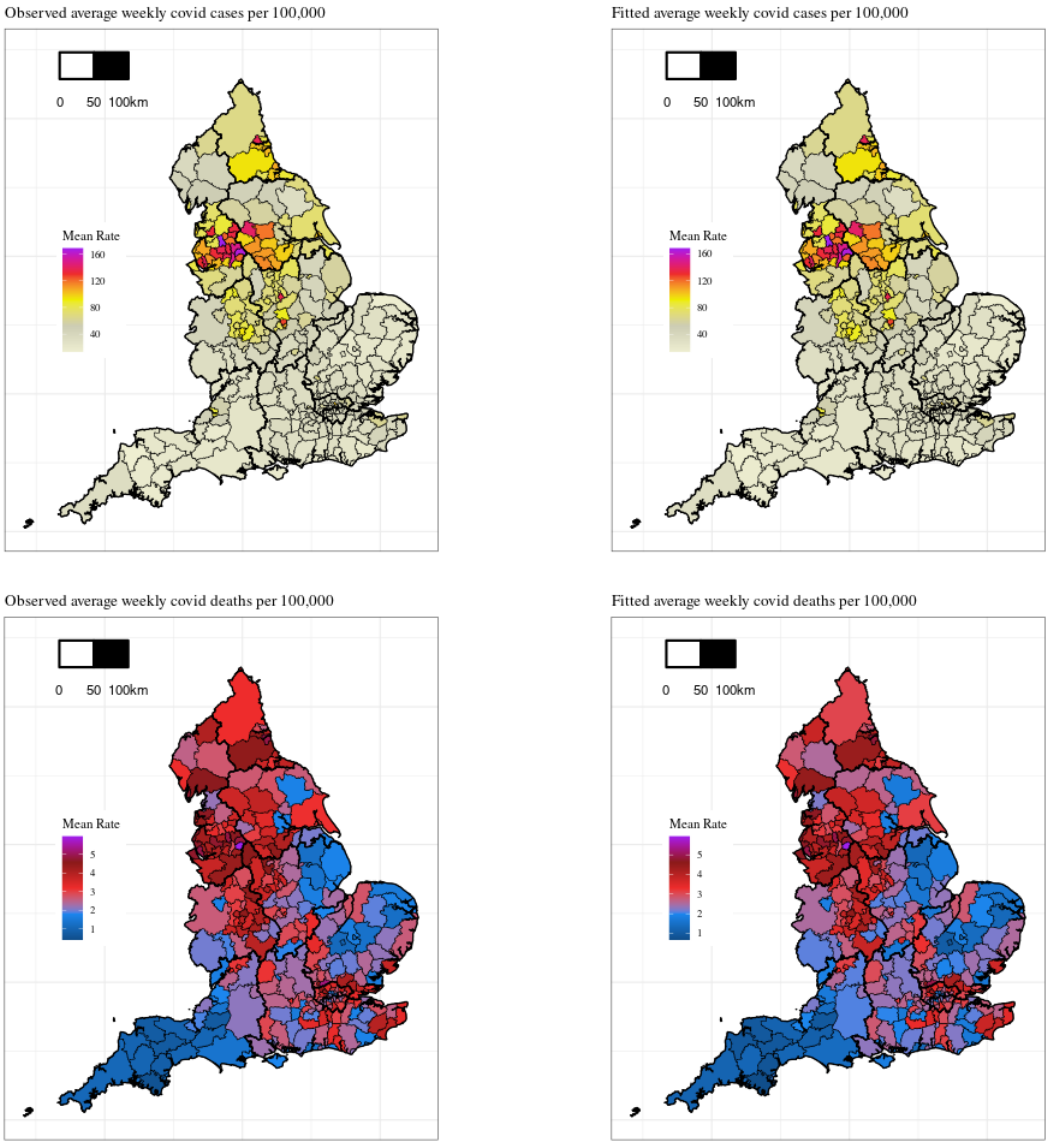

Figure 7: Maps of observed and fitted average covid case and death rates.

Random effects

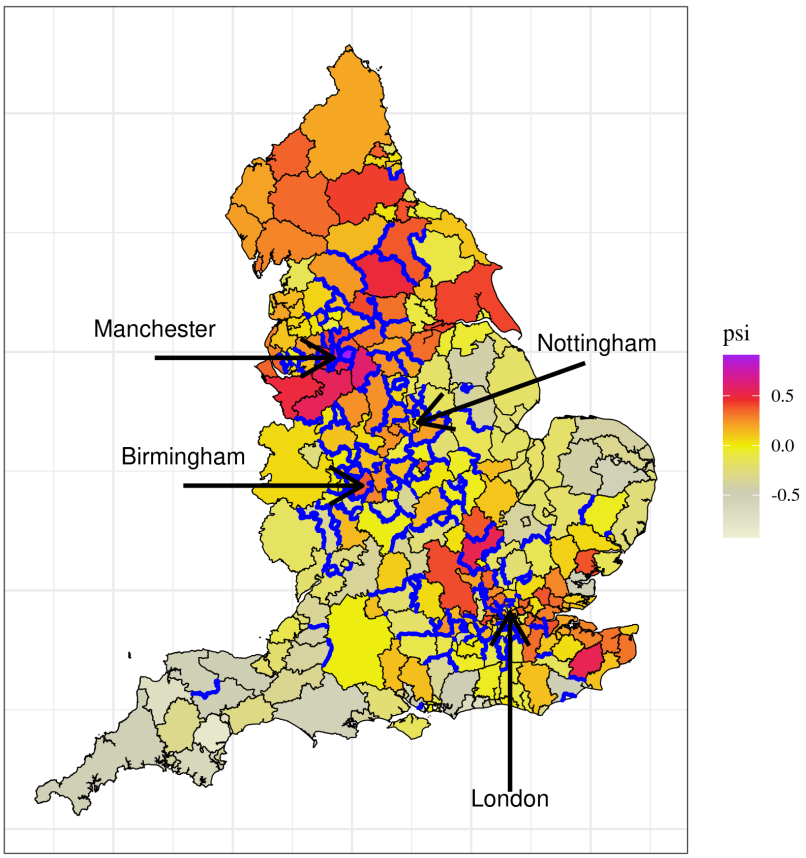

Figure 8: Highlighted borders where there are step changes in the death rates.

the Anova model M3 since we also illustrate forecasting using the chosen model C3 (Anova) for the cases. We illustrate forecasting of the number of deaths a week in advance for the final week number 46. For this, we fit the model M3 using data upto and including the week number 45. The fitted model is then used to forecast for week 46 - the very last week in our data set.

The observations and forecasts are plotted in Figure 9. These forecasts agree well with the observed data with an average mean absolute error (MAE) of 3.7 and a coverage rate of 99% for the associated 95% forecast intervals. The forecasts for the two areas of Cornwall (bottom left corner of the map) and Buckinghamshire (slightly north of London) seem to be slightly higher than the observed. However, this is not a systematic failure as the model has under predicted the death counts for the County Durham in the North. The standard deviations of the predictions (not shown for brevity) for these areas are higher as well - confirming a higher level of uncertainty in these predictions.

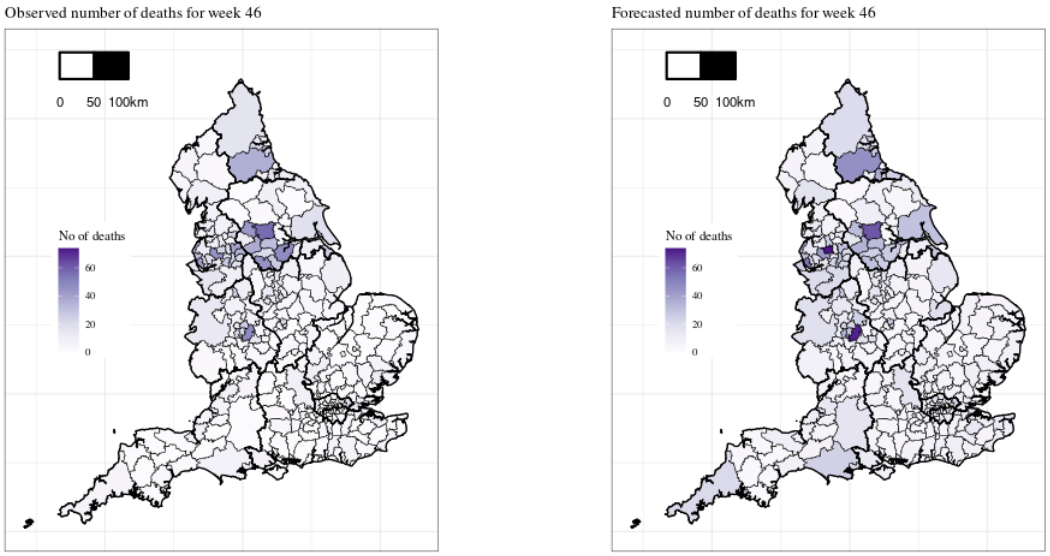

Figure 9: Observations and forecasts for the number of deaths in week 46.

Figure 10 plots the observed and fitted covid case and death rates by week. Here also we see a very high degree of agreement between the plots of the observed and fitted weekly summaries. Moreover, the upper and lower limits for the fitted values are very close to the fitted values which indicates a very high degree of accuracy due to the large sample size of the spatio-temporal data set.

We have also compared these forecasts with the ones obtained using a standard generalised linear models and also the BYM-AR1 model described in Section 4.6, using the R-INLA package[6]. This model also provides comparable fits and model validation error statistics with an MAE of 3.2

[6]https://www.r-inla.org/

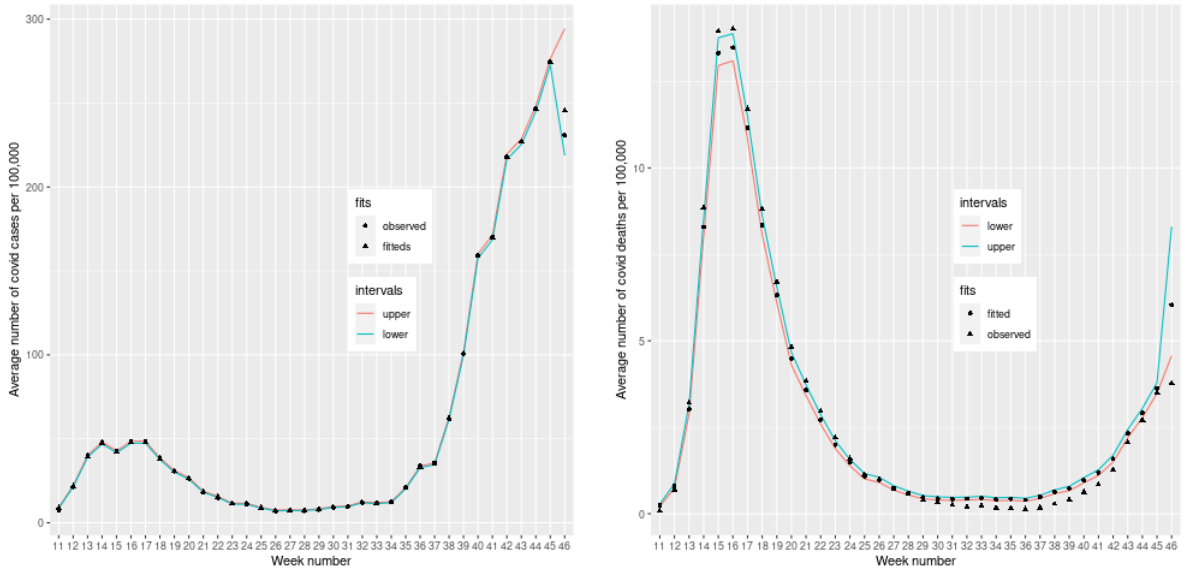

Figure 10: Observed and fitted covid case and death rates by week

but it results in a lower level of coverage, 66% for the associated 95% forecast intervals.

In principle, and in theory, it is possible to forecast the number of cases and deaths many more time steps ahead in the future. However, the uncertainty levels in the forecasts increase very rapidly when we attempt to make those forecasts more than a few weeks ahead of current time. This is because all the models are highly parameterised random effects models which are designed to provide a very good fit but those are not expected to perform accurately in predicting long-term future course of the pandemic. This is a clear limitation of the proposed models. Further investigation in developing forecasting methods in disease epidemiology may bear interesting and fruitful research.

6. Discussion

Methodologically this paper has developed a complete modelling toolkit for joint analysis of case and death rates of the Covid-19 pandemic. The paper has proposed novel methods to estimate the weekly expected case and death rates which are required in the very first step of practical disease mapping. Having completed this task, the primary focus of this paper has been to develop a joint model for cases and deaths as those are the two most important statistics that the authorities as well as the public would like to control to end the pandemic. Based on the state of the art modelling methodology in this area, the joint spatio-temporal model has been able to investigate the fixed effects of a number of relevant socio-economic and environmental factors simultaneously. The models have been shown to be very successful in their stated aims based on actual observed data for local administrative regions in England.

The fitted adaptive Bayesian model for number of weekly Covid-19 deaths has been able to

identify step changes in death rates of selected neighbouring areas in big cities such as London, Birmingham, Manchester and Nottingham. The public health authorities in the pairs of neighbouring areas, where these step changes were identified, are able to contrast the rates and thereby identify the possible causes of these step changes which can potentially reduce the death rate in the area with the higher value. Another potential use of this modelling work lies in the ability to predict the number of weekly deaths and cases at least a week in advance. Such advance warning can potentially be used to manage all kinds of services a public health authority is responsible for providing to its residents.

The modelling methodology developed in this article can be used to model other statistics of interests, e.g. weekly hospitalisation counts. For example, the methodology can be directly applied to conditionally model number of hospitalisation given the number of cases and then a marginal model for number of cases. Moreover, a hierarchical tri-variate joint model such as:

$$\begin{gathered}[\text{cases}] \\ [\text{hospitalisation} \mid \text{cases}], \\ [\text{deaths} \mid \text{hospitalisation, cases}],\end{gathered}$$

can be developed, where $[\cdot]$ denotes the probability distribution of its argument. Each model in the hierarchy can be parameterised by a number of shared socio-economic and environmental covariates.

The case study reported in this paper, however, has several limitations. For example, this study does not model case and death data for different age groups who are clearly experiencing different levels of severity. Further multivariate modelling with number of weekly cases and deaths for different age groups as the multivariate components will need to be developed and experimented. Such modelling, though attractive, has to tackle various challenging issues such as the presence of high number of zero deaths for individual age groups during the suppressed times of the pandemic during the summer months. Further sub-division of the spatial and/or temporal resolutions will also exacerbate the problem due to the sparsity of the data.

The data analysis presented here misses the more severe second peak of deaths in the UK during the months of January and February in 2021. It has not been possible to collect and process all the data in real-time to perform analysis of the latest data. The modelling framework developed here, however, can be applied to the more recent data sets once those have been compiled in future. In order to facilitate this we will publish the code used in the paper.

Being ecological in nature the modelling developments do not allow us to present the results as a cause-effect analysis would have done. The study presented here attempts to find significant statistical association between socio-economic and environmental factors on the pandemic. However, the models do reveal significant spatial disparities, i.e. clustering, in the rate of death. Specifically, the modelling results show a significant north-south divide with much higher rates of cases and deaths in the northern regions in England.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in: Petrov, B.N., Csáki, F. (Eds.), 2nd International Symposium on Information Theory, Budapest: Akadémiai Kiadó. pp. 267–281.

Anderson, C., Ryan, L.M., 2017. A comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in New South Wales, Australia. Int J Environ Res Public Health 14(2): 146, 10.3390/ijerph14020146. doi:10.3390/ijerph14020146.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2015. Hierarchical Modeling and Analysis for Spatial Data. 2nd ed., CRC Press, Boca Raton.

Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration with two applications in spatial statistics. Annals of the Institute of Statistics and Mathematics 43, 1–59.

Bland, M., 2000. An Introduction to Medical Statistics. 3rd ed., Oxford University Press: Oxford.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-temporal Bayesian Models with R - INLA. Chichester: John Wiley and Sons.

Böhning, D., Dietz, E., Schlattmann, P., 2000. Space-time mixture modeling of public health data. Statistics in Medicine 19, 2333–44. doi:10.1002/1097-0258(20000915/30)19:17/18<2333::AID-SIM573>3.0.CO;2-Q.

Boyle, P., 2008. Favorable trends in cancer mortality in the European Union but no room for complacency. Ann Oncol. 19, 605–6. doi:10.1093/annonc/mdm598.

Clayton, D., Hills, M., 1993. Statistical Models in Epidemiology. Oxford University Press: Oxford.

Faraway, J., 2006. Extending the Linear Model With R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Boca Raton: Chapman and Hall/CRC Press.

Finkenstädt, B., Held, L., Isham, V., 2007. Statistical Methods for Spatio-Temporal Systems. Boca Raton: Chapman & Hall/CRC.

Keiding, N., 1987. The method of expected number of deaths, 1786-1886-1986. International Statistical Review 55, 1–20.

Knorr-Held, L., 2000. Bayesian modelling of inseparable space-time variation in disease risk. Statistics in Medicine 19, 2555–2567. doi:https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-\#.

Konstantinoudis, G., Padellini, T., Bennett, J., Davies, B., Ezzati, M., Blangiardo, M., 2020. Long-term exposure to air-pollution and covid-19 mortality in england: a hierarchical spatial analysis. Environment International 146, 10.1016/j.envint.2020.106316. doi:10.1016/j.envint.2020.106316.

Last, J.M., 2000. A Dictionary of Epidemiology. 3rd ed., Oxford University Press: Oxford.

Lawson, A., Böhning, D., Lesaffre, E., Biggeri, A., Viel, J.F., Bertollini, R., 1999. Disease Mapping and Risk Assessment for Public Health. Chichester: Wiley.

Lawson, A.B., 2000. Cluster modelling of disease incidence via rjmcmc methods: a comparative evaluation. Statistics in Medicine 19, 2361–2376.

Lawson, A.B., 2018. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. Chapman & Hall: Boca Raton.

Lee, D., 2021. Carbayes version 5.2.3: An r package for spatial areal unit modelling with conditional autoregressive priors. Technical Report , –URL: https://cran.r-project.org/web/packages/CARBayes/vignettes/CARBayes.pdf.

Lee, D., Mukhopadhyay, S., Rushworth, A., Sahu, S.K., 2017. A rigorous statistical framework for spatio-temporal pollution prediction and estimation of its long-term impact on health. Biostatistics 18, 370–385. doi:10.1093/biostatistics/kxw048.

Lee, D., Robertson, C., Diogo, M., 2021. A spatio-temporal covid-19 syndromic surveillance tool for scotland using telehealth data. Spatial Statistics , NA.

Lee, D., Rushworth, A., Napier, G., 2018. Spatio-temporal areal unit modeling in r with conditional autoregressive priors using the carbayesst package. Journal of Statistical Software 84, 10.18637/jss.v084.i09.

Lee, D., Rushworth, A., Sahu, S.K., 2014. A bayesian localised conditional auto-regressive model for estimating the health effects of air pollution. Biometrics 70, 419–429.

Leroux, B.G., Lei, X., Breslow, N., 2000. Estimation of disease rates in small areas: A new mixed model for spatial dependence, in: Halloran, M.E., Berry, D. (Eds.), Statistical Models in Epidemiology, the Environment, and Clinical Trials. New York: Springer-Verlag, pp. 179–191. doi:http://dx.doi.org/10.1007/978-1-4612-1284-3_4.

Lilienfeld, A.M., Lilienfeld, D.E., 1980. Foundations of Epidemiology. 2nd ed., Oxford University Press: Oxford.

Miettinen, O.S., 1985. Theoretical Epidemiology. Principles of Occurrence Research in Medicine. Wiley: New York.

Mukhopadhyay, S., Sahu, S.K., 2018. A bayesian spatio-temporal model to estimate long term exposure to outdoor air pollution at coarser administrative geographies in England and Wales. Journal of the Royal Statistical Society, Series A 181, 465–486. doi:10.1111/rssa.12299.

Pickle, L.W., 2000. Exploring spatio-temporal patterns of mortality using mixed effects models. Statistics in Medicine 19, 2351–2263. doi:10.1002/1097-0258(20000915/30)19:17/18<2251::aid-sim567>3.0.co;2-m.

Polo, G., Acosta, C.M., Soler-Tovar, D., Villamil, J.F.P., Palencia, N.P., Penagos, M., Martinez, J.M., Bobadilla, J.N., Martin, L.V., Durán, S., Rodriguez, M., Carvajalino, C.M., Villamil, L.C., Ortiz1, E.B., 2020. Bayesian spatio-temporal modeling of covid-19: Inequalities on case-fatality risk. medRxiv preprint , –doi:https://doi.org/10.1101/2020.08.18.20171074.

Rabe-Hesketh, S., Skrondal, A., 2012. Multilevel and Longitudinal Modeling Using Stata, Volume II: Categorical Responses, Counts, and Survival. 3rd ed., College Station, TX: Stata Press.

Ramirez-Aldana, R., Gomez-Verjan, J.C., Bello-Chavolla, O.Y., 2020. Spatial analysis of covid-19 spread in iran: Insights into geographical and structural transmission determinants at a province level. Plos Neglected Tropical Diseases 14, 10.1371/journal.pntd.0008875. doi:10.1371/journal.pntd.0008875.

Rushworth, A., Lee, D., Sarran, C., 2017. An adaptive spatio-temporal smoothing model for estimating trends and step changes in disease risk. Journal of the Royal Statistical Society, Series C 66, 141–157.

Schwartz, G.E., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Snow, J., 1854. On the Mode of Communication of Cholera. 2nd ed., London: Churchill Livingstone.

Spiegelhalter, S.D., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society B 64, 583–639.

Wah, W., Ahern, S., Earnest, A., 2020. A systematic review of bayesian spatial–temporal models on cancer incidence and mortality. Int J Public Health 65, 673–682. doi:https://doi.org/10.1007/s00038-020-01384-5.

Watanabe, S., 2010. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research 11, 3571–3594.

Wood, S., 2016. Generalized Additive Models: An Introduction with R. Boca Raton: Chapman and Hall.