A Review of Spatiotemporal Models for Count Data in R Packages. A Case Study of COVID-19 Data

María Victoria Ibañez, Marina Martínez-García, Amelia Simó.

mibanez@uji.es martigar@uji.es simo@uji.es

Department of Mathematics-IMAC. Universitat Jaume I. Castelló. Spain

March 9, 2021

Abstract

Several R-packages to deal with spatiotemporal count areal data are currently available. Each package is based on different models and/or statistical methodologies. Unfortunately, results generated by these models are rarely comparable due to differences in notation and methods which ultimately hinder the robustness and effectiveness of such models. These models have become crucial nowadays, because their ability to analyze COVID-19 related data. This prompted us to revisit three spatiotemporal count models and monitor their performance under the same conditions (on a same data set and under a consensus notation).

The three different models assessed are representative of the main two extended methodologies used to analyse spatiotemporal count data i.e., the classical approach (based on Penalised Likelihood meanwhile and based to Estimating Equations) and the Bayesian point of view.

As a case study, we have used these packages to analyze COVID-19 data corresponding to 24 health regions within the Valencian Community, Spain. In particular, daily COVID-19 positive individuals are used to predict daily hospitalisations. Data corresponds to the period from 28 of June to 13 of December 2020. keyword COVID-19; count data; spatiotemporal models; R packages

1 Introduction

Since December 2019, when the first cases of COVID-19 were reported in Wuhan, China, the SARS-CoV-2 virus has spread worldwide. Up to now the virus has caused more than 100 million infections and 2,16 millions of deaths around the
world and, currently, it is impossible to predict how many people will be affected by it.

An effective way to control the spread of the infection is to understand and predict not only the number of cases and COVID-19 severity in a given area at a given time but also where and when the spread may become uncontrollable bringing health care systems to the brink of collapse due to an increase of the number of patients needing hospitalisation. Robust prediction models are therefore vital to support decisions on population and community-level interventions to control the spread of the virus and to prevent the collapse of health services.

The robustness of a model is intrinsically linked to the data quality, that is, the availability of reliable and comparable count data from across the different countries world-wide. Moreover, given the contribution of people’s motility to the spread of the virus, these models should be able to adjust for people’s motility within and between neighbouring health regions or constituencies as well as nation-wide and/or worldwide.

Epidemic models to study the spread of infectious diseases date back to the beginning of the twentieth century. The susceptible-infected-recovered (SIR) models developed by Kermack and McKendrick ([Kermack and McKendrick(1927)]) were the first mathematical models developed to study the transmission dynamics of infectious diseases. The SIR and SEIR (Susceptible, Exposed, Infectious and Removed) models have been improved and used for analysing and characterising the COVID-19 epidemic ([Fang et al.(2020)Fang, Nie, and Penny], Kucharski et al.(2020)Kucharski, Russell, Diamond, Liu, Edmunds, Funk, Eggo, Sun, Jit, Munday, et al., Tang et al.(2020)Tang, Wang, Li, Bragazzi, Tang, Xiao, and Wu) also in Spain ([Guirao(2020), López and Rodo(2020)]). Predicting the future course of epidemics has been another great challenge through time and have become particularly challenging with the rise of new infectious diseases [Held and Meyer(2019)].

In addition to the crucial role played by those epidemic models described above, other models are also being adapted to examine different aspects of the COVID-19 pandemic [Fronterre et al.(2020)Fronterre, Read, Rowlingson, Bridgen, Alderton, Diggle, and Jewell, Dunbar and Held(2020)]. In particular, mathematical modelling of patient hospitalisation is essential as it may help raising awareness of a possible collapse of the health care systems. Models for this type of data are rather different from the epidemic models (SIR and SEIR) as the prediction of hospitalisations requires previously obtained COVID-19 data such as the number of people tested and/or infected, that is, count data.

It is well known that for count data modelling, the linear model with normal assumption is not appropriate and generalised linear models with Poisson, binomial or negative binomial distributions should be used ([Nelder and Wedderburn(1972)]).

Historically, spatial and temporal models for count data are needed in a variety of settings: agriculture production ([Besag and Higdon(1999)]), fishing captures ([Paradinas et al.(2017)Paradinas, Conesa, López-Quírez, and Bellido]), volcanic eruptions ([Gusev(2008)]), crime counts ([Law et al.(2014)Law, Quick, and Chan]), cases of pulmonary disease ([Choi et al.(2011)Choi, Lawson, Cai, and Hossain]), etc.

The reasons to model these data are diverse and can range from estimating the effect of a risk factor to a response, identifying clusters of contiguous areas.
units or to forecasting future observations. To deal with this array of scenarios comprising spatiotemporal data different modelling strategies have been proposed. Strategies become more complex when the aim is to build a spatiotemporal model for the joint analysis of different 58 variables which include specific and shared spatial and temporal effects \cite{Gomez-Rubio2019}.

A common challenge when modelling count data is that of spatiotemporal autocorrelation, namely that observations from geographically close areal units and temporally close time periods tend to have more similar values than units and time periods that are further apart. In addition, due to the complicated correlation structure, the parameter estimation is not straightforward and different approaches are needed.

In the context of COVID-19 data we have to manage observations over time in several geographical areas. In particular, this review deals with areal unit data, which is a type of data where observations are related to a set of K contiguous but non-overlapping spatial areal units, such as health departments. Each observation relates to an entire areal unit and in this case referred as aggregated count measures for the entire unit. So we have a spatiotemporal count data (number of new infected people in a health department) to predict the number of COVID-19 hospitalisations.

The diversity of applications, data types and conceptual approaches leads us to a broad spatiotemporal modelling literature. Two excellent books that provide a gradual entry to the methodological aspects of spatiotemporal statistics and outline some of the standard techniques used in this area are \cite{Cressie2015} and \cite{Wikel2019}. An overview of different spatiotemporal modelling approaches can also be found in \cite{Anderson2017}.

Here we revisit three different models which are representative of the three more extended methodologies to analyse this type of data. The two first follow the classic statistical paradigm and the last the Bayesian point of view \cite{Berger2013}. Of the two classical approaches one is based on Penalised Likelihood meanwhile and the other one on Estimating Equations \cite{Liang1986}.

This work does not claim to provide a comprehensive coverage of all existing methods to deal with count data but to describe and compare the three most relevant approaches that can be found implemented R packages \cite{R Core Team 2020}.

The performance of the existing models to predict hospitalisation can only be fully assessed and compared when each model is tested against the same data set under a unique consensus. So, we applied new daily positive cases and daily hospitalisations on the 24 health department where Valencian Community, Spain, is divided to the models above explained.

The article is organised as follows: in Section 2 a descriptive analysis of the data is perform. Following by the mathematical details of the three models, Section 3. Then the application to COVID-19 data of the three models, Section 4. Finally the conclusions are discussed in Section 5.

In this section, we describe a spatiotemporal approach to monitoring COVID-19 (daily new cases and total number of people hospitalised) on Valencia Community, Spain, we explore some possible temporal biases that data may have.
This section begins by describing the data sources, before describing the analytic model and fitting process.

The data set comprises the number of daily new positive cases of COVID-19 (tested via PCR (polymerase chain reaction) or the antigen test), and the daily number of people at hospital by COVID-19 (daily hospitalisations) in the Valencian Community, that comprises a big area in the eastern part of Spain, during 533 days. This area is organised in $K = 24$ different health departments, and data set is conformed with both temporal series (hospitalised/new positive cases) for all departments. The spatial distribution of the health departments so as their population ($\times 100000$ people) are shown in Figure 1, as can be observed the populations across regions is very heterogeneous.

Figure 1: Spatial distribution of the 24 health departments with their population, unit 100,000

The number of new daily positive cases by health department is published regularly on an open data platform of the Generalitat Valenciana ([GVA()]). However, the number of daily hospitalisations is public only at the aggregated level of the entire Valencian Community, so, as a result we cannot show maps nor identifiable information of this variable. As an illustration, in this work we will show the data related to eight of the 24 health departments. These departments have been anonymized, and will be used to illustrate all the steps of the different analysis. However, the models are fitted using the whole data set and the estimations of the parameters and the goodness of fit measurements showed in the paper are related to all of them.

Although these data may be subject to temporal biases due to changing testing regimes, among other problems, the spatial weekly case prevalence (number of cases divided by population size) for weeks 27, 38, 49 plotted in Figure 2 shows strong variation over Valencian Community.
The goal of this paper is to predict the number of hospitalisation in time given the number of verified infected people in each of the 24 regions. Figure 3 a) and b) shows time series plots for both variables of interest; the number of daily new positive cases and the number of daily hospitalisations per health department. From now on, with daily hospitalizations we will mean total number of people at hospital by COVID-19 each day. Figure 3 a) shows a peak of new cases in late August, and another in mid-November, two and a half moths apart, and the same temporal pattern is showed in the series of hospitalisations. Figure 3 c) shows the total number of new positive cases and hospitalisations per day (adding the values of the 24 health departments). From Figure 3 c) we can see that there is a time lag between both time series.
Figure 3: Time series plot for (a) new positive cases; (b) daily hospitalisations, one time point for each of the 24 health department. Fig (c) shows the sum of the 24 health departments each day.

To estimate the delay between both time series a cross-correlogram (see Figure 4) have been used. It plots a measure of similarity of both time series (in these case Pearson correlation), as a function of the displacement (days) of daily positives relative to the daily hospitalisations. Figure 4 shows that there are two maximums in the correlation at the time lag of 9 and 5 days.

These data may be subject to temporal biases due to under reporting on weekend and/or non-working days. Figure 5 a) shows a great variability in number of positives, depending on day of the week. As we can see in Figure 5 b) this effect does not hold for number of daily hospitalisations. So, this effect is an artefact, due to when the official data is reported, rather than a real effect of the virus. To minimise this effect, henceforth, we will take as a daily hospitalisations, the mean of the last 4 days. By doing this smoothing we reduce this effect, as we can see from Figure 5 c).

Finally, both effects are corrected. Figure 6 shows the smoothed number of positive daily cases (in red) jointly with the number of people hospitalised
Figure 4: Cross-correlogram, between daily new positives and daily hospitalisations. It shows the Pearson correlation between both series, as a function of the displacement (days) of daily positives relative to the daily hospitalisations.

by COVID-19 (in blue) for the eight illustrative health departments, with a time delay of 9 days between both series. The spatial adjacency matrix among these departments is shown in Table 1. In this table values equal to 1 signal neighbouring regions, i.e. those that share a geographical border.

To conclude, when looking for the best model, we explored the relationship between the mean and the variance of the hospitalisation data collected at each instant of time. Figure 7 shows that there exist a potential relationship between mean and variance. This relationship is crucial for the modelling as we will see in the following sections.

2 Models and methods

Throughout this section $Y_{kt}$ will denote the observation taken at the $k$-th areal unit $S_k$ at time $t$, for $k = 1, \ldots, K$ and $t = 1, \ldots, N$. Then, $Y_t = (Y_{1t}, \ldots, Y_{Kt})$, $(t = 1, \ldots, N)$, will be a count spatiotemporal serie, i.e. count data are recorded at the areal units on consecutive discrete time periods. We assume that we have also space-time varying covariates $X_t = (X_{1t}, \ldots, X_{Kt})$ recorded at the same times and locations. Our main objective will be to predict future observations
of the spatiotemporal time series $Y_t$, by taking into account the spatiotemporal covariates $X_t$ and the spatial and temporal relations between the observations.

Among the R packages that implement models and procedures to work with spatiotemporal data, we are going to review three packages that allow to work with count areal data: package **surveillance** ([Meyer et al.](Meyer et al., 2016)Meyer, Held, and Höhle), package **Mcglm** ([Bonat](Bonat, 2018)) and package **CARBayesST** ([Lee et al.](Lee et al., 2018)Lee, Rushworth, and Napier). Package **surveillance** will be reviewed in Section 2.1 and applied to the COVID-19 data set in Section 3.1. The model implemented in this package is based on a likelihood model, working on the classical methodology. Package **Mcglm** ([Bonat](Bonat, 2018)) will be reviewed in Section 2.2 and applied to the COVID-19 data set in Section 3.2. The model implemented in this package is based on estimating equations, working on the classical methodology. Finally, package **CARBayesST** ([Lee et al.](Lee et al., 2018)Lee, Rushworth, and Napier) based on the Bayesian methodology, will be reviewed in Section 2.3 and applied to the COVID-19 data set in Section 3.3.

We will see that the second model is the most flexible for modelling $Y_{kt}$, allowing any type of function of time $t$ in the expression of the mean and a huge variety of spatiotemporal neighbourhoods to model both the mean and
|    | A   | B   | C   | D   | E   | F   | G   | H   |
|----|-----|-----|-----|-----|-----|-----|-----|-----|
| A  | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   |
| B  | 1   | 0   | 1   | 1   | 0   | 0   | 0   | 0   |
| C  | 0   | 1   | 0   | 1   | 1   | 1   | 0   |     |
| D  | 0   | 1   | 1   | 0   | 1   | 1   | 1   |     |
| E  | 0   | 0   | 1   | 1   | 0   | 1   | 1   |     |
| F  | 0   | 0   | 1   | 1   | 1   | 0   | 1   |     |
| G  | 0   | 0   | 1   | 1   | 1   | 1   | 0   | 0   |
| H  | 0   | 0   | 0   | 1   | 1   | 1   | 0   | 0   |

Table 1: Spatial adjacency matrix of the health departments used to illustrate the paper.

Figure 6: Time series of new positives (in red) vs hospitalisations (in blue) in eight different health departments, with a time lag of 9 days in the series of new positive cases, and a smooth of 4 days.

the variance-covariance matrix (when necessary). In the first model we can also use any function of time in the expression of the mean, but the temporal relationship with the neighbours only allows an autorregressive (AR) structure of order
This model allows to choose between a Poisson distribution for modelling the observations or a Negative Binomial model when data show overdispersion. Regarding to modelling the mean as a function of $t$, in the third model we only have the possibility of linear relationships. In this third model, spatiotemporal relations between the observations can be introduced in the model of the mean using random effects with different correlation structures.

2.1 Endemic-epidemic models. R package surveillance ([Meyer et al.(2016)Meyer, Held, and H"ohle]).

The endemic-epidemic (EE) models, are a class of statistical time series models for multivariate surveillance counts proposed by [Held et al.(2005)Held, H"ohle, and Hofmann] and extended in [Paul and Held(2011) and Meyer et al.(2014)Meyer, Held, et al.]. It has been applied to a multitude of infectious diseases (see [Dunbar and Held(2020)] for references there in) such as Influenza [Paul et al.(2008)Paul, Held, and Toschke], Norovirus [Held et al.(2017)Held, Meyer, and Bracher], and COVID-19 [Fronterre et al.(2020)Fronterre, Read, Giuliani et al.(2020)Giuliani, Dickson, Espa, and Santi]. In fact, a regularly updated table of use cases is maintained by S. Meyer at [https://github.com/rforge/surveillance/blob/master/www/applications_EE.csv](https://github.com/rforge/surveillance/blob/master/www/applications_EE.csv).

On its current formulation and implementation in the R package surveillance ([Meyer et al.(2016)Meyer, Held, and H"ohle]), the EE framework uses incidence from the preceding week, $t-1$, to explain the incidence in week $t$. So, the counts, $Y_{kt}|Y_{t-1}$, are assumed to be Poisson or Negative Binomial distributed.
with conditional mean:

\[ \mu_{kt} = e_{kt}\nu_{kt} + \lambda_k Y_{k,t-1} + \phi_k \sum_{q \neq k} w_{qk} Y_{q,t-1}, \quad \nu_{kt}, \lambda_k, \phi_k > 0, \]

and overdispersion parameter, in the Negative Binomial case, \( \psi_k > 0 \).

The first component of the summation is called endemic component and captures information not directly linked to observed cases from the previous day. This component can cover exogenous factors such as temporal trends, seasonality, sociodemography, and/or population. As an example, in spatial applications, \( e_{kt} \) can refer to the fraction of the population living in region \( k \) at time \( t \). The remaining terms in Eq. 1 constitute the epidemic component and describe how the incidence in region \( k \) is linked to previous cases in the same and in the adjacent regions. The two terms of this epidemic component are usually denoted as ‘autoregressive’ and ‘spatiotemporal’ component, respectively.

The parameters \( \nu_{kt}, \lambda_k \) and \( \phi_k \) are constrained to be non-negative and can be modelled by allowing for log-linear predictors in all three components, as sine-cosine terms to account for seasonality [Held and Paul(2012)], long-term temporal trends or/and covariates [Bauer and Wakefield(2018)]Cheng et al.(2016)Cheng, Lu, Wu, Liu, and Huang.

\[
\begin{align*}
\log(\nu_{kt}) &= \alpha^{(\nu)} + b^{(\nu)}_{k} + \beta^{(\nu)} z^{(\nu)}_{kt} \\
\log(\lambda_{k}) &= \alpha^{(\lambda)} + b^{(\lambda)}_{k} + \beta^{(\lambda)} z^{(\lambda)}_{kt} \\
\log(\phi_{k}) &= \alpha^{(\phi)} + b^{(\phi)}_{k} + \beta^{(\phi)} z^{(\phi)}_{kt}
\end{align*}
\]

This form allows for fixed intercepts \( \alpha^{(\cdot)} \), region-specific intercepts \( b^{(\cdot)}_{k} \), and exogenous covariates \( z^{(\cdot)}_{kt} \) in each model compartment. Population fraction, population density, border effects, etc. can be used as covariates. The region-specific intercepts \( b^{(\cdot)}_{k} \), can be treated as fixed effects or as random effects accounting for heterogeneity between the regions. When they are treated as random effects, they are assumed to be independent and identically distributed across \( k \), but can be correlated across the model components, following a Gaussian distribution:

\[ b_k := (b^{(\nu)}_{k}, b^{(\lambda)}_{k}, b^{(\phi)}_{k}) \sim N((0, 0, 0), \Sigma_k). \]

We will see this part in a more detailed way in Section 3.1.

Mamimun likelihood (ML) estimates are obtained using penalised likelihood approaches.

This basic model, has been extended to cover other different aspects of disease modelling (see [Dunbar and Held(2020)] for references there in). Recent extensions include methodology to adjust for under reporting ([Dunbar and Held(2020], Bracher(2019a)], or to allow different lags in the auto-regressive part of the model (package hhh4addon ([Bracher(2019b]Bracher and Held(2017)]), modelling the conditional mean \( Y_{kt}|Y_{t-1}, \cdots Y_{t-D} \) as:

\[ \mu_{kt} = e_{kt}\nu_{kt} + \lambda_k \sum_{d=1}^{D} \sum_{q} w_{qk} Y_{q,t-d}, \quad \nu_{kt}, \lambda_k > 0, \]
where $D$ is the maximum lag considered.

In [Meyer and Held(2017)], authors extend the basic endemic-epidemic spatiotemporal model to fit multivariate time series of counts $y_{gkt}$ stratified by (age) groups in addition to spatial regions. So they define a contact matrix $C = (c_{g'g})$, where $c_{g'g} \geq 0$ quantifies the average number of contacts of an individual of group $g'$ with individuals of group $g$, and the spatiotemporal model is now modelled as:

\[
\mu_{gkt} = e_{gkt}\nu_{gkt} + \lambda_{gkt} \sum_{g'q} c_{g'g'} w_{g'q} Y_{g'q,t-1}, \quad \nu_{kt}, \lambda_k > 0,
\]

where both the endemic and epidemic predictors may gain group-specific effects.

This model is implemented in the R-package hhh4contacts ([Bracher(2019c)]).

Forecast

Package surveillance uses the function hhh4 to fit the models, and implements the function oneStepAhead() that computes successive one-step-ahead predictions for the fitted model, providing also confident intervals for the predictions and plot methods. The associated scores-method computes a number of (strictly) proper scoring rules based on such one-step-ahead predictions; see [Paul and Held(2011)] for details.

A discussion of suitable measures to evaluate the quality of a point forecast can be found in [Gneiting(2011)] and several scoring rules based on the one-step-ahead predictions [Paul and Held(2011)] are implemented in the function scores, although we will consider the root mean squared error (RMSE). Other functions implemented in the package related with the function oneStepAhead() is the function calibrationTest that implements calibration tests for Poisson or Negative Binomial predictions of count data are based on proper scoring rules and described in detail in [Wei and Held(2014)].

Long term predictions do not have much sense in our context because we do not know the long term evolution of the covariates.

2.2 Multivariate covariance generalised linear models. R package Mcglm

Under the same previous assumption of predicting $Y_t = (Y_{1t}, \ldots, Y_{Kt})$ in terms of spatiotemporal correlations and $X_t = (X_{1t}, \ldots, X_{Kt})$ covariates we can use the multivariate covariance generalised linear model (McGLMs) introduced in [Bonat and Jørgensen(2016)]. This model is a general and flexible statistical model to deal with multivariate count data that explicitly models the marginal covariance matrix combining a covariance link function and a matrix linear predictor composed of known matrices.

Let $Y_{K \times N} = \{Y_1, \ldots, Y_N\}$ be the outcome matrix and let $M_{K \times N} = \{\mu_1, \ldots, \mu_N\}$ denote the corresponding matrix of expected values.

Given $\Sigma_t$ the $K \times K$ covariance matrix within the response variable $Y_t$ for $t = 1, \ldots, N$ and $\Sigma_b$ be the $N \times N$ correlation matrix whose components denote the correlation between outcomes, the McGLM as proposed by
Bonat and Jørgensen(2016) is given by

\[ E(Y) = (\mu_1, \cdots, \mu_N) = (g_1^{-1}(X_1\beta_1), \ldots, g_N^{-1}(X_N\beta_N)) \]  

\[ Var(Y) = C = \Sigma_t \bigotimes G \Sigma_b, \]

where \( g_t \) are monotonic differentiable link functions, \( X_t \) denotes a \( K \times k_t \) design matrix, \( \beta_t \) is a regression parameter vector to be estimated, and

\[ \Sigma_N \bigotimes \Sigma_b = Bdiag(\hat{\Sigma}_1, \cdots, \hat{\Sigma}_N)(\Sigma_b \bigotimes I)Bdiag(\hat{\Sigma}^{-T}_1, \cdots, \hat{\Sigma}^{-T}_N) \]

is the generalised Kronecker product [Martinez-Beneito(2013)]. The matrix \( \hat{\Sigma}_t \) denotes the lower triangular matrix of the Cholesky decomposition of \( \Sigma_t \). The operator \( Bdiag \) denotes a block diagonal matrix, and \( I \) denotes an \( K \times K \) identity matrix.

A key point for specifying McGLMs for mixed types of outcomes is the specification of the covariance matrix within outcomes \( \Sigma_k \). Following [Bonat and Jørgensen(2016)], we define the covariance within outcomes by

\[ \Sigma_t = V(\mu_t; p_t)^{\frac{1}{2}}(\Omega(\tau_t))V(\mu_t; p_t)^{\frac{1}{2}}, \]

for modelling count outcomes they propose to adopt the Poisson-Tweedie dispersion function [Jørgensen and Kokonendji(2016)] so that

\[ V(\mu_t; p_t) = diag(\mu_t^{p_t}) \]

is a diagonal matrix whose main entries are given by the power variance function. The Poisson-Tweedie family of distributions provide a rich class of models to deal with count outcomes, since many important distributions appear as special cases, examples include the Hermite (\( p = 0 \)), Neyman Type A (\( p = 1 \)), Negative Binomial (\( p = 2 \)) and Poisson-inverse Gaussian (\( p = 3 \)).

The dispersion matrix \( \Omega(\tau_t) \) describes the part of the covariance within outcomes that does not depend on the mean structure. Jørgensen et al. [Jørgensen and Kokonendji(2016)] among others, propose to model the dispersion matrix using a matrix linear predictor combined with a covariance link function, i.e.

\[ h(\Omega(\tau_t)) = \tau_{00}Z_{00} + \cdots + \tau_{0D}Z_{0D}, \]

where \( h \) is the covariance link function, \( Z_{td} \) with \( d = 0, \cdots, D \) are known matrices reflecting the covariance structure within the response variable \( Y_t \), and \( \tau_t = (\tau_{t0}, \cdots, \tau_{tD}) \) is a \( (D + 1) \times 1 \) vector of dispersion parameters.

McGLMs are fitted based on the estimating function approach described in detail by [Bonat and Jørgensen(2016)] and [Jørgensen and Knudsen(2004)]. A general overview of the algorithm and the asymptotic distribution of the estimating function estimators can be found in [Bonat(2018)]. As a method to select the components of the matrix linear predictor (variable selection), the
score information criterion (SIC) is proposed. This is an important tool to assist with the selection of the linear and matrix linear predictors components, but unfortunately it is less useful for comparing models fitted using different link, variance or covariance functions.

**Forecasting**
Unfortunately the mcglm package does not have implemented any function to predict future observations. If we don’t know to implement any more sophisticated routine, once estimated the model, we can use the function mc link function that returns the inverse of the link function applied on the linear predictor i.e. $\mu = g^{-1}(X\beta)$, as an approximation of the predictions sought.

### 2.3 Bayesian Hierarchical generalised linear models. CARst package

A great variety of spatiotemporal models for count data using generalised linear models (GLM) can be found on the Bayesian literature. To modelling these data a hierarchical model with spatiotemporal structured prior distributions is used. The spatiotemporal structure is modelled via sets of autocorrelated random effects with conditional autoregressive and its spatiotemporal extensions priors. An excellent review can be found in Lee et al. (2018). These methods have had a remarkable development, especially in disease mapping, thanks to the availability of estimation methods based on Monte Carlo Markov Chain. With respect to the software packages for implementing these models, although a great quantity of software packages can be found for implementing purely spatial models, such as BUGS Lunn et al. (2009) and R-INLA Rue et al. (2009), software for spatiotemporal modelling is much less well developed, and mainly focused on geostatistical data. This was the motivation to developing the R package CARBayesST Lee et al. (2018). This package can fit several models for count data with different spatiotemporal structures. A useful tutorial is provided by Lee (2020).

The general Bayesian Hierarchical model for spatiotemporal count data is as following:

\[
Y_{kt} \sim f(y_{kt}|\mu_{kt}, \nu) \quad (9)
\]

\[
g(\mu_{kt}) = X_{kt}\beta + \psi_{kt} \quad (10)
\]

\[
\beta \sim N(\mu_{\beta}, \Sigma_{\beta}). \quad (11)
\]

The probability function $f$ is in the exponential family (not necessarily a Gaussian distribution), $\beta$ is the vector of covariate regression parameters, and a multi-variate Gaussian prior is assumed. $g$ can be any monotonic differentiable link function, and $\psi_{kt}$ is a latent component for areal unit $k$ and time period $t$ encompassing one or more sets of spatiotemporally autocorrelated, we denote $\psi_t = (\psi_{1t}, \ldots, \psi_{Kt})$. 

---

14
In this paper we are just focusing on the models implemented in the CARBayesST package. In this package binomial, Gaussian and Poisson data models can be used for the first level of the model, $f$ in Eq. 10, and different spatiotemporal structures for $\psi_{kt}$ in Eq. 11 are given.

All models implemented in this package use random effects to introduce spatial autocorrelation into the response variable. For this purpose, CAR-type prior distributions and their space-time extensions are used. Spatial autocorrelation is induced via a non-negative symmetric matrix of adjacency $W = (w_{kj})$, where $w_{kj}$ represents the spatial closeness between units ($S_k, S_j$). Larger valued elements represent spatial closeness between the two areas in question, and spatially autocorrelated random effects whereas zero values correspond to areas that are not spatially close are conditionally independent random effects given the remaining. The models are outlined in Table 2 extracted of [Lee et al.(2018) Lee, Rushworth, and Napier]

In all cases inference is based on Markov chain Monte Carlo (MCMC) simulation.

**Forecasting**

In order to predict future observations of the response variable, simulations from the posterior predictive distribution (the distribution of possible unobserved values conditional on the observed values) have to be obtained.

This predictive density can be approximated by Monte Carlo integration. If we denote by $\theta$ the vector with all the parameters of the model:

$$f(Y_{N+h}|(Y_1, \ldots, Y_N)) \approx \frac{1}{n} \sum_T^n f(\hat{Y}_{N+h}|\theta^{(j)}, (Y_1, \ldots, Y_N)).$$

If a representative value is wanted, the mean of the predictive density can be obtained taking into account the property that $E_{\theta}E(Y_{N+h} | \theta)$ and approxi-
mating again the mean with respect to $\theta$:

$$E(Y_{N+h}|(Y_1, \ldots, Y_N)) = \frac{1}{n} \sum_{j} E(Y_{N+h}|\theta^{(j)}, (Y_1, \ldots, Y_N))$$

Unfortunately the CARBayesST package has not implemented a function to predict future observations, for $h=1$ is easily implemented using the samples of the posterior distributions of the parameters. For $h > 1$ the simulation of the posterior distribution of the random effects should be implemented. For non experts in R programming users just a approximations can be obtained approximating the value of the random effect by the that of the previous prediction. This approximation will be used in Section 3.3.

3 Application

In this section all the models reviewed in Section 2 are applied to COVID-19 data described in Section ???. Following with the notation introduced, $Y_t = (Y_{1t}, \ldots, Y_{K_t})$ refers to the observed data of hospitalisations at time $t$ in the $K$ health departments of Valencian Community, $X_t = (X_{1t}, \ldots, X_{K_t})$ refers to the positive cases at time $t - lag$, where lag is the lag previously determined and $p_k$ is the population in risk in the $k$-th health department.

Since our primary goal is forecasting, the mean squared errors of the predictions (RMSE) up to a five-day horizon are calculated. This is a classical approach to measure their performance. If we want to use real data of positive cases the horizon of prediction is limited, taking a horizon of five is considered enough, within our possibilities, to prevent the collapse of hospitals.

Each model is adjusted using the corresponding statistical methodology included in its corresponding R package. Additional specific measures of goodness of fit are given, these other measures will be useful to compare different models within the same R package.

3.1 Endemic-epidemic models

As stated in Section 2.1, let us consider that the counts of daily hospitalisations at the $k$-th health department, the $t$-th day, $Y_{kt}$, follows a Poisson distribution with mean as in Eq.1.

If $p_k$ denotes the population of the $k$-th health department, we assume in Eq. 1 known population fractions

$$e_{kt} = e_k = \frac{p_k}{\sum_{k=1}^{24} p_k}, \forall t,$$

and that $w_{qk} = I(q \sim k)$, i.e, $w_{qk} = 1$ if both health districts have a common geographic border (assuming that epidemic only arrives from adjacent health areas) and 0 otherwise. Weights $w_{qk} = \frac{w_{qk}}{\sum_{q} w_{qk}}$ are normalised and restricted to be positive.
Covariates such as number of positive cases, can be added to the model in different ways (Herzog et al. (2011) Meyer, Held, and Höhle). The simplest is to include the covariates $x_{kt}$, in the formulation in the endemic part of the model, for example considering:

- **Model 1:**
  \[
  \mu_{kt} = e_k \nu_{kt} + \lambda_{kt} Y_{k,t-1} + \phi_{kt} \sum_{q \neq k} w_{qk} Y_{q,t-1}, \quad \nu_{kt}, \lambda_k, \phi_k > 0,
  \]

  \[
  \log(\nu_{kt}) = \alpha^\nu + \beta^\nu x_{kt}, \forall k
  \]
  \[
  \log(\lambda_{kt}) = \alpha^\lambda, \forall t, k
  \]
  \[
  \log(\phi_{kt}) = \alpha^\phi, \forall k, t
  \]
  \[
  w_{qk} = I(q \sim k).
  \]

We are going to consider two possibilities regarding to the covariates. Case 1: consider as covariate the smoothed number of new positive cases at a lag 9. Case 2: consider as covariates the smoothed number of positive new cases at a lag 9 and the smoothed number of new positive cases at a lag 5. Many works include a seasonal effect in the model of the parameters $\log(\nu_{kt}), \log(\lambda_k)$ and/or $\log(\phi_k)$ but we would expect that this seasonal effect be included in the covariates (time series of positive cases), so we include it only in the model of $\log(\lambda_k)$ and $\log(\phi_k)$.

All unknown parameters are estimated directly by maximising the corresponding log-likelihoods using numerical optimisation routines (see Paul et al. (2008) Paul, Held, and Toschke). The estimates obtained so as several goodness of fit measurements for this model are shown in Table 3.

Once estimated the parameters of the model, the fitted mean can be compared with the observed counts in order to check the goodness of the fit, but additionally, we can see the contribution to this fitted mean of the endemic and autorregressive components. The average of the proportions of the mean explained by the different components are also shown in Table 3. Note that the proportion explained by the epidemic component is in both cases around 97%, being with difference the component with a greater influence in the value of the total fit. So, there is a high influence of the within - health area autoregressive component with very little contribution of adjacent areas and a rather small endemic incidence.

We have assumed a Poisson distribution to model the observations but as seen in Figure 7, there is a clear overdispersion in the data set. To account this overdispersion, the Poisson distribution may be replaced by two alternatives included in the hhh4 function: 'NegBin1', a Negative Binomial model with a common overdispersion parameter $\psi$ for all districts and 'NegBinM' with different overdispersion parameters ($\psi_i$) for the different health areas, but these distributions does not improve the fit (see Table 4).
|           | Model 1.1 |            | Model 1.2 |            |
|-----------|-----------|------------|-----------|------------|
|           | Estimate  | Std. Error | Estimate  | Std. Error |
| $\alpha^\lambda$ | 0.985     | 0.006      | 0.985     | 0.006      |
| $\alpha^\phi$  | 0.002     | 0.0008     | 0.002     | 0.0008     |
| $\alpha^\nu$   | 4.061     | 0.709      | 4.085     | 0.7117     |
| $\beta_1$      | 1.019     | 0.003      | 1.005     | 0.019      |
| $\beta_2$      | -         | -          | 1.014     | 0.0153     |
| Log-likelihood:| -8559.39  |            | -8558.74  |            |
| AIC:         | 17126.78  |            | 17127.48  |            |
| BIC:         | 17151.64  |            | 17158.56  |            |
| RMSE        | 6.13      |            | 6.18      |            |
| endemic     | 1.39 %    |            | 1.47 %    |            |
| epi.own     | 97.28 %   |            | 97.21%    |            |
| epi.neighbours | 1.33 %    |            | 1.32 %    |            |

Table 3: Estimations, goodness of fit measurements and contribution of the endemic and autoregressive components to the global fit.

|           | Poisson | NegBin1 | NegBinM |
|-----------|---------|---------|---------|
| Log-likelihood:| -8539   |         |         |
| AIC:        | 17177.99| 31554.37| 31309.46|
| BIC:        | 17488.74| 31566.80| 31464.84|

Table 4: Goodness of fit comparison between Poisson and two Negative Binomial distributions.
Table 5: Goodness of fit measurements and contribution of the endemic and autoregressive components to the global fit.

|                      | Model 2.1 | Model 2.2 |
|----------------------|-----------|-----------|
| Log-likelihood:      | -8539     | -8538.49  |
| AIC:                 | 17177.99  | 17178.98  |
| BIC:                 | 17488.74  | 17495.94  |
| RMSE                 | 5.76      | 5.78      |
| endemic              | 1.89 %    | 2.04 %    |
| epi.own              | 94.73 %   | 94.60 %   |
| epi.neighbours       | 3.38 %    | 3.36 %    |

This basic model can be refined. Paul et al. [Paul and Held(2011)] introduced random effects for endemic-epidemic models, which are useful if the districts exhibit heterogeneous incidence levels not explained by observed covariates allowing for district-specific intercepts in the endemic or epidemic components. Due to the heterogeneity shown in the different health departments, Model 2 considers a baseline specific district incidence, $\nu_k$; the population fraction $e_k$ has been included as a multiplicative offset and $\phi_k$ reflects the mean spatial force of influence of the neighbouring health districts.

Both $\nu_k$ and $\phi_k$ have been modelled as fixed effects.

- Model 2:

\[
\begin{align*}
\mu_{kt} & = e_k \nu_{kt} + \lambda_{kt} Y_{k,t-1} + \phi_{kt} \sum_{q \neq k} w_{qk} Y_{q,t-1}, \quad \nu_{kt}, \lambda_{kt}, \phi_{kt} > 0, \\
\log(\nu_{kt}) & = \alpha_k \nu + \beta_k \nu x_{kt}, \quad \forall k \\
\log(\lambda_{kt}) & = \alpha \lambda, \quad \forall t, k \\
\log(\phi_{kt}) & = \alpha_k \phi, \quad \forall t \\
w_{qk} & = I(q \sim k). 
\end{align*}
\]

So, considering a Poisson distribution for the observations, goodness of fit measures of the Model 2 are shown in Table 5 so as the contribution of the three components of the mean in the global fit. Once again we show the results considering as covariates the smoothed number of positive new cases at a lag 9 (Model 2.1) and the smoothed number of new positive cases at a lags 9 and 5 (Model 2.2). Individual estimates for the parameters are not shown now, because the quantity of parameters to estimate.

As can be seen in Tables 3 and 5, there are no much differences between the four models. In all of them, the largest portion of the fitted mean results from
the within-district autoregressive component (among 94 and 97\%) with very little contribution of cases from adjacent districts and a rather small endemic incidence. There are also no great differences in the goodness of fit parameters provided by the Likelihood inference (Log-likelihood, Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC)) or in the values of the RMSE.

Figure 8 shows 5-days predictions obtained with Model 2.2, jointly with the fitted and the true observed values.

3.2 Multivariate Covariance generalised Linear Models

In this section we apply the MCGLM approach to analyse the multivariate count data set that was presented in Section ??, Eq. 5.

As seen in Eq. 5, MCGLM take non-normality into account defining a variance function, and modelling the mean structure by means of a link function and a linear predictor. In this application we have daily observations on 24 health departments (\(K = 24\)) on \(N = 533\) consecutive days.

Despite the previous model, now we are not dealing with an autoregressive
model, but this autoregressive effect can be included in the linear predictor expression together with other covariates.

\[ g(\mu_{kt}) = \beta_0 + \beta_1 x_{kt} + \beta_2 Y_{k,t-1} + \beta_3 \sum_{q \neq k} w_{qk} Y_{q,t-1} + e_{kt}. \]  

(14)

with \( g() \) being in this case the log-link function, and \( e_{kt} \) the offset.

A key point for specifying McGLMs for mixed types of outcomes is the specification of the covariance matrix within outcomes \( \Sigma_k \). Following [Bonat and Jørgensen(2016)] we define the covariance within outcomes by

\[ \Sigma_t = (\Omega(\tau_t))^{1/2} V(\mu_t; p_t)^{1/2}. \]  

(15)

The matrix linear predictor is defined as: \( h(\Omega(\tau_t)) = \tau_0 I_{n \times n} + \tau_1 Z_1 + \tau_2 Z_2 \), where \( n \) denotes the total number of observations in the data set and \( I_{n \times n} \) is the \( n \times n \) identity matrix. \( \tau_0 \) is the intercept of the covariance linear model. If \( n_t \) denotes the number of observations in time in each spatial region, \( Z_1 = I_{24 \times 24} \otimes_G \Gamma_t \), where \( \otimes_G \) denotes the Kronecker product and \( \Gamma_t = (\gamma(i,j))_{i,j \in \{1, \ldots, n_t\}} \) with \( \gamma(i,j) = 1 \) if \( j \in \{i - 1, i + 1\} \) and 0 otherwise. \( \tau_1 \) measures the effect of the ‘time’. Finally, \( Z_2 = W \otimes_G I_{n_t \times n_t} \) with \( W \) being the spatial adjacency matrix between the 24 health departments.

In this case we have obtained the best fits, when using as \( h() \) the exponential covariance link function.

To define \( V(\mu_t; p_t) \), we will adopt the Poisson-Tweedie dispersion function ([Jørgensen and Kokonendji(2016)]) so that

\[ V(\mu_t; p_t) = diag(\mu_t^{p_t}) \]  

(16)

in agreement with the plot of Figure 7.

We employed a step-wise procedure for selecting the components of the linear predictor. As in the previous modelling, we are going to consider as potential covariates in Eq. (14) the number of new positive cases at a lag 9, the number of new positive cases at a lag 5, the number of hospitalizations at lag 1, and an additional categorical covariate ‘Health Department’ to allow different intercepts (\( \beta_0 k \) instead of \( \beta_0 \)) in Eq. (14). The population of each health district will also be used as an offset. The SIC using penalty \( \delta = 2 \) and the Wald test were used in the forward and backward steps, respectively. We defined a stopping criterion for the selection procedure as \( \text{SIC} > 0 \), since in that case the penalty is larger than the score statistics.

The SIC is an important tool to assist with the selection of the linear and matrix linear predictors components, but it is less useful for comparing models fitted using different link, variance or covariance functions. The mcglm package implements the SIC for selection of the linear and matrix linear predictors components, by the functions mc.sic and mc.sic_covariance. In this case, the SIC values indicate that all components, excepting the value of the daily hospitalizations with lag 1 at neighboring regions should be included in the model (SIC < 0).
To compare the goodness of non-nested models, the function \texttt{gof} provides the pseudo Akaike information criterion (pAIC), the pseudo Bayesian information criterion (pBIC) and the pseudo Kullback-Leibler information criterion (pKLIC). We are going to use it to select the bet structure for the matrix linear predictor. Results can be seen in Table 6. We will fit the model considering \( h(\Omega(\tau_t)) = \tau_0 I_{K \times K} + \tau_1 Z_1 + \tau_2 Z_2 \).

Figure 9 shows 5-days predictions obtained with the resulting model, jointly with the fitted and the true observed values. In this case we have obtained a RMSE equal to 5.78. Although the RMSE obtained is similar to the obtained with the other models/packages, we can see in Fig. 9 that in this case there are health departments where the model provides neither a good fit of the observations nor precise predictions, while the fits and predictions of others health departments are very accurate. With the information available, we have not found a model with better fits and predictions for all the health departments.

3.3 Bayesian spatiotemporal models

In this section we apply different models included in CARBayesST package to COVID-19 data given in Section 3.1. For the reasons explained in this section we use the Poisson log-linear model for \( Y_{kt} \), Eq. 10. Overdispersion cannot be controlled with this package. Although that could be regarded as a handicap, as we will see in the results sections the adjusted and forecasted values are quite good.

In order to induce spatial smoothness between the random effects, we use the binary adjacency matrix \( W \) used in previous sections. Element \( w_{ik} = 1 \) if areas \( i \) and \( k \) share a common border and \( w_{ik} = 0 \) otherwise, whereas \( w_{ii} = 0 \) for all \( i \).

Taking into account again the characteristics of our data, the plots shown in Section 3.1 and the fact that our main objective is the prediction of future observations, just two of the spatiotemporal correlation structures included in this package (see Table 2) have been fitted for \( \psi_{kt} \) in Eq. 11: the CARar and CAR adaptive structures. Neither spatially varying linear time trends model and spatiotemporal clustering models are appropriated for our data and CARanova and CARsepspatial strucutres assume a symmetric temporal correlation that not allow to obtain future predictions.
Figure 9: Observed values (in blue) jointly with fitted values (in brown) and predictions (in green).
In both cases the spatiotemporal structure is modelled with a multivariate first order autoregressive process with a spatially correlated precision matrix:

\[ \psi_{kt} = \phi_{kt} \]  
\[ \phi_t | \phi_{t-1} \sim N(\rho_T \phi_{t-1}, \tau^2 Q(W, \rho_S)^{-1}) \]  
\[ \phi_1 \sim N(0, \tau^2 Q(W, \rho_S)^{-1}), \]  
\[ \tau^2 \sim \text{Inverse - Gamma}(a, b), \]  
\[ \rho_S, \rho_T \sim \text{Uniform}(0, 1), \]

\( \phi_t = (\phi_1, \ldots, \phi_K) \) is the vector of random effects for time period \( t \), the precision \( Q(W, \rho_S) \) corresponds to the CAR models proposed in [Leroux et al.(2000)Leroux, Lei, and Breslow] and has the expression:

\[ Q(W, \rho_S) = \rho_S (\text{diag}(W^1) - W) + (1 - \rho_S)I, \]

\( 1 \) is the \( K \times 1 \) vector of ones and \( I \) the \( K \times K \) identity matrix. \( (\rho_S, \rho_T) \) respectively control the levels of spatial and temporal autocorrelation, with values of 0 corresponding to independence while a value of 1 corresponds to strong autocorrelation.

The random effects from CARar have a single level of spatial dependence that is controlled by the parameter \( \rho_S \). That means that all pairs of adjacent areal units will have the same degree of autocorrelation: strongly if \( \rho_S \) is close to one, while no spatial dependence will exist if \( \rho_S \) is close to zero.

The CARadaptative model allows for localized spatial autocorrelation, that is, it allows to be stronger in some parts of the study region. This fact could be adequate for our data because it would be possible that spatial correlation between adjacent health departments to be correlated or conditionally independent, depending on, for example, whether these departments are in the same big city or not or according to the socioeconomic characteristics of its inhabitants.

The CARadaptative model allows this spatial autocorrelation heterogeneity by allowing spatially neighbouring random effects which is achieved by modelling the non-zero elements of the neighbourhood matrix \( W \) as unknown parameters, rather than assuming they are fixed constants.

With respect to covariates to be included on the mean, for the same reasons explained in the previous sections, we tried again two possibilities. Case 1: the smoothed number of new positive cases at a lag 9. Case 2: the smoothed number of positive new cases at a lag 9 plus the smoothed number of new positive cases at a lag 5.

Additionally, as a basic district specific measure of disease incidence, the population fraction \( e_k \) has been also included as an offset.

Inference for all models is based on thinning (by 10) 60000 posterior samples including a burn-in period of a further 1000 samples. Plots of convergence assured that it was reached in all cases.

Table 7 displays the overall fit of each model by presenting the deviance information criterion (DIC) and the effective number of parameters (pd). It
Table 7: Deviance information criterion (DIC) and the effective number of parameters (pD) for each model and scenario.

| Model                  | CARar case 1 | CARar case 2 | CARadaptative case 1 | CARadaptative case 2 |
|------------------------|--------------|--------------|-----------------------|-----------------------|
| DIC                    | 18730.92     | 18741.31     | 18671.93              | 18678.58              |
| pD                     | 793.5918     | 787.9688     | 780.8599              | 770.0481              |
| RMSE                   | 5.610704     | 5.854628     | 5.861029              | 5.402237              |

Table 8

| Parameter | CARar case 1 | CARar case 2 | CARadaptative case 1 | CARadaptative case 2 |
|-----------|--------------|--------------|-----------------------|-----------------------|
| Intercept | 5.7011       | 5.5108       | 5.6233                | 5.6955                |
|           | 5.7307       | 5.2602       | 5.6621                | 5.7247                |
| Posit9    | 0.0022       | 0.0011       | 0.0027                | 0.0022                |
|           | 0.0108       | 0.0018       | 0.0100                | 0.0012                |
| Posit5    | -            | -            | -                     | -                     |
|           | -            | -            | -                     | -                     |
| tau2      | 0.0225       | 0.0197       | 0.0222                | 0.0136                |
|           | 0.0313       | 0.0193       | 0.0313                | 0.0206                |
| rho.S     | 0.9432       | 0.8767       | 0.9666                | 0.9652                |
|           | 0.9591       | 0.9349       | 0.9784                | 0.9225                |
| rho.T     | 0.9895       | 0.9811       | 0.9376                | 0.9898                |
|           | 0.9949       | 0.8420       | 0.9548                | 0.9836                |
| tau2,w    | -            | -            | -                     | -                     |

As can be seen, the estimated parameters of spatial and temporal correlations shows in all cases a strong spatial and temporal correlation. Regarding the covariates both are significative. Finally, with respect to the number of step-changes between two spatially adjacent areas detected in the CARadaptative models only one change step is detected in case 1 meanwhile no changes are detected in case 2. From all these data we can conclude that there are no large differences between the four fitted models. Since our main objective is the prediction, we calculate the mean square error of the prediction up to 5 days for the four cases. They can be found in the last row of Table 7 and again not great differences are observed. In this case the best result is obtained with the CARadaptative model with both covariates: positive cases at lag 9 and 5. Adjusted, observed and predicted values of this models can be seen in Figure 10.

4 Discussion

The aim of this work is to review three different spatiotemporal models for count data, implemented in R packages, and test their performance on an actual case.
study, from three completely different approaches.

For three models tested, significative positive parameters for the covariates are obtained. Those parameters that indicate temporal correlation show high values for all models. Regarding to the spatial correlation, there is difference between the models depending on how the spatial neighbourhood has been included in the model.

Due to these different statistical methodologies, the different packages provide different goodness of fit measures, there being no measure in common between them. So, as the final aim of our case study has been the short term prediction of the evolution of the hospitalisations in the different spatial areas, the mean squared prediction error (RMSE) has been obtained in all cases. We can achieve really good results using each of the packages reviewed. The RMSE of the best model in each package ranges from 5.4 (using CARBayesST package) until 5.78 (using mcglm package). There is no much difference between them. These results are very promising and can promote the use of these models to help in short-term government decision making regarding preventive measures against the hospital collapse. Additionally, they can provide tools to know in advance if it is necessary to expand hospital capacity, in terms of beds and/or workers.

Our objective in this work has not been to say that one package or one type of model is better than others, but to show possibilities that can be used in practice to analyse this type of data. The choice of the type of model to use will depend on the application at hand.
Acknowledgements

This research is supported by Ayudas Fundación BBVA a Equipos de Investigación Científica SARS-CoV-2 y COVID-19 and FONDO SUPERA COVID19 by Banco Santander. We would like to thank to Dr. Nuria Oliver (chief head of Generalitat Valenciana Task force) and Generalitat Valenciana for their support providing the data set.

References

[Kermack and McKendrick(1927)] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 115(772):700–721, 1927.

[Fang et al.(2020)] Yaqing Fang, Yiting Nie, and Marsha Penny. Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions: A data-driven analysis. Journal of medical virology, 92(6):645–659, 2020.

[Kucharski et al.(2020)] Adam J Kucharski, Timothy W Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M Eggo, Fiona Sun, Mark Jit, James D Munday, et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. The lancet infectious diseases, 2020.

[Tang et al.(2020)] Biao Tang, Xia Wang, Qian Li, Nicola Luigi Bragazzi, Sanyi Tang, Yanni Xiao, and Jianhong Wu. Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. Journal of clinical medicine, 9 (2):462, 2020.

[Guirao(2020)] Antonio Guirao. The covid-19 outbreak in spain. a simple dynamics model, some lessons, and a theoretical framework for control response. Infectious Disease Modelling, 5:652–669, 2020.

[López and Rodo(2020)] Leonardo López and Xavier Rodo. A modified seir model to predict the covid-19 outbreak in spain and italy: simulating control scenarios and multi-scale epidemics. Available at SSRN 3576802, 2020.

[Held and Meyer(2019)] Leonhard Held and Sebastian Meyer. Forecasting based on surveillance data. Handbook of Infectious Disease Data Analysis. Chapman & Hall/CRC, 2019.

[Fronterre et al.(2020)] Claudio Fronterre, Jonathan M Read, Barry Rowlingson, Jessica Bridgen,
Simon Alderton, Peter J Diggle, and Chris P Jewell. Covid-19 in england: spatial patterns and regional outbreaks. medRxiv, 2020.

[Dunbar and Held(2020)] M Bekker-Nielsen Dunbar and L Held. Endemic-epidemic framework used in covid-19 modelling. REVSTAT–Statistical Journal, 18(5):565–574, 2020.

[Nelder and Wedderburn(1972)] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. Journal of the Royal Statistical Society: Series A (General), 135(3):370–384, 1972.

[Besag and Higdon(1999)] Julian Besag and David Higdon. Bayesian analysis of agricultural field experiments. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(4):691–746, 1999.

[Paradinas et al.(2017)Paradinas, Conesa, López-Quílez, and Bellido] Iosu Paradinas, David Conesa, Antonio López-Quílez, and José María Bellido. Spatio-temporal model structures with shared components for semi-continuous species distribution modelling. Spatial Statistics, 22:434–450, 2017.

[Gusev(2008)] AA Gusev. Temporal structure of the global sequence of volcanic eruptions: Order clustering and intermittent discharge rate. Physics of the Earth and planetary interiors, 166(3-4):203–218, 2008.

[Law et al.(2014)Law, Quick, and Chan] Jane Law, Matthew Quick, and Ping Chan. Bayesian spatio-temporal modeling for analysing local patterns of crime over time at the small-area level. Journal of quantitative criminology, 30(1):57–78, 2014.

[Choi et al.(2011)Choi, Lawson, Cai, and Hossain] Jungsoon Choi, Andrew B Lawson, Bo Cai, and Md Monir Hossain. Evaluation of bayesian spatiotemporal latent models in small area health data. Environmetrics, 22(8):1008–1022, 2011.

[Gómez-Rubio et al.(2019)] Virgilio Gómez-Rubio, Palmí-Perales, López-Abente, Ramis-Prieto, and Fernández-Navarro. Bayesian joint spatio-temporal analysis of multiple diseases. SORT-Statistics and Operations Research Transactions, pages 51–74, 2019.

[Cressie and Wikle(2015)] Noel Cressie and Christopher K Wikle. Statistics for spatio-temporal data. John Wiley & Sons, 2015.

[Wikle et al.(2019)] Christopher K Wikle, Andrew Zammit-Mangion, and Noel Cressie. Spatio-temporal Statistics with R. CRC Press, 2019.
[Anderson and Ryan(2017)] Craig Anderson and Louise M Ryan. A comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in new south wales, australia. *International journal of environmental research and public health*, 14(2):146, 2017.

[Berger(2013)] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[Liang and Zeger(1986)] Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[R Core Team(2020)] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

[GVA()](https://dadesobertes.gva.es/es/dataset) GVA. Portal de dades obertes. https://dadesobertes.gva.es/es/dataset

[Meyer et al.(2016)Meyer, Held, and Höhle] Sebastian Meyer, Leonhard Held, and Michael Höhle. hhh4: Endemic-epidemic modeling of areal count time series. *Journal of Statistical Software*, 2016.

[Bonat(2018)] Wagner Hugo Bonat. Multiple response variables regression models in r: The mcglm package. *Journal of Statistical Software*, 84(4), 2018.

[Lee et al.(2018)Lee, Rushworth, and Napier] Duncan Lee, Alastair Rushworth, and Gary Napier. Spatio-temporal areal unit modelling in r with conditional autoregressive priors using the carbayesst package. *Journal of Statistical Software*, 84(9), 2018.

[Held et al.(2005)Held, Höhle, and Hofmann] Leonhard Held, Michael Höhle, and Mathias Hofmann. A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical modelling*, 5(3): 187–199, 2005.

[Paul and Held(2011)] Michaela Paul and Leonhard Held. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Statistics in medicine*, 30(10):1118–1136, 2011.

[Meyer et al.(2014)Meyer, Held, et al.] Sebastian Meyer, Leonhard Held, et al. Power-law models for infectious disease spread. *The Annals of Applied Statistics*, 8(3):1612–1639, 2014.

[Paul et al.(2008)Paul, Held, and Toschke] Michaela Paul, Leonhard Held, and André M Toschke. Multivariate modelling of infectious disease surveillance data. *Statistics in medicine*, 27(29):6250–6267, 2008.

[Held et al.(2017)Held, Meyer, and Bracher] Leonhard Held, Sebastian Meyer, and Johannes Bracher. Probabilistic forecasting in infectious disease epidemiology: the 13th armitage lecture. *Statistics in medicine*, 36(22):3443–3460, 2017.

29
[Giuliani et al.(2020)] Giuliani, Dickson, Espa, and Santi. Modelling and predicting the spatio-temporal spread of coronavirus disease 2019 (covid-19) in Italy. *Available at SSRN 3559569*, 2020.

[Held and Paul(2012)] Leonhard Held and Michaela Paul. Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal*, 54 (6):824–843, 2012.

[Bauer and Wakefield(2018)] Cici Bauer and Jon Wakefield. Stratified space–time infectious disease modelling, with an application to hand, foot and mouth disease in china. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(5):1379–1398, 2018.

[Cheng et al.(2016)] Qing Cheng, Xin Lu, Joseph T Wu, Zhong Liu, and Jincai Huang. Analysis of heterogeneous dengue transmission in guangdong in 2014 with multivariate time series model. *Scientific reports*, 6:33755, 2016.

[Bracher(2019a)] J. Bracher. hhh4underreporting. r package. [https://github.com/jbracher/hhh4underreporting](https://github.com/jbracher/hhh4underreporting), 2019a.

[Bracher(2019b)] J. Bracher. hhh4addon: extending the funcionality of surveillance: hhh4. r package. [https://github.com/jbracher/hhh4addon](https://github.com/jbracher/hhh4addon), 2019b.

[Bracher and Held(2017)] Johannes Bracher and Leonhard Held. Periodically stationary multivariate autoregressive models. *arXiv preprint arXiv:1707.04635*, 2017.

[Meyer and Held(2017)] Sebastian Meyer and Leonhard Held. Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, 18(2):338–351, 2017.

[Bracher(2019c)] J. Bracher. hhh4contacts: Age-structured spatio-temporal models for infectious disease counts. r package. [https://cran.r-project.org/web/packages/hhh4contacts](https://cran.r-project.org/web/packages/hhh4contacts), 2019c.

[Gneiting(2011)] Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

[Wei and Held(2014)] Wei Wei and Leonhard Held. Calibration tests for count data. *Test*, 23(4):787–805, 2014.

[Bonat and Jørgensen(2016)] WH Bonat and Bent Jørgensen. Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society, Series C (applied Statistics)*, 65(5):649–675, 2016.

[Martinez-Beneito(2013)] Miguel A Martinez-Beneito. A general modelling framework for multivariate disease mapping. *Biometrika*, 100(3):539–553, 2013.
[Jørgensen and Kokonendji(2016)] Bent Jørgensen and Célestin C Kokonendji. Discrete dispersion models and their tweedie asymptotics. *ASTA Advances in Statistical Analysis*, 100(1):43–78, 2016.

[Jørgensen and Knudsen(2004)] Bent Jørgensen and Sven Jesper Knudsen. Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, 31(1):93–114, 2004.

[Lunn et al.(2009)Lunn, Spiegelhalter, Thomas, and Best] David Lunn, David Spiegelhalter, Andrew Thomas, and Nicky Best. The bugs project: Evolution, critique and future directions. *Statistics in medicine*, 28(25):3049–3067, 2009.

[Rue et al.(2009)Rue, Martino, and Chopin] Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392, 2009.

[Lee(2020)] Duncan Lee. A tutorial on spatio-temporal disease risk modelling in r using markov chain monte carlo simulation and the carbayesst package. *Spatial and Spatio-temporal Epidemiology*, 34:100353, 2020.

[Polo et al.(2020)Polo, Acosta, Soler-Tovar, Villamil, Palencia, Penagos, Martinez, Bobadilla, Martin, Durán, Gina Polo, Carlos Mera Acosta, Diego Soler-Tovar, Julián Felipe Porras Villamil, Natalia Polanco Palencia, Marco Penagos, Juan Meza Martinez, Juan Nicolás Bobadilla, Laura Victoria Martin, Sandra Durán, et al.] Bayesian spatio-temporal modeling of covid-19: Inequalities on case-fatality risk. *medRxiv*, 2020.

[Bernardinelli et al.(1995)Bernardinelli, Clayton, Pascutto, Montomoli, Ghislandi, and Songini] Luisa Bernardinelli, D Clayton, Cristiana Pascutto, Chiara Montomoli, Mauro Ghislandi, and Marco Songini. Bayesian analysis of space–time variation in disease risk. *Statistics in medicine*, 14(21-22):2433–2443, 1995.

[Knorr-Held(2000)] Leonhard Knorr-Held. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, 19(17-18):2555–2567, 2000.

[Napier et al.(2016)Napier, Lee, Robertson, Lawson, and Pollock] Gary Napier, Duncan Lee, Chris Robertson, Andrew Lawson, and Kevin G Pollock. A model to estimate the impact of changes in mmr vaccine uptake on inequalities in measles susceptibility in scotland. *Statistical methods in medical research*, 25(4):1185–1200, 2016.

[Rushworth et al.(2014)Rushworth, Lee, and Mitchell] Alastair Rushworth, Duncan Lee, and Richard Mitchell. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in greater london. *Spatial and spatio-temporal epidemiology*, 10:29–38, 2014.
[Rushworth et al.(2017)] Rushworth, Alastair, Duncan Lee, and Christophe Sarran. An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(1):141–157, 2017.

[Lee and Lawson(2016)] Duncan Lee and Andrew Lawson. Quantifying the spatial inequality and temporal trends in maternal smoking rates in glasgow. *The annals of applied statistics*, 10(3):1427, 2016.

[Herzog et al.(2011)] SA Herzog, M Paul, and L Held. Heterogeneity in vaccination coverage explains the size and occurrence of measles epidemics in german surveillance data. *Epidemiology & Infection*, 139(4):505–515, 2011.

[Leroux et al.(2000)] Brian G Leroux, Xingye Lei, and Norman Breslow. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 179–191. Springer, 2000.