



Research article

Modeling the spread of COVID-19 in spatio-temporal context

S. H. Sathish Indika¹, Norou Diawara^{2,*}, Hueiwang Anna Jeng³, Bridget D. Giles⁴ and Dilini S. K. Gamage²

¹ Department of Mathematics, Virginia Peninsula Community College, Hampton, VA 23666, USA.

² Department of Mathematics & Statistics, Old Dominion University, Norfolk, VA 23529, USA.

³ School of Community & Environmental Health, Old Dominion University, Norfolk, VA 23529, USA.

⁴ Hampton Roads Biomedical Research Consortium Research, Virginia Modeling Analysis and Simulation Center, Old Dominion University, Suffolk, VA 23435, USA.

* **Correspondence:** Email: ndiawara@odu.edu; Tel: +17576833886.

Abstract: This study aims to use data provided by the Virginia Department of Public Health to illustrate the changes in trends of the total cases in COVID-19 since they were first recorded in the state. Each of the 93 counties in the state has its COVID-19 dashboard to help inform decision makers and the public of spatial and temporal counts of total cases. Our analysis shows the differences in the relative spread between the counties and compares the evolution in time using Bayesian conditional autoregressive framework. The models are built under the Markov Chain Monte Carlo method and Moran spatial correlations. In addition, Moran's time series modeling techniques were applied to understand the incidence rates. The findings discussed may serve as a template for other studies of similar nature.

Keywords: COVID-19; conditional autoregressive model; Bayesian analysis; Moran statistics; dynamics

1. Introduction

The novel coronavirus (COVID-19) disease has developed rapidly in the United States of America and around the world. It has led many nations and localities to reassess their preparedness in case of disease outbreaks. Many phases of the disease have been noticed where the count of COVID-19 cases and related deaths are characterized as too high to continue daily normal activities. There are locations

where the severe outbreak has led to drastic decisions and complete shutdowns except for essential services. Some cities have implemented partial shutdowns. The phases of the shutdowns are related to the geographical locations of outbreaks. The latter cannot be ignored in any modeling scheme as they may shed light on the count effects and dynamics of the disease over time and its transmission. Such models will help in the monitoring and prediction of the disease.

In this paper, we make use of a statistical Bayesian approach [1–3]. In such Bayesian approach, where the inherent spatio-temporal structure is modeled, the random effects are autocorrelated. We will use the CARBayesST statistical package in R for the inferential statistical analysis. The CARBayesST statistical package allows spatio-temporal areal unit modeling with conditional autoregressive priors. This is the first dedicated software package in R for this purpose. This software package is very flexible in the sense that it allows the researcher to fit a wide range of models, focusing on different aspects of space-time modeling. The CARBayesST statistical package can be used to estimate the overall space and time trends. It can be used to identify clusters.

The data of interest include the total COVID-19 case counts in the state of Virginia, as reported by the Johns Hopkins University (JHU) website (<https://coronavirus.jhu.edu/map.html>). The type of real and temporal unit data relates to 93 counties in the state of Virginia. Specifically, the random response variable of interest is the daily COVID-19 case counts in these 93 counties. The 93 counties are considered to have non-overlapping areal units. Realizations of the random response and the case counts for the 93 counties have been considered for 490 consecutive days, beginning from January 22, 2020. The total cases of COVID-19 in the state of Virginia, as reported from the JHU, are a type of summary measure. Many authors have used summary measures in analyzing areal and temporal unit data. For example, the total yield in sectors in an agricultural field trial was considered by Besag and Higdon [4]. Choi and Lawson [5] considered the total number of individuals suffering from chronic obstructive pulmonary disease in counties of the US state of Georgia.

The manuscript describes a novel setting for the application of statistical methods in COVID-19 spread and management, with the aim to identify clusters [6–8]. This will allow public health officials and decision makers to take timely and effective preventive measures. Our motivation for modeling this spatio-temporal data is to estimate the effect of risk factor(s) on the random response variable. Land area (size of the county) and population density were among the list of potential significant risk factors. Another motivation is to identify clusters and areal units that exhibit elevated values for the response variable, i.e., areas where an elevated risk of the disease is present, such as elderly nursing facilities. These values give an idea of clustering or dispersion, i.e., spatial autocorrelation [9,10]. Linking the changes in the total case counts of the disease to local counties with executive orders will play a key role in dealing with the spread. The dynamic progression with the availability of high-quality data stands as a novel indicator for public health and surveillance. To measure the global spatial autocorrelation in the random response variable of interest, we calculate Moran's I statistic [11] and graph the Moran's I statistic over time. A variety of disease mapping methods from simulated to real data, in the context of goodness-of-fit, have been considered [12–15]. They have applied the Moran's I statistic to the Pearson residuals from the model fit. To explain differences based on locations, a simple global Moran's value cannot explain the whole growth. As such, [16] proposed to partition the global Moran value into a sum of local indices of spatial association (LISA); the individual Moran values show local trends that are not shared globally.

A zero-inflated Poisson autoregressive model and zero-inflated negative binomial autoregressive model of the COVID-19 data in Ghana have been proposed by Tawiah et al. [17]. Their paper investigated the trend of daily count of COVID-19 deaths in Ghana. Chang et al. [18] have introduced a metapopulation susceptible–exposed–infectious–removed (SEIR) model. This model integrates

dynamic mobility networks to simulate the spread of COVID-19 in ten of the largest metropolitan areas in the USA. The mobility networks considered are derived from mobile phone data. Comess et al. [19] have considered testing problems in the context of pooled testing during the COVID-19 pandemic. They have discussed the need of integrating field sampling information with statistical modeling. This integration is to optimize pooled testing procedures. The Institute for Health Metrics and Evaluation (IHME) COVID-19 Forecasting Team [20] has considered COVID-19 case counts and mortality data from February 1, 2020, to September 21, 2020. The SEIR model was used to predict possible trajectories of severe COVID-19 infections and the effects of non-pharmaceutical interventions from September 22, 2020, through February 28, 2021, in the US. Moreover, Moreau [21] studied the COVID-19 outbreak in Brazil in the context of the Weibull distribution model. Their SIR model did not fit well in the COVID-19 context data. The susceptible, infected and recovery compartments that are needed in this disease dynamic are difficult to obtain. Likewise, an extension of the SIR model includes the *exposed* compartment (SEIR). As mentioned in Tolles and Luong [22], alternative models become more reliable in drawing informed public policy. Their model was used to forecast COVID-19 cases, in scenarios based on the daily number of new deaths as a function of time. Finally, Woolf et al. [23] used a Poisson regression model for the mortality data from March to August 2020 and predicted the number of deaths attributed to COVID-19 in USA in 2020.

Bayesian conditional autoregressive (CAR) models and approaches offer innovative and computational routines that can be adjusted and applied to longitudinal data. One more feature of CAR models is that they allow an update of the prior information in fitting present data and building predictions about the future. Bayesian models evaluate procedures over repeated samplings. In this paper, we use the Bayesian approach, where the inherent spatio-temporal structure is modeled using a set of random effects. This approach is widely used in the modeling of disease data [24]. Conditional autoregressive priors and spatio-temporal extensions have been typically assigned to these models to capture spatial and temporal autocorrelations [4,25]. The model proposed by Knorr-Held [26] provided an ANOVA style decomposition of the data into overall spatial trends, temporal trends and spacetime interaction. While frequentist models have been popular in the literature, the main challenges faced are related to the model accuracy estimates and the design matrix experiment. The idea is presented under the likelihood principle as described in Boedeker [27]. Giudici et al. [28] have considered a class of birth process called self-exciting point in their research to measure the spread in space and time (spatio-temporal network model). They also considered covariates, such as spatio-temporal mobility, and sociodemographic factors (vaccination, policy). The parameters of the model were estimated using the maximum likelihood method via an expectation maximization (EM) algorithm. Giudici et al. [29] proposed a Bayesian time-dependent Poisson autoregressive model. They have considered varying non-pharmaceutical interventions (NPIs) implemented by most countries as their explanatory variables. Paez et al. [30] investigated the influence of environmental factors, such as temperature, humidity and sunshine, on the COVID-19 pandemic in Spain. They have used a seemingly unrelated regressions (SUR) approach and reported results from a spatio-temporal model. The manuscript by De Angelis et al. [31] describes the spatio-temporal spread of COVID-19 from a highway transportation network that dictates mobility. Park et al. [32] proposed a spatio-temporal model using a Dirichlet process mixture model. Aggregated data may not always accurately capture the fine-scale spatial heterogeneity of disease transmission, and this can lead to biased decisions by decision-makers and public health officials. Inaccurate models and unreliable preliminary data can also have significant impacts on decision-making, as seen in the examples provided by Bizzarri et al. [33] and by Guenther et al. [34]. In Bizzarri et al. [33], the unreliable preliminary data, as well as inaccurate models, significantly affected the political decisions of the

Italian administration. Another example in Guenther et al. [34] documents a superspreader event in Germany in a meat processing plant. Modeling such an event requires accounting for the precise plant location information, and aggregated data may miss such crucial local information. To address these challenges, it is important to collect and analyze data at specified spatial and temporal scales. Our paper proposes another model to complement the literature. In the Bayesian context, the main challenges are related to prior selection and possible non-robustness of the estimates. Data from the state of Virginia are given consideration, where the area is partitioned into counties. Spatial dependence among the counties is captured by looking at the neighborhood matrix, denoted by W , of closeness distance. Moran measures of spatial correlations are computed. We apply the Poisson count process to provide a method of relating count time series and spatial covariates by specifying different kernels in the W matrix of the measure of proximity.

The manuscript is organized as follows. In Section 2, we provide a description of the data set. We propose preliminary analyses in Section 3. The CARBayes statistical modeling is described in Section 4. In Section 5, Moran's statistic of spatial dependence and spatial analysis are presented. The findings are discussed in Section 6. The conclusion is provided in Section 7.

2. The data set

Data on total COVID-19 case counts from January 22, 2020, to May 25, 2021, were collected using the R package called "covid19.analytics". Total case counts were accounted for each day, and there was a total of 490 days per county. The aggregate data are also available under the US Centers for Disease Control and Prevention (US CDC). Given the association between population size and COVID-19, data on time in days and area sizes were collected as covariates. Daily counts based on a particular day (starting Monday, January 22) over the time were collected, for a total of 490 daily counts for each county.

3. Preliminary analyses

This type of spatial data has an inherent spatio-temporal autocorrelation. Modeling the spatial autocorrelation and the temporal autocorrelation is important to accurately model and classify the growth of total case counts. So, by looking at the overall trend, it is difficult to classify the growth of total case counts without advanced statistical models. From a single time point of cases, no conclusion can be drawn about the trends or the growth of total cases over time. The histogram of total cases in the county of Henrico and its neighbor Chesterfield County reveal trends that are unforeseen. Nearby counties are not necessarily alike. The histogram of the county of Henrico can reach values that are almost twice those of the neighboring county of Chesterfield. See Figure 1.

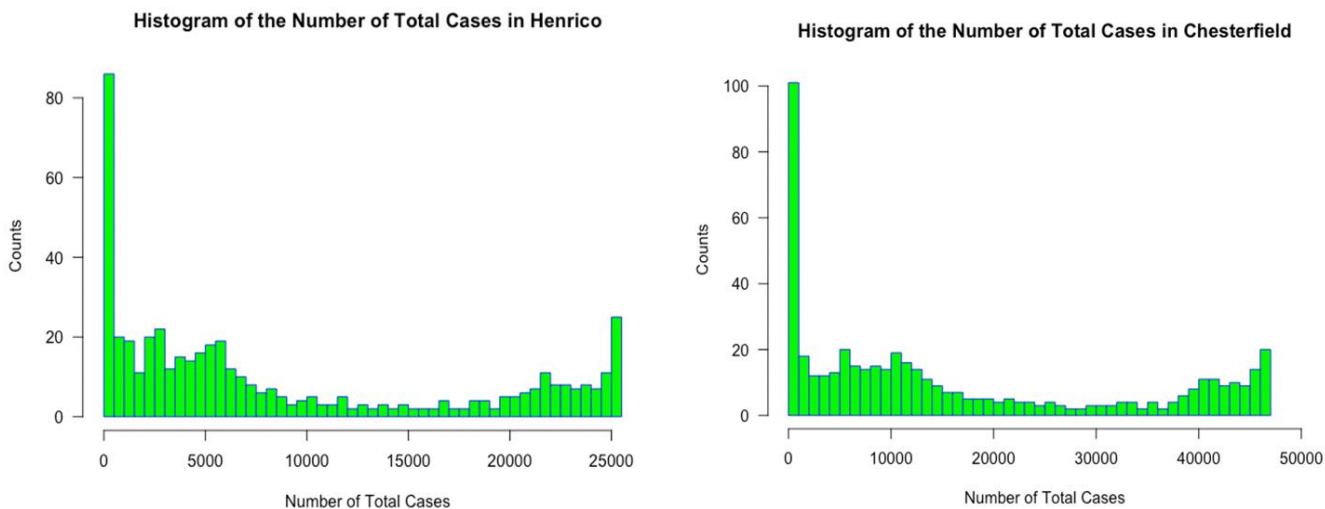


Figure 1. Histograms of COVID-19 case counts in Henrico and Chesterfield Counties.

For example, the county of Henrico has higher counts of total cases than Richmond, although Richmond is the capital city in terms of population size and resources. The histogram of the total number of cases shown in Figure 2 is now broken up into three other figures (Figure 3) based on the population density. The population density was obtained by dividing the 2019 population (in thousands) by land area (in square miles). Three categories low (5.3–47.1), medium (47.7–96.5) and high (102.8–9118.4) were generated. The tails of those histograms are heavier for medium- and high-density areas, where we observed the highest number of total cases. We also plotted the data for total cases per county.

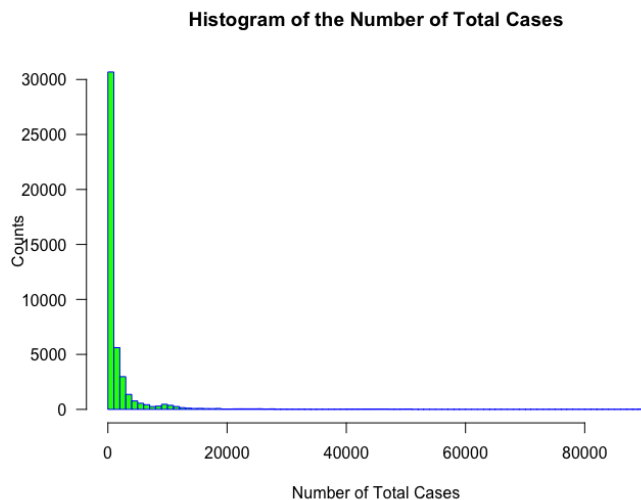


Figure 2. Histogram of the COVID-19 case counts aggregated across 93 counties.

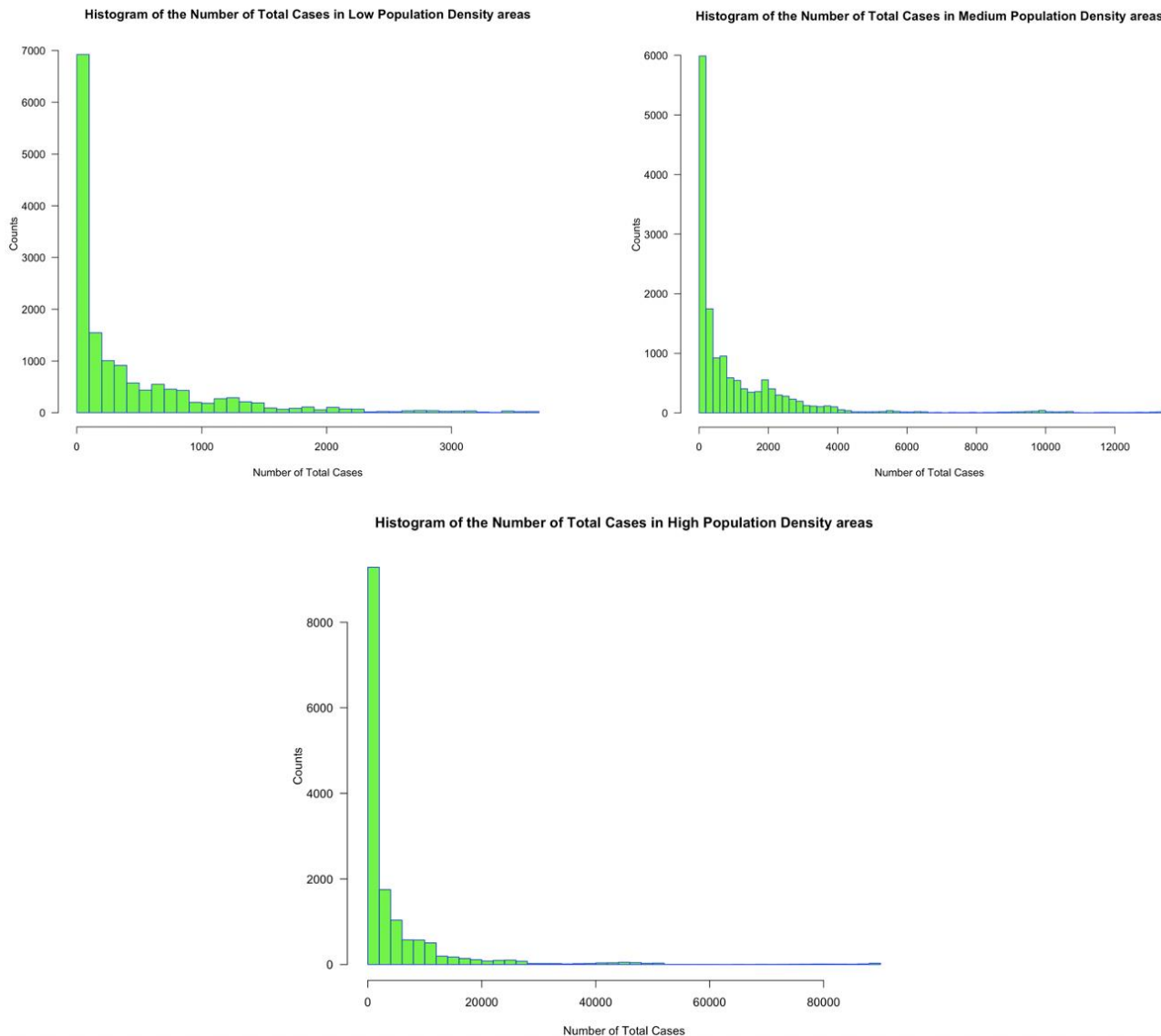


Figure 3. Histograms of COVID-19 case counts by population density.

The progression of COVID-19 cases is not accurately tied to the population size in the counties of Virginia. Figures 3 and 4 of the number of cases in high, medium and low-density areas have similar shapes. Such results show the complex temporal variation where the feature inputs are still not well understood. Day and land area are used in our analysis. The range of the total number of cases in Figure 4 changes from 1000 total cases (Amelia) to 2000 total cases (Alleghany) to 3000 (Accomack) to 10000 (Albemarle) and 80000 (Arlington). Understanding the spatial dependence of the dataset becomes challenging. So, a simple model is not enough to understand the trends in total case counts in the state. An alternate approach is to look at Moran's spatio-temporal values along with Bayesian analysis methods. The first gives a measure of spatial correlation, an indication of hot spots and a sense of dependence. The second is obtained after running the CARBayes model.

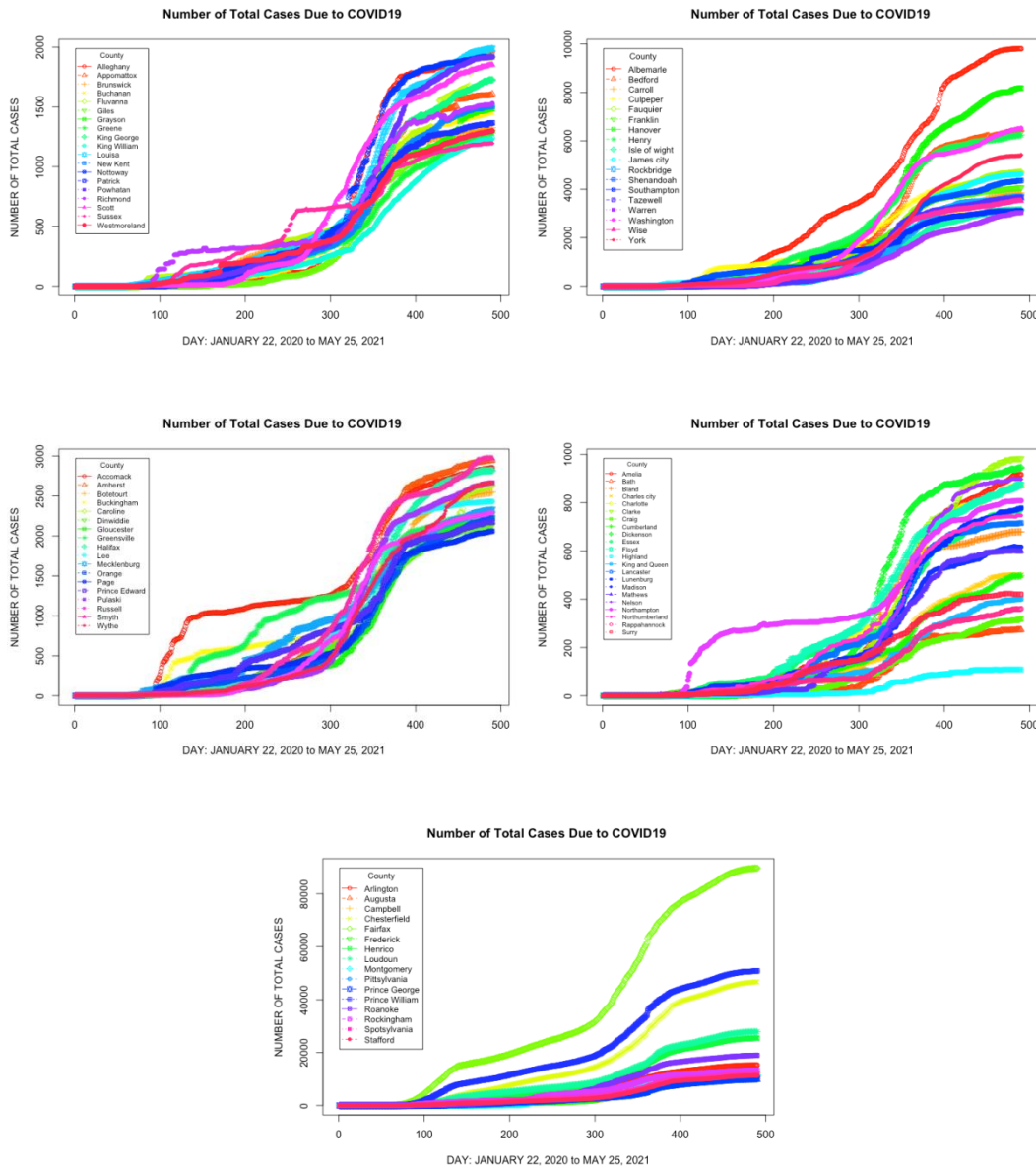


Figure 4. The ranges of the total number of cases.

4. CARBayes statistical modeling

The 93 counties in the state of Virginia, as non-overlapping areal units, are considered the study region. The data are available for each areal unit for 490 consecutive time periods. Here, we are considering 93 counties ($K = 93$) and 490 consecutive days ($N = 490$) from January 22, 2020, to May 25, 2021.

Let Y_{kt} be the number of people who are infected with COVID-19 in the k^{th} county at day t .

$$k = 1, 2, \dots, 93 \text{ and } t = 1, 2, \dots, 490.$$

We consider p (in our case $p = 2$) known covariates (day and area).

Let $\mathbf{x}_{kt} = (x_{kt1}, x_{kt2}, \dots, x_{ktp})$, $k = 1, 2, \dots, K$ and $t = 1, 2, \dots, N$. \mathbf{x}_{kt} is a vector of p known covariates for k^{th} county and day t .

The most generalized form of the model that we consider is [35]

$$Y_{kt} | \mu_{kt} \sim f(y_{kt} | \mu_{kt}, v^2) \quad \text{for } k = 1, \dots, K, t = 1, \dots, N,$$

$$g(\mu_{kt}) = \mathbf{x}_{kt}^T \boldsymbol{\beta} + O_{kt} + \phi_{kt}, \quad \boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \quad (1)$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ is a vector of covariate regression parameters. $\boldsymbol{\beta}$ is assumed to be a multivariate normal prior with mean $\boldsymbol{\mu}_\beta$ and covariance matrix $\boldsymbol{\Sigma}_\beta$. The prior parameters can be chosen by the user. The ψ_{kt} term in the linear model is a latent component term to encompass areal unit k and time t . The ψ_{kt} term will account for the spatio-temporally autocorrelated random effects. v^2 represents the variance of the data.

The complete descriptions of these spatio-temporally autocorrelated random effects and offset terms are as follows:

$$\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_t, \dots, \boldsymbol{\phi}_N), \text{ where } \boldsymbol{\phi}_t = (\phi_{1t}, \dots, \phi_{kt}, \dots, \phi_{Kt}), \text{ and}$$

$$\mathbf{O} = (\mathbf{O}_1, \dots, \mathbf{O}_t, \dots, \mathbf{O}_N)_{K \times N}, \text{ where } \mathbf{O}_t = (o_{1t}, \dots, o_{kt}, \dots, o_{Kt}).$$

We use

$$Y_{kt} \sim \text{Poisson}(\mu_{kt}) \text{ and } g(\mu_{kt}) = \ln(\mu_{kt}).$$

The spatio-temporal structure that we consider is the model ST.CARar [36]. This spatio-temporal structure consists of a multivariate first order autoregressive process with a spatially correlated precision matrix. The precision matrix is the inverse of the covariance matrix [37,38]. The precise definition of the spatio-temporal structure is as follows:

$$\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1} \sim N(\rho_T \boldsymbol{\phi}_{t-1}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho_s)^{-1}), t = 2, \dots, N, \quad (2)$$

$$\boldsymbol{\phi}_1 \sim N(\mathbf{0}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho_s)^{-1}),$$

$$\tau^2 \sim \text{Inverse - Gamma}(a, b),$$

$$\rho_s, \rho_T \sim \text{Uniform}(0, 1),$$

where $\boldsymbol{\phi}_t = (\phi_{1t}, \dots, \phi_{kt}, \dots, \phi_{Kt})$ is a vector of random effects for time t ($t = 1, \dots, N$). $\boldsymbol{\phi}_t$ will evolve over time via a multivariate autoregressive process. Here, we are using the precision matrix ($\mathbf{Q}(\mathbf{W}, \rho_s)$) proposed by [39]. The parameter τ^2 controls the overall variability of $\boldsymbol{\phi}_1$. ρ_s and ρ_T represent the spatial and temporal correlations, respectively. The algebraic definition of the precision matrix is as follows:

$$\mathbf{Q}(\mathbf{W}, \rho_s) = \rho_s [\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}] + (1 - \rho_s)\mathbf{K},$$

where the vector $\mathbf{1}$ is the $K \times 1$ vector of ones, and \mathbf{K} is the $K \times K$ identity matrix. \mathbf{W} is called the symmetric non-negative $K \times K$ neighborhood, weight or adjacency matrix.

$$\mathbf{W} = (w_{kj}), \text{ where } w_{kj} = \begin{cases} 1 & \text{if } (S_k, S_j) \text{ share a common border} \\ 0 & \text{otherwise} \end{cases}$$

w_{kj} can be described as the spatial closeness between the counties S_k and S_j , for $k = 1, \dots, K$ and $j = 1, \dots, K$.

The specific model from Eq (1) that we consider in this paper is

$$\mathbf{x}_{kt}^T \boldsymbol{\beta} = \beta_{0t} + \beta_{1t} \text{day}_{kt} + \beta_{2t} \text{area}_{kt}, \quad (3)$$

for $k = 1, \dots, K, t = 1, \dots, N$, where $Y_{kt} \sim \text{Poisson}(\mu_{kt})$, where μ_{kt} is the mean for the k^{th} county at time t .

5. Moran's statistic

The Moran values can indicate spatial similarities and dissimilarities in the progression of the disease. We use the Moran values as a comparative tool for the CARBayes model. Although the assumptions under Moran values may not be met, they provide a basic validation of the results in the Bayesian context [40]. We compare the performance of the CARBayes model and the change in Moran values over time using the spatial weight matrix. This matrix, denoted as \mathbf{W} , links the counties' geographical distance. It is the same as the adjacency matrix described in Section 4.

Moran's statistic (referred to as Moran's I) [11] is defined as

$$I = \frac{K}{W} \frac{\sum_{i=1}^K \sum_{j=1}^K w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^K (y_i - \bar{y})^2}, \quad (4)$$

where

\mathbf{W} is the symmetric non-negative $K \times K$ neighborhood, weight or adjacency matrix.

y_i is the total number of people infected with COVID-19 in the i^{th} county, $i = 1 \dots K$,

and $\bar{y} = \frac{\sum_{i=1}^K y_i}{K}$.

Positive spatial autocorrelation occurs when similar counts cluster around the same geographical area. Negative spatial autocorrelation occurs when dissimilar counts arise next to each other in the same geographical area. This helps to understand relationships of the observed counts from one county to another and validate if the observations are independent. We are using Moran's statistic to understand the spatial trends of the COVID-19 cases in Virginia.

6. Statistical findings

The total case count keeps increasing but at different rates over time, as shown in Figure 5. Covid-19 case counts in different locations cannot be assumed to be independent. If there is some dependence, the Moran value, Eq (4), can lead us to the cluster structure. We develop an accurate classifier of the coefficient and the covariance matrix by analyzing the system as a whole. Other predictors, such as the number of intensive care unit beds available (hospitalization counts) or the population size and density in that area, could have been used.

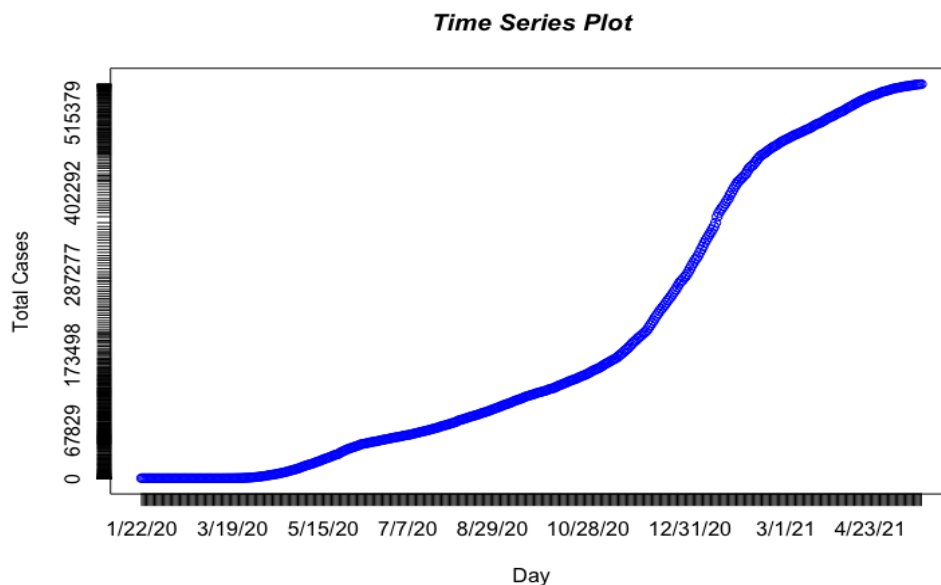


Figure 5. The trend of COVID-19 case counts from January 22, 2020, to March 23, 2021.

The analysis is challenging because the measurements are not independent, and clustering is present. We use the linear mixed model to incorporate the temporal trends and patterns under non-normal data with varying intercepts and slope models. Parameters are estimated under CARBayes using conditional auto-regressive priors under Markov Chain Monte Carlo (MCMC) simulations. Slightly similar variations of the models were also considered. An offset was added to allow the Poisson response in a predicted simple linear (or equivalently log linear) regression. This model is known as the quasi-Poisson log linear model, which provides the statistics associated with the day and land area estimates as 0.0024 and -0.00004 , respectively (Table 1). They are both significant at the 0.05 level. The signs indicate the spread of total case counts and the pervasive nature of the virus transmission and of people in that area at that time. The case count increases per day but decreases as the location area gets larger. The model is described in the formula of Eq (1).

Table 1. Estimates of the parameters under Poisson log linear model.

	Estimate	Std. Error	T value	Pr(> t)
Intercept	-9.239×10^{-1}	1.525×10^{-2}	-60.568	$< 2 \times 10^{-16}$
Day	2.396×10^{-3}	3.520×10^{-5}	68.062	$< 2 \times 10^{-16}$
Land Area	-3.943×10^{-5}	1.655×10^{-5}	-2.383	0.0172

The trace plots (history plots) and the density plots' MCMC convergence of the parameters associated with intercept, day and land area are represented for the model in Figure 6, respectively. Figure 6 shows the monitoring plots and kernel density estimates for the parameters β_0 , β_1 , β_2 and produced by the CARBayes model Eqs (2) and (3). We ran three independent Gibbs sample chains, and we picked the model chain with the lowest deviance information criterion (DIC). See [41]. The corresponding estimated effective number of parameters (p.d) and its corresponding estimated log of marginal predictive likelihood (LMPL) are also reported. The plots are based on 440,000 iterations, with 40,000 first points removed as burning points and a thinning rate of 20 indicative of 40,000 posterior estimates. Although the convergence is difficult to achieve, as the density plots of the parameters in Figure 6 indicate, it highlights the evolution of the disease as a function of time. No one single distribution function can estimate the parameters of the model. Rather, a mixture of distributions will be a good measure of performance of the variables. The value of the DIC was about 370,566. The trace plots of the densities (Figure 7) show that the autocorrelation between the successive collected data cannot be ignored. The values of spatial and temporal correlations ($\rho.S$ and $\rho.T$) are about 0.7933 and 0.9950.

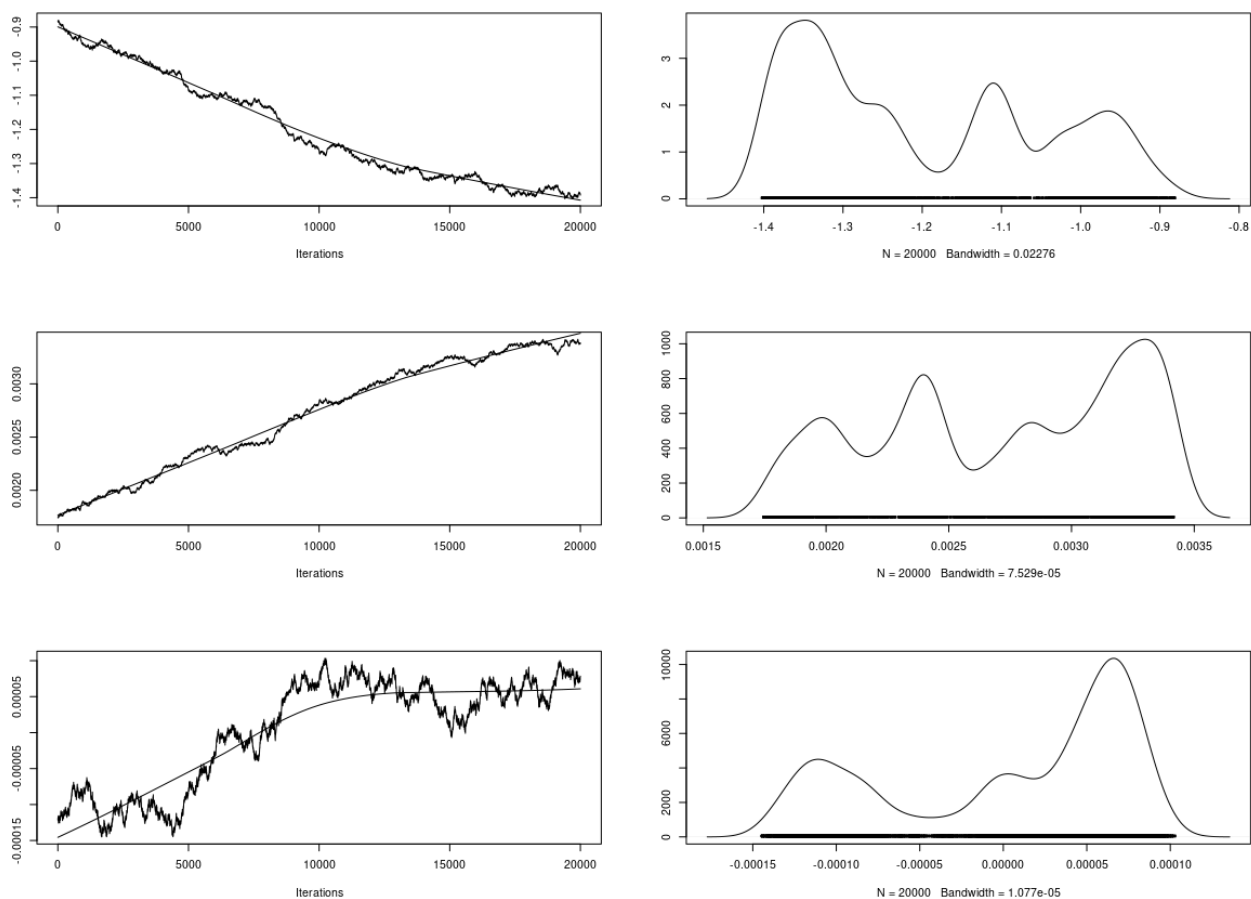


Figure 6. Trace plots and density plots of parameters related to the parameters in Table 1: intercept (a), day (b) and area (c).

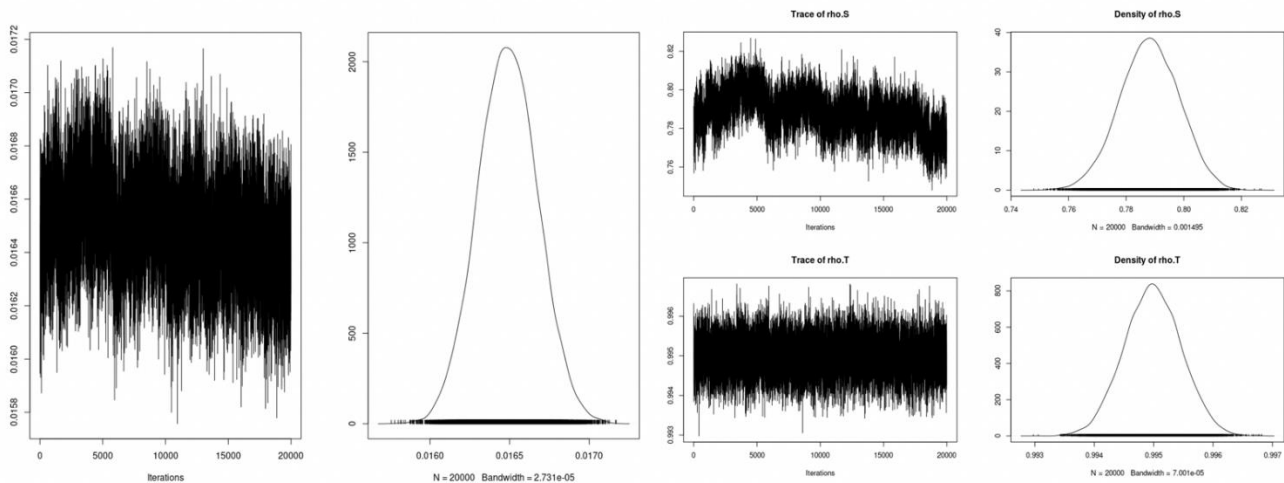


Figure 7. Values (trace plot and density plot) of overall variance, spatial and temporal correlations of Equation 2 built from Table 2 (τ^2 , ρ_S and ρ_T).

The proposed model was also evaluated based on the Geweke diagnostic values. The values of the posterior quantities’ median scores, the comparisons between the first 2.5% and last 97.5% of the chain quantiles, along with the effective length of the chain are displayed in Table 2. The results in Table 2 reveal that the estimates of the parameters and the autocorrelation are all significant. The spatial dependencies are also all significant. The posterior 95% credible intervals do not contain zero. The exponential of the median for day and land area is 1.0027 and 0.9999, respectively. These values suggest a relative risk increase of cases by 0.27% additional total cases per day, whereas the corresponding relative risk for a one square mile increase in area size is decreased by 0.01%. Thus, we found that the spread of the disease is increasing with time, and decreasing the area size is also related to the spread of the disease.

Table 2. Posterior quantities for selected parameters and DIC.

	Median	2.5%	97.5%	n.effective	Geweke.diag
(Intercept)	-1.0376	-1.3071	-0.9020	1.8	5.4
Day	0.0023	0.0017	0.0032	1.3	-6.5
Land area	-0.0001	-0.0002	0.0000	4.8	2.2
τ^2	0.0165	0.0162	0.0169	341.1	-1.9
ρ_S	0.7933	0.7695	0.8146	35.4	-0.9
ρ_T	0.9950	0.9941	0.9959	6335.0	-0.7

*Note: DIC = 370566.4, p.d = 23376.46, LMPL = -186084.1.

The total cases may vary substantially based on geographical location and time. Based on the structure results using Bar and Wells [42], most sampling locations can be pooled into one cluster (Figure 8). This shows the significance of the spatial correlation and the growth of the total number of cases in the region as a whole. Such results can help to guide healthcare regulations. The results in Figure 8 show that the total cases cannot be discretized and confirm that there is no spatial cluster, and in fact the spread is for the whole region. In Figure 9, the spread is visualized, where the red shows the highest counts of cases at six selected times between January 22, 2020, and May 25, 2021. The scales are changed over the six time points to highlight the counties with high counts of total cases of

COVID-19. The pattern can be associated to the Governor's executive orders (EO), as shown in the graph of the Moran values (Figure 10).

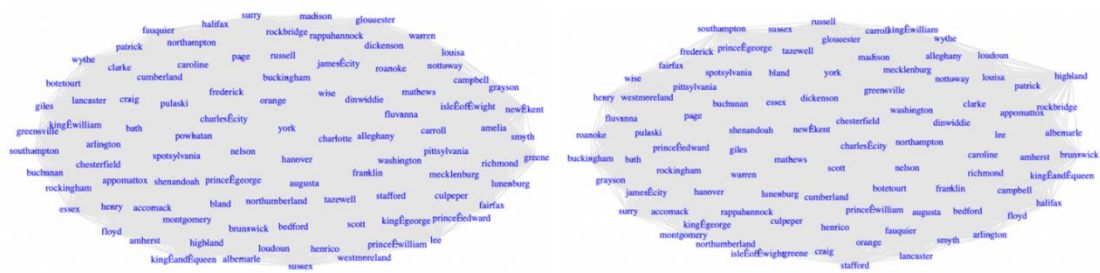


Figure 8. COVID-19 total case count networks – Dec 15, 2020, to Dec 21, 2020 (left) and Jan 7, 2021, to Jan 14, 2021 (right).

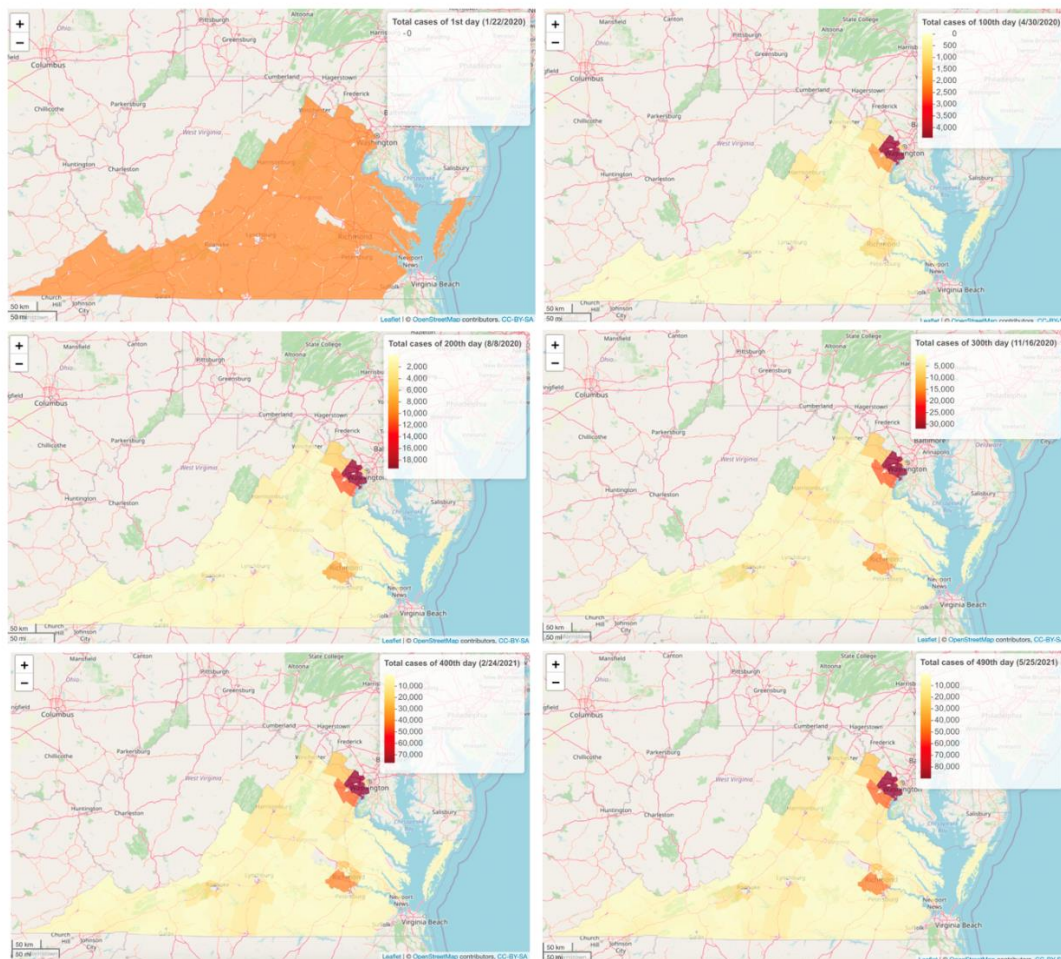


Figure 9. Maps depicting the pattern of COVID-19 over time in 93 counties at the six selected times: January 22, 2020 (1st day), April 30, 2020 (100th day), August 8, 2020 (200th day), November 16, 2020 (300th day), February 24, 2021 (400th day) and May 25, 2021 (490th day).

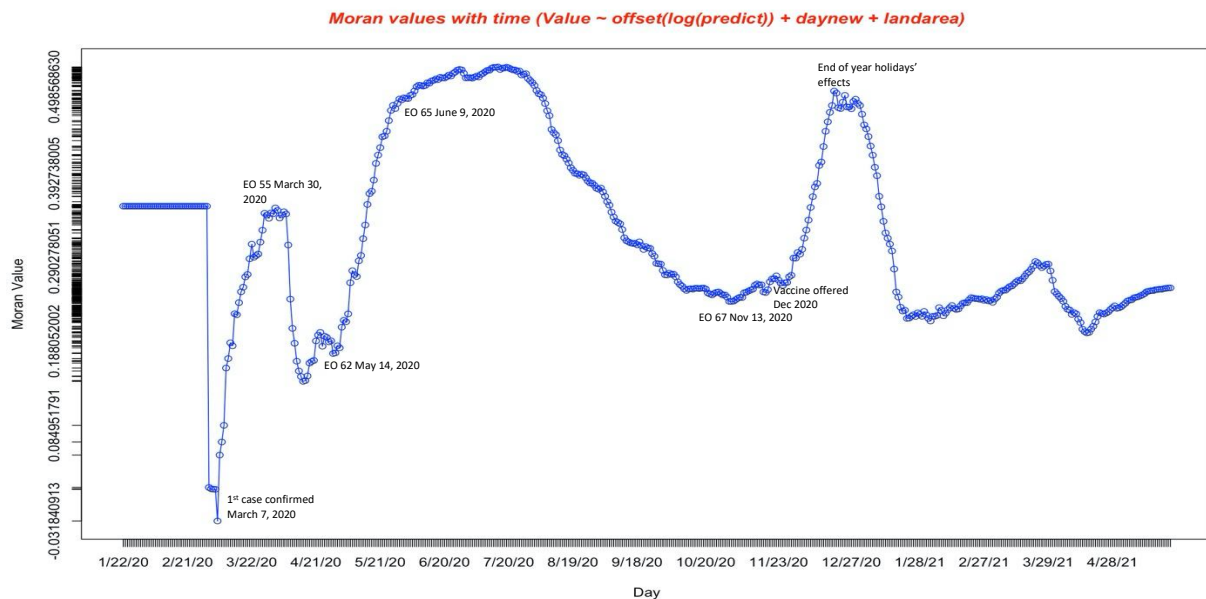


Figure 10. Moran values over time and state decisions. EO 55, March 30, 2020: Temporary stay at home order due to novel coronavirus. EO 62, May 14, 2020: Amended order of public health emergency four; jurisdictions temporarily delayed from entering phase one. EO-65: June 9, 2020: Phase two easing of certain temporary restrictions due to novel coronavirus; EO 67, Nov 13, 2020: Phase three easing of certain temporary restrictions due to COVID-19.

The values of Moran’s statistic that are significantly below, equal to or above $-\frac{1}{(K-1)} = -\frac{1}{92} = -0.01087$ imply that there is a negative, zero or positive spatial autocorrelation, respectively [43].

Almost all our computed Moran values are greater than -0.01087 (Figure 10). This shows that there is a positive spatial correlation. There is an inherent relationship between total cases and the predictive variables (day and area). Geographical distance plays a main factor in the COVID-19 cases in short time periods, as mentioned in [44]. We can also see that the Moran values closely connect with the EOs and “end of year effects.”

7. Conclusions

This paper models the progression of the coronavirus (COVID-19) pandemic in 93 counties in the state of Virginia. The daily count data related to the total cases of COVID-19 in these 93 counties are analyzed. The statistical framework of spatio-temporal context with conditional autoregressive (CAR) modeling has been used from the R package CARBayesST for the descriptive and inferential statistical analysis. The Moran statistic values have also been computed to compare spatial properties of the total cases of COVID-19. The results confirm the relationship of the total cases of COVID-19 in space and time. The total cases at a specific time point are impacted by and linked to the Governor’s executive orders. We show the connection of the Moran values in state decisions via the plot of the Moran values. With our proposed outlook, we have shown that the Bayesian approach is a solid approach for the modeling of the total cases of COVID-19 in space and time. This is quite novel in the literature of COVID-19 data.

The research literature and applications for spatio-temporal modeling of total cases of

COVID-19 are expanding. As new statistical techniques for spatio-temporal modeling develop, the properties of such geostatistical models will be tested and validated. Future work will include conditional and multivariate space-time models that have efficient and tractable computational capabilities.

We have considered two covariates (time and location) in our research. Adding more covariates, such as the number of hospitalizations each day for each county or the intensive or critical care units availability related to COVID-19 in the respective counties, may be useful to understand the disease progression. Also, changing the covariance structure may be considered in future studies.

Acknowledgments

The authors (N.D., H.A.J. and B.D.G.) are grateful for the resources provided by the ODU Advanced Computing Core and grant support from the ODU Research Foundation.

Conflicts of interest

The authors declare there is no conflict of interest.

References

1. A. Gelman, J. Carlin, H. Stern, D. B. Rubin, *Bayesian data analysis*, 3rd edition, Chapman & Hall, London, 2013.
2. M. V. Ibañez, M. Martínez-García, A. Simo, A review of spatiotemporal models for count data in R packages. A case study of COVID-19 Data, *Mathematics*, **9** (2021), 1–23. <https://doi.org/10.3390/math9131538>
3. J. Tomal, S. Rahmati, S. Boroushaki, L. Jin, E. Ahmed, The impact of COVID–19 on students' Marks: A Bayesian hierarchical modeling approach, *METRON*, **79** (2021), 57–91. <https://doi.org/10.1007/s40300-021-00200-1>
4. J. Besag, D. Higdon, Bayesian analysis of agricultural field experiments, *J. R. Statist. Soc. B*, **61** (1999), 691–746. <https://doi.org/10.1111/1467-9868.00201>
5. J. Choi, A. B. Lawson, Evaluation of Bayesian spatial-temporal latent models in small area health data, *Environmetrics*, **22** (2011), 1008–1022. <https://doi.org/10.1002/env.1127>
6. M. M. Hossain, A. B. Lawson, B. Cai, J. Choi, J. Liu, R. S. Kirby, Space-time areal mixture modelling: Relabeling algorithm and model selection issues, *Environmetrics*, **25** (2014), 84–96. <https://doi.org/10.1002/env.2265>
7. C. Anderson, D. Lee, N. Dean, Identifying clusters in Bayesian disease mapping, *Biostatistics*, **15** (2014), 457–469. <https://doi.org/10.1093/biostatistics/kxu005>
8. C. Anderson, D. Lee, N. Dean, Bayesian cluster detection via adjacency modeling, *Spatial Spatio-Temporal Epidemiol.*, **16** (2016), 1–20. <https://doi.org/10.1016/j.sste.2015.11.005>
9. A. O. Finley, S. Banerjee, B. P. Carlin, spBayes: An R package for univariate and multivariate hierarchical point-referenced spatial models, *J. Stat. Software*, **19** (2007), 1–24. <https://doi.org/10.18637/jss.v019.i04>
10. A. O. Finley, S. Banerjee, A. Gelfand, spBayes for large univariate and multivariate point-referenced Spatio-Temporal data models, *J. Stat. Software*, **63** (2015), 1–28. <https://doi.org/10.18637/jss.v063.i13>

11. P. A. P. Moran, Notes on continuous stochastic phenomena, *Biometrika*, **37** (1950), 17–23. <https://doi.org/10.2307/2332142>
12. A. B. Lawson, A. B. Biggeri, D. Boehning, E. Lesaffre, J. F. Viel, A. Clark, et al., Disease mapping models: an empirical evaluation. Disease Mapping Collaborative Group, *Stat. Med.*, **19** (2000), 2217–2241. [https://doi.org/10.1002/1097-0258\(20000915/30\)19:17/18%3C2217::AID-SIM565%3E3.0.CO;2-E](https://doi.org/10.1002/1097-0258(20000915/30)19:17/18%3C2217::AID-SIM565%3E3.0.CO;2-E)
13. T. McKinley, A. R. Cook, R. Deardon, Inference in epidemic models without likelihoods, *Int. J. Biostat.*, **5** (2009), 1–40. <https://doi.org/10.2202/1557-4679.1171>
14. S. Ng, N. E. Basta, B. J. Cowling, Association between temperature, humidity and ebolavirus disease outbreaks in Africa, 1976 to 2014, *Euro Surveill.*, **19** (2014), 1–11. <https://doi.org/10.2807/1560-7917.ES2014.19.35.20892>
15. R. J. Millettich, N. Diawara, H. A. Jeng, Modelling of the deaths due to Ebola virus disease outbreak in western Africa, *Int. J. Stats. Med.*, **4** (2015), 306–321. <https://doi.org/10.6000/1929-6029.2015.04.04.1>
16. L. Anselin, Local indicators of spatial association – LISA, *Geogr. Anal.*, **27** (1995), 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
17. K. Tawiah, W. A. Iddrisu, K. A. Asosega, Zero-Inflated time series modelling of COVID-19 deaths in Ghana, *J. Environ. Public Health*, (2021), 1–9. <https://doi.org/10.1155/2021/5543977>
18. S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, et al., Mobility network models of COVID-19 explain inequities and inform reopening, *Nature*, **589** (2021), 82–87. <https://doi.org/10.1038/s41586-020-2923-3>
19. S. Comess, H. Wang, S. Holmes, C. Donnat, Statistical modeling for practical pooled testing during the COVID-19 pandemic, preprint, arXiv: 2107.05619.
20. Institute for Health Metrics and Evaluation (IHME) COVID-19 Forecasting Team, Modeling COVID-19 scenarios for the United States. *Nat. Med.*, **27** (2021), 94–105. <https://doi.org/10.1038/s41591-020-1132-9>
21. V. H. Moreau, Forecast predictions for the COVID-19 pandemic in Brazil by statistical modeling using the Weibull distribution for daily new cases and deaths, *Braz. J. Microbiol.*, **51** (2020), 1109–1115. <https://doi.org/10.1007/s42770-020-00331-z>
22. J. Tolles, T. Luong, Modeling epidemics with compartmental models, *JAMA*, **323** (2020), 2515–2516. <https://doi.org/10.1001/jama.2020.8420>
23. S. H. Woolf, D. A. Chapman, R. T. Sabo, E. B. Zimmerman, Excess deaths from COVID-19 and other causes in the US, March 1, 2020, to January 2, 2021, *JAMA*, **325** (2021), 1786–1789. <https://doi.org/10.1001/jama.2021.5199>
24. A. B. Lawson, *Bayesian Disease Mapping: Hierarchical modelling in spatial epidemiology*, 2nd edition, CRC Press, Boca Raton, FL, 2013.
25. F. Rigat, M. D. Gunst, J. V. Pelt, Bayesian modelling and analysis of spatio-temporal neuronal networks, *Bayesian Anal.*, **1** (2006), 733–764. <https://doi.org/10.1214/06-BA124>
26. L. Knorr-Held, Bayesian modelling of inseparable Space-Time variation in disease risk, *Stat. Med.*, **19** (2000), 2555–2567. [https://doi.org/10.1002/1097-0258\(20000915/30\)19:17/18<2555::AID-SIM587>3.0.CO;2-%23](https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-%23)

27. P. Boedeker, Hierarchical linear modeling with maximum likelihood, restricted maximum likelihood, and fully Bayesian estimation, *Pract. Assess. Res. Evaluation*, **22** (2017). <https://doi.org/10.7275/5vvy-8613>
28. P. Giudici, P. Pagnottoni, A. Spelta, Network self-exciting point processes to measure health impacts of COVID-19, *J. R. Stat. Soc. Ser. A: Stat. Soc.*, 2023. <https://doi.org/10.1093/jrsssa/qnac006>
29. P. Giudici, B. Tarantino, A. Roy, Bayesian time-varying autoregressive models of COVID-19 epidemics, *Biomet. J.*, **65** (2023), 2200054. <https://doi.org/10.1002/bimj.202200054>
30. A. Paez, F. A. Lopez, T. Menezes, R. Cavalcanti, M. G. D. R. Pitta, A Spatio-Temporal analysis of the environmental correlates of COVID-19 incidence in Spain, *Geogr. Anal.*, **53** (2021), 397–421. <https://doi.org/10.1111/gean.12241>
31. M. De Angelis, C. Durastanti, M. Giovannoni, L. Moretti, Spatio-temporal distribution pattern of COVID-19 in the Northern Italy during the first-wave scenario: The role of the highway network, *Transp. Res. Interdiscip. Perspect.*, **15** (2022), 100646. <https://doi.org/10.1016%2Fj.trip.2022.100646>
32. J. Park, S. Yi, W. Chang, J. Mateu, A Spatio-Temporal dirichlet process mixture model for coronavirus disease-19, preprint, arxiv: 2207.06587.
33. M. Bizzarri, M. D. Traglia, A. Giuliani, A. Vestri, V. Fedeli, A. Prestininzi, New statistical RI index allow to better track the dynamics of covid-19 outbreak in Italy, *Sci. Rep.*, **10** (2020). <https://doi.org/10.1038/s41598-020-79039-x>
34. T. Guenther, M. Czech-Sioli, D. Indenbirken, A. Robitailles, P. Tenhaken, M. Exner, et al., Investigation of a superspreading event preceding the largest meat processing plant-related SARS-Coronavirus 2 outbreak in Germany, *SSRN J.*, 2020. <https://doi.org/10.2139/ssrn.3654517>
35. D. Lee, A. Rushworth, G. Napier, Spatio-Temporal areal unit modelling in R with conditional autoregressive priors using the CARBayesST package, *J. Stat. Softw.*, **84** (2018), 1–39. <https://doi.org/10.18637/jss.v084.i09>
36. A. Rushworth, D. Lee, R. Mitchell, A Spatio-Temporal model for estimating the Long-Term effects of air pollution on respiratory hospital admissions in greater London, *Spat. Spatiotemporal Epidemiol.*, **10** (2014), 29–38. <https://doi.org/10.1016/j.sste.2014.05.001>
37. A. Das, A. Sampson, C. Lainscsek, L. Muller, W. Lin, J. Doyle, et al., Interpretation of the precision matrix and its application in estimating sparse brain connectivity during sleep spindles from human electrocorticography recording, *Neural Comput.*, **29** (2017), 603–642. https://doi.org/10.1162/NECO_a_00936
38. B. P. Carlin, T. A. Louis, *Bayesian methods for data analysis*, 3rd edition, CRC Press, Boca Raton, FL, 2008.
39. B. G. Leroux, X. Lei, N. Breslow, Estimation of disease rates in small areas: A new mixed model for spatial dependence. Statistical models in epidemiology, the environment, and clinical trials, *The IMA Volumes in Mathematics and its Applications*, Springer, New York, NY, **116** (2000), 179–191. https://doi.org/10.1007/978-1-4612-1284-3_4
40. T. Zhang, G. Lin, On Moran's I coefficient under heterogeneity, *Comput. Stat. Data Anal.*, **95** (2016), 83–94. <https://doi.org/10.1016/j.csda.2015.09.010>
41. D. Spiegelhalter, N. Best, B. Carlin, A. Van der Linde, Bayesian measures of model complexity and fit, *J. R. Stat. Soc. Ser. B*, **64** (2002), 583–639. <https://doi.org/10.1111/1467-9868.00353>

42. H. Y. Bar, M. T. Wells, On graphical models and convex geometry, preprint, arXiv: 2106.14255.
43. P. D. Jong, C. Sprenger, F. V. Veen, On extreme values of Moran's I and Geary's c (spatial autocorrelation), *Geogr. Anal.*, **16** (1984), 17–24.
<https://doi.org/10.1111/j.1538-4632.1984.tb00797.x>
44. J. M. Kim, Copula dynamic conditional correlation and functional principal component analysis of COVID-19 mortality in the United States, *Axioms*, **11** (2022), 619.
<https://doi.org/10.3390/axioms11110619>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>).