Received: 29 November 2021 | Revised: 24 May 2022 | Accepted: 4 June 2022

DOI: 10.1111/tbed.14623

Transboundary and Emerging Diseases WILEY

ORIGINAL ARTICLE

Exploring the risk factors of COVID-19 Delta variant in the United States based on Bayesian spatio-temporal analysis

Shaopei MA[1] | **Xueliang Zhang**[2] | **Kai Wang**[2] | **Liping Zhang**[2] | **Lei Wang**[2] | **Ting Zeng**[2] | **Man-Lai Tang**[3] | **Maozai Tian**[1,2]

[1]Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, China

[2]Department of Medical Engineering and Technology, Xinjiang Medical University Urumqi, China

[3]Mathematical Sciences, Brunel University, Uxbridge, London, UK

Correspondence
Maozai Tian, Department of Medical Engineering and Technology, Xinjiang Medical University Urumqi, 830011, China.
Email: mztian@ruc.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 11861042; China Statistical Research Project, Grant/Award Number: 2020LZ25

Abstract

The transmission of coronavirus disease-2019 (COVID-19) epidemic is a global emergency, which is worsened by the genetic mutations of SARS-CoV-2. However, till date, few statistical studies have researched the COVID-19 spread patterns in terms of the variant cases. Hence, this paper aims to explore the associated risk factors of Delta variant, the most contagious strain of COVID-19. The study collected the state-level COVID-19 Delta variant cases in the United States during a 12-week period and included potential environmental, socioeconomic, and public prevention factors as independent variables. Instead of regarding the covariate effects as constant, this paper proposes a flexible Bayesian hierarchical model with spatio-temporally varying coefficients to account for data heterogeneity. The method enables us to cluster the states into distinctive groups based on the temporal trends of the coefficients and simultaneously identify significant risk factors for each cluster. The findings contribute novel insight into the dynamics of covariate effects on the COVID-19 Delta variant over space and time, which could help the government develop targeted prevention measures for vulnerable regions based on the selected risk factors.

KEYWORDS
Bayesian analysis, COVID-19 variant cases, space-time model, spatial clusters, variable selection

1 | INTRODUCTION

The coronavirus disease 2019 (COVID-19) was declared a global pandemic by the World Health Organization on 11 March 2020 (WHO, 2020) and has caused a severe disruption to society as well as a great loss to life by overwhelming health care systems in many countries. Just as some signs of optimism were observed about the prevention and control of COVID-19, it was found that SARS-CoV-2, the cause of COVID-19, had gone through genetic mutations over time, resulting in the propagation of new virus variants. One of these new variants, the Delta (B.1.617.2) variant, was first identified in India in late 2020 and has spread rapidly around the globe. As of 14 October 2021, the Delta variant has been reported in 187 out of 194 World Health Organization member countries, becoming the predominant lineage worldwide (Mirror, 2021).

According to new estimates released by the Centers for Disease Control and Prevention (CDC), the highly contagious Delta strain could account for more than 80% of COVID-19 new cases in parts of the United States, leading to a huge surge of hospitalization in most states. Due to its extremely high transmissibility, the Delta variant has been listed as a variant of concern (VOC) by WHO (Ministry of Health, 2021), outcompeting the incipient and other variant lineages of COVID-19, such as B.1.617.1 (Kappa) and B.1.1.7 (Alpha) (Mlcochova et al., 2021). Therefore, researchers around the world pay particular attention to

the transmission of the Delta virus in the environment. This paper is focused on exploring the potential risk factors that may have significant effects on the spread of Delta variant.

There have been extensive studies on the contributing factors of the incipient COVID-19 virus, albeit with conflicting results. Many researchers found that environmental factors, such as temperature, humidity, wind speed and air pollution, play an important role in the transmission of COVID-19 epidemic (Ahmadi et al., 2020; Bashir et al., 2020; Tosepu et al., 2020; Xu et al., 2020). Demographic and socioeconomic factors, including population density, age distribution, income and so on, are also found crucial in controlling COVID-19 spread, with increased cases more likely observed in densely populated areas (Díaz de León-Martínez et al., 2020; Zheng et al., 2020). However, some other studies showed no evidence of association between COVID-19 and absolute humidity or temperature (Jamil et al., 2020; Xie & Zhu, 2020). In the research of Hamidi et al. (2020), population density is found not to be a significant driver, contrary to the previous results. Thus, there is still much uncertainty about the risk factors of COVID-19 transmission due to the rapid changes of data in various research scopes.

Many statistical methods, such as partial correlation coefficients (PCCs) (Ahmadi et al., 2020), ordinary least squares (OLS) and Bayesian hierarchical models (Millett et al., 2020) have been utilized to analyse the correlations between COVID-19 cases and the potential contributing factors. For example, Gayawan et al. (2020) used the Poisson hurdle model to explain excess zero counts of COVID-19 cases. Briz-Redón and Serrano-Aroca (2020) employed the separable random effects model to incorporate structured and unstructured components. However, the mentioned models often assume that covariate effects are constant and do not change over space and time, which may neglect the possible dynamics and staged significance of the coefficients. Based on geographically weighted regression models (GWRs), Luo et al. (2021) proposed a local regression method, showing the spatial variability of the correlations between multiple risk factors and the COVID-19 death rate. But the study only focused on the spatial dimension of the data, lacking consideration of temporal trends. Hence, it is necessary in this research to discover the spatio-temporal patterns of the covariate effects to help monitor health care access.

In this paper, preliminary analysis shows spatial heterogeneity in the distribution of Delta variant cases in the United States. Sannigrahi et al. (2020) pointed out that the uneven distribution of the COVID-19 confirmed cases across Europe could be attributed to the discrepant socio-demographic factors such as age and income. Inspired by their finding, this paper constructs a flexible Bayesian hierarchical model with spatio-temporally varying coefficients, which allows for heterogeneous temporal patterns of the covariate effects. Applied to the COVID-19 Delta variant cases in the United States, the proposed method enables us to cluster the states into distinctive groups based on the variation of covariate effects over time and to identify significant risk factors separately for each cluster. The study will pave way for better understanding of the transmission of COVID-19 Delta virus and may help the government establish appropriate precautions according to the conditions of the regions.

2 | MATERIALS AND METHODS

2.1 | Study site

In this paper, the spatial domain of interest includes the lower 48 contiguous states and the District of Columbia in the United States, excluding the states of Alaska and Hawaii from analysis, since they do not border the continental United States. The total population of the United States is 330 million, of which California and Texas are the two most populous states. Most of the states are on the east or west coast and have fairly dense populations.

2.2 | Data collection

The detection of Delta variant needs viral genomes sequencing in the laboratory. We obtained the weekly proportion of Delta variant cases relative to all samples sequenced in each state of the United States from the website of cov-spectrum (https://cov-spectrum.ethz.ch/explore/United%20States/AllSamples/AllTimes) for the period of week 22 (24 May 2021) to week 33 (15 August 2021). Then, the weekly number of Delta variant cases was calculated via the proportion of the variant multiplied by the cumulative number of all confirmed COVID-19 cases in the corresponding week, which was collected by state from https://outbreak.info/location-reports?loc=USA_US-GA.

Potential contributing factors of the Delta variant are considered as independent variables, including the environmental factors, demographics and socioeconomic factors and public interventions. The environmental variables contain maximum temperature (tem_max), average temperature (tem_ave), minimum temperature (tem_min), average relative humidity (hum_ave), average wind speed (wind_ave) and cumulative precipitation (preci_sum). These data were collected on a week scale to match the response and obtained by integrating the daily information from 'Weather Underground' website from 10 May 2021 to 16 August 2021, which is 2 weeks prior to the confirmed Delta cases since the COVID-19 virus usually has a 7-day or longer incubation period (Briz-Redón & Serrano-Aroca, 2020).

The demographics and socioeconomic factors include state-level population density, proportion of elderly people (≥ 65-years-old) and per capita personal income in the first quarter of 2021, following Hamidi et al. (2020) and Pequeno et al. (2020). Data on population and income were collected from Bureau of Economic Analysis (https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=1&acrdn=2). The proportion of the elderly in each state was obtained from Population Reference Bureau (PRB), which is available at https://www.prb.org/usdata/indicator/age65/snapshot.

Public interventions of primary interest in this paper include weekly COVID-19 vaccination coverage and per cent change in community mobility in the period of week 20–33. To explore the effect of vaccination, the cumulative share of population vaccinated with 1+ doses in each state was collected from KFF COVID-19 Vaccine Monitor (https://www.kff.org/coronavirus-covid-19/issue-brief/state-covid-19-data-and-policy-actions/). The per cent change in

community mobility was obtained from Google website, relative to a baseline index for a typical day before the onset of COVID-19, which is available at https://www.google.com/covid19/mobility/. Since there is strong multi-collinearity among the different community mobility patterns, the paper uses only transit station mobility. Note that meteorological variables and public intervention factors vary with space and time while the demographics and social factors only change across different states.

2.3 | Data analyses

2.3.1 | Bayesian spatio-temporal modelling

Let $\{Y_{it}\}$ denote the observed number of cases in the ith area ($i = 1, \ldots, N$) and the tth time point ($t = 1, \ldots, T$), which is assumed to follow a Poisson distribution as

$$Y_{it} \mid \theta_{it} \sim \text{Poisson}\left(E_{it}\theta_{it}\right), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

where θ_{it} is the relative risk and E_{it} is the expected case count. The logarithm of relative risk is then modeled with spatio-temporally varying coefficients as follows:

$$\log\left(\theta_{it}\right) = \beta_0 + \sum_{p=1}^{P_1} X_{ip}\beta_{itp} + \sum_{p=P_1+1}^{P} X_{i(t-\text{lag})p}\beta_{itp} + \varepsilon_{it}$$

where β_0 is the intercept, X_{ip} ($p = 1, \ldots, P_1$) are the demographics or socioeconomic covariates only changing with state i, $X_{i(t-\text{lag})p}$ ($p = P_1 + 1, \ldots, P$) are the time-varying meteorological and intervention covariates with temporal lag, and $\boldsymbol{\beta}_{it} = (\beta_{it1}, \ldots, \beta_{itP})$ are the corresponding coefficients in area i at time t. The space–time random component ε_{it} represents the structure not captured by the covariate effects. The temporally lagged specification is utilized for consideration of the COVID-19 incubation period, which establishes the relationship between the relative risk at time t and the meteorological and intervention covariates at time t − lag. On the basis of the existing research, the incubation time of COVID-19 is usually 7–14 days, and thus in our real data analysis, the temporal lag parameter is considered as lag = 1 or 2 (weeks).

According to the previous studies (Lawson et al., 2014; Lee et al., 2017), the time-varying coefficients may have locally homogeneous trends due to spatial adjacency. Thus, this paper models the regression coefficients as cluster-specific in the sense that the areas in the same cluster have common covariate effects with particular temporal patterns, which is different from the patterns in other clusters. Following Choi et al. (2012), the variable coefficients $\boldsymbol{\beta}_{it}$ is modelled as

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{C(i),t}$$

where $C(i)(= 1, \ldots, S)$ denotes the spatial cluster of area i and S is the number of clusters.

A multi-nomial distribution is assumed for $C(i)$ as follows

$$C(i) \sim \text{Multi}\left(w_{i1}, \ldots, w_{iS}\right)$$

where $w_{is} \geq 0$ indicates the probability of area i being assigned to cluster s, and thus $\sum_{s=1}^{S} w_{is} = 1$. For modelling purposes, unstandardized weight $w_{is}^* \geq 0$ is utilized in place of the probability w_{is}, with the transformation of $w_{is} = \frac{w_{is}^*}{\sum_{s=1}^{S} w_{is}^*}$. Then the weight w_{is}^* is modeled with the following log-normal distribution to account for the spatial dependence in the small areas,

$$\log\left(w_{is}^*\right) \sim N\left(\eta_{is}, \sigma_s^2\right)$$

where η_{is} is the mean with spatial correlation and σ_s^2 is the variance. The widely used intrinsic conditional autoregressive (ICAR) distribution (Besag et al., 1991) is employed to further model the mean η_{is} as

$$\eta_{is} \mid \eta_{i's}, i' \neq i \sim N\left(\frac{1}{n_i}\sum_{i' \sim i}\eta_{i's}, \frac{\sigma_{\eta_s}^2}{n_i}\right)$$

where n_i is the number of neighbours for area i, $i' \sim i$ represents that areas i' and i are neighbours and $\sigma_{\eta_s}^2$ controls the degree of spatial volatility. The full conditional distribution given earlier, denoted as $\eta_{is} \sim \text{ICAR}(\sigma_{\eta_s}^2)$, holds independently for each s.

When area i belongs to cluster s, the coefficient vector in area i is denoted as $\boldsymbol{\beta}_{st} = (\beta_{st1}, \ldots, \beta_{stP})'$. To select significant covariates for each spatial cluster, the pth covariate effect β_{stp} is hierarchically modelled as

$$\beta_{stp} = \delta_{sp} \times \lambda_{stp}, \quad s = 1, \ldots, S;\ t = 1, \ldots, T;\ p = 1, \ldots, P$$

where δ_{sp} is a binary entry parameter controlling the selection of the pth covariate in cluster s and λ_{stp} denotes the temporal pattern of the covariate effect. If $\delta_{sp} = 0$, then the pth covariate X_p is not significant in cluster s. Otherwise, $\delta_{sp} = 1$ denotes that X_p is selected in cluster s, with temporally dependent coefficient $\beta_{stp} = \lambda_{stp}$.

The binary selector δ_{sp} flexibly switches in or out of the pth covariate in cluster s, following a Bernoulli distribution as

$$\delta_{sp} \mid \phi_{sp} \sim \text{Ber}\left(\phi_{sp}\right),$$

where ϕ_{sp} denotes the selection probability. Since spatial dependence across clusters is generally not evident (Adin et al., 2019), an unstructured prior distribution is used to model ϕ_{sp} as

$$\text{logit}\left(\phi_{sp}\right) = \mu_p + \xi_{sp},$$

where μ_p is the intercept and $\xi_{sp} \sim N(0, \sigma_{\xi_{sp}}^2)$ is the spatially uncorrelated term.

The temporal patterns of λ_{stp} vary with spatial clusters and many alternative structures, such as autoregressive models and random walks, could be taken into consideration. In this paper, an AR(1) process

is employed to model the temporal trend of λ_{stp} as

$$\lambda_{stp}|\lambda_{s,t-1,p} \sim N\left(\rho_{\lambda_{sp}}\lambda_{s,t-1,p}, \sigma^2_{\lambda_{sp}}\right), \quad s = 1, \ldots, S; \quad p = 1, \ldots, P.$$

Following from Briz-Redón and Serrano-Aroca (2020), the random component ε_{it} is modelled as $\varepsilon_{it} = u_i + v_i + \gamma_t + \zeta_t$, where u_i is a spatially unstructured random effect following a normal distribution as $u_i \sim N(0, \sigma_u^2)$, v_i is a spatially structured effect following an ICAR distribution as $v_i \sim \text{ICAR}(\sigma_v^2)$, γ_t is a temporally unstructured effect following $\gamma_t \sim N(0, \sigma_\gamma^2)$ and ζ_t is a temporally structured effect following an AR(1) distribution as $\zeta_t \sim N(\rho_\zeta \zeta_{t-1}, \sigma_\zeta^2)$ with $0 < \rho_\zeta < 1$.

Our main focus is on the regression coefficients and spatial clustering, i.e., the estimation of $\boldsymbol{\beta}_{it}$, $C(i)$, δ_{sp}, and λ_{stp}, which can be computed from the posterior distributions of the parameters based on the hierarchical models above.

2.3.2 | Bayesian inference

The likelihood function of the observed counts $\mathbf{Y} = \{Y_{it}, i = 1, \ldots, N, t = 1, \ldots, T\}$ is expressed as follows

$$L(\Theta|\mathbf{Y}) = \prod_{i=1}^{N}\prod_{t=1}^{T} \text{Poisson}\left(Y_{it}|E_{it}, \mathbf{X}_{it}, \beta_0, \boldsymbol{\beta}_{C(i),t}\right)$$

where Θ is the set of parameters in the hierarchical models. The prior distributions of the intercept parameters β_0 and μ_p are specified as $\beta_0 \sim N(0, \sigma_0^2)$ and $\mu_p \sim N(0, \sigma_{\mu_p}^2)$, respectively. All the standard deviation parameters are assumed to follow uniform distributions as $\sigma_s, \sigma_{\eta_s}, \sigma_{\lambda_{sp}}, \sigma_{\xi_{sp}} \sim \text{Uniform}(0, c)$, where c is a constant (Gelman, 2006). The temporal parameter $\rho_{\lambda_{sp}}$ in the structure of λ_{stp} is considered to follow $\rho_{\lambda_{sp}} \sim \text{Uniform}(0, 1)$.

Bayesian inference is conducted for the parameters of interest based on the posterior distributions. MCMC algorithm is carried out via software R and WinBUGS. The convergence of posterior sampling is checked by autocorrelation functions, trace plots and Geweke convergence diagnostics.

Since the indicator $C(i)$ is a categorical variable, its posterior mode is used as estimates. As for variable selection in each cluster, the marginal posterior selection probability, $P(\delta_{sp} = 1|\mathbf{Y})$, is compared with the cut-off value of 0.5 (Barbieri & Berger, 2004; Lawson et al., 2012). If the estimated probability is greater than 0.5, then $\hat{\delta}_{sp} = 1$ and X_p is selected in cluster s; otherwise, X_p is viewed as unimportant for cluster s and excluded from the model. The remaining parameters are estimated by the posterior means.

2.4 | Ethics statement

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. No ethical approval was required as this study did not collect any identifiable personal information.

3 | RESULTS

3.1 | Descriptive analysis for COVID-19 Delta cases

Figure 1 displays the transmission of COVID-19 Delta variant in the United States. The left panel shows the cumulative number of Delta variant cases (per 100,000 population) in the period between week 22 and week 33. It is seen that the Delta virus has spread throughout the United States with the southeastern areas suffering the most severe invasion. Many states such as Nevada and Wyoming in the western district are also in a serious condition. The right panel depicts the weekly standardized morbidity ratio (SMR) of each state, defined as the number of observed Delta cases divided by the number of expected cases calculated by the internal standardization method (Banerjee et al., 2003), with larger values representing higher disease risks. The SMRs in several states reach the peak at the beginning of the study period and gradually fall down with time while some others show a slowly upward tendency, such as Louisiana and Florida.

Figure 2 displays the standardized incidence maps for the United States in 6 weeks. There is obvious clustering phenomenon possibly due to the spatial adjacency effects and the influence of common risk factors. To further quantify the strength of spatial correlation among the state-level incidences, we first apply Global Moran's I statistic (Banerjee et al., 2003) to the SMRs of the 49 regions. It is shown that the value of Global Moran's I is reasonably large for each week and the significance test suggests very strong evidence for positive spatial correlation. Then Local Moran's I is employed to identify spatial clusters. Taking the SMRs in week 25 for example, the test results indicate that the states of Utah, Nevada, Wyoming, Missouri and Arkansas are significant hotspots, which is consistent with the finding from the top-middle panel of Figure 2. Similar results are obtained for the other weeks and are omitted for brevity. It is easy to find that the hotspots gradually move from the west and central to southeastern districts, which facilitates the tendency for the Delta virus to spread nationwide. The SMRs have an evident spatio-temporal variation throughout the country and hence, the dynamic changes motivate us to explore the contributing risk factors in each state.

3.2 | Exploratory analysis for the covariates

To avoid multi-collinearity issues, the Spearman-rank correlation test is applied to the six environmental variables. The results are shown in Figure 3. Not surprisingly, the three temperature-related variables, tem_max, tem_ave and tem_min are highly correlated to each other. There is also a close relationship between humidity (hum_ave) and precipitation (preci_sum) and both of them are negatively connected with tem_max. Therefore, only three environmental factors, tem_ave, hum_ave and wind_ave are chosen to be incorporated into the model.

To explore the variation tendency of the three climate variables, boxplots are drawn in Figure 4 for all the states from week 21 to week

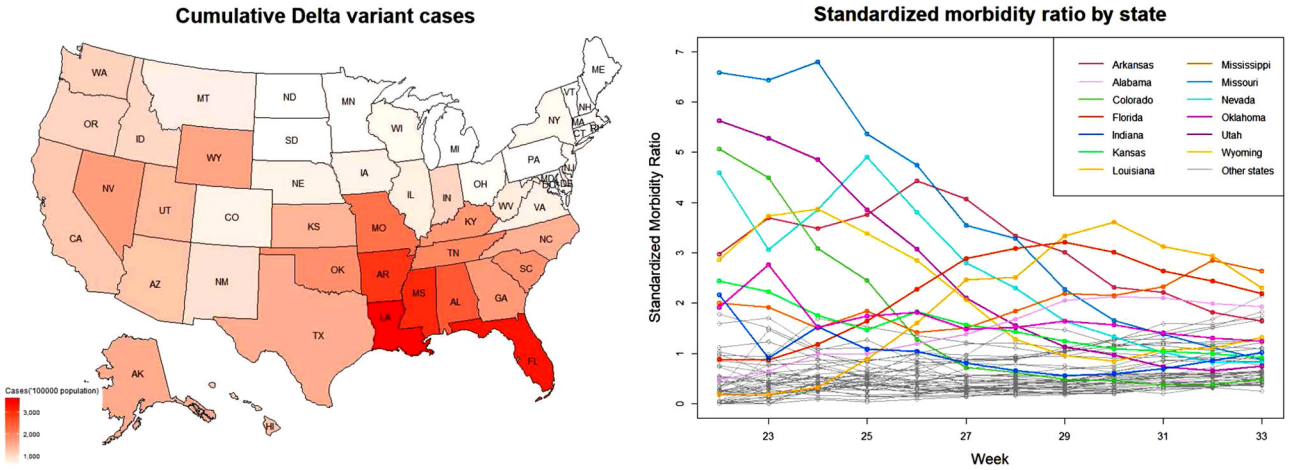

FIGURE 1 The spread of COVID-19 Delta variant in the United States. Left: Cumulative Delta cases between week 22 and week 33. Right: The weekly standardized morbidity (incidence) ratios for the states in the study period

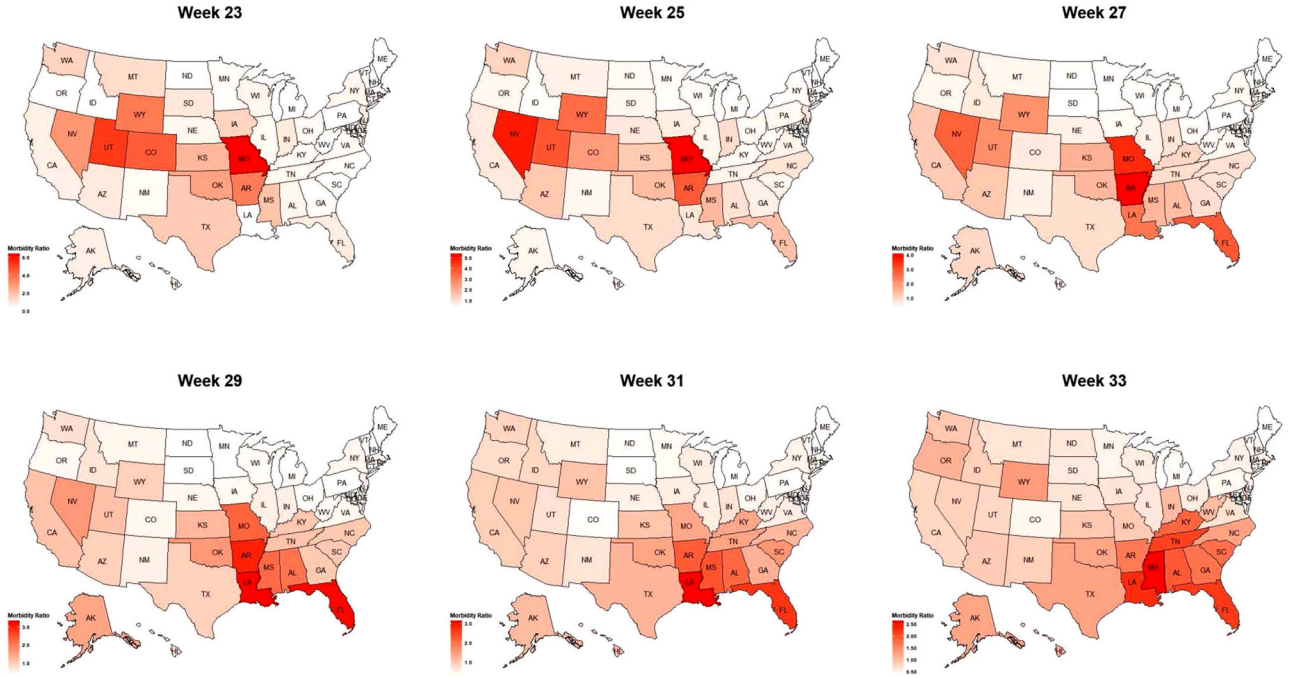

FIGURE 2 State maps of standardized morbidity ratios of COVID-19 Delta variant in six selected weeks in the United States

32, which is 1 week prior to the confirmed cases. The average temperature jumps upward at first and climbs slowly in the subsequent weeks. On the whole, the average humidity and average wind speed exhibit earlier increase and later decrease trends, although the fluctuations are not in sync. The maps of these considered environmental and demographics and socio-economic covariates, which are omitted for conciseness, also show significant spatial effects. Thus, it is reasonable to consider the spatio-temporally varying coefficients of these risk factors to explore the underlying patterns of significant covariates for the Delta variant cases.

3.3 | Spatio-temporal modelling analysis

3.3.1 | Comparisons with the alternative models

Based on MCMC algorithms, we obtain Bayesian posterior samples of 30,000 iterations, with 10,000 burn-in samples. A thinning rate of 10 is used to avoid high sample autocorrelations; thus, parameter estimation is finally conducted with 2000 samples.

The temporal lag parameter and the number of clusters are determined based on model assessment criteria: deviance information

FIGURE 3 Spearman-rank correlation of the six environmental variables in the analysis of COVID-19 Delta cases

criterion (DIC_3) (Celeux et al., 2006), marginal predictive likelihood (MPL) (Gelfand & Dey, 1994) and mean squared prediction error (MSPE). In general, the model with smaller values of DIC_3, MSPE and a larger value of MPL provides better fitness for the data. The considered range for the number of clusters is from 4 to 13. It is found that overall, the models with lag = 1 fit the data better than the ones with lag = 2. And in the setting of lag = 1, DIC_3 and MSPE prefer the model with 7 clusters, while MPL achieves its optimum with S = 8. Since there is no significant difference between the MPL values of the two models, the number of clusters S = 7 is chosen to be the best. Therefore, the subsequent studies are conducted based on the specification of lag = 1 and S = 7.

The proposed model is also compared with four alternative models, which are constructed as follows:

TABLE 1 Comparison results of the considered models in terms of the assessment criteria

Model	DIC_3	MPL	MSPE
Model 1	5894.2	−2920.4	399.6
Model 2	5782.6	−2915.7	397.2
Model 3	5452.9	−2761.3	252.3
Model 4	5403.7	−2784.2	241.5
Model 5	5361.4	−2717.5	236.8

1. Model 1: Bayesian hierarchical model with only variable selection procedures, but no spatial clustering: $\beta_{itp} = \delta_{ip} \times \lambda_{tp}$; δ_{ip} is a binary variable selector, similar to δ_{sp} in the previous section, and λ_{tp} follows an AR(1) process.
2. Model 2: Bayesian hierarchical model with spatial cluster indicators, but no variable selection: $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{C(i),t}$ as proposed, but $\beta_{stp}|\beta_{s,t-1,p} \sim N(\rho_{sp}\beta_{s,t-1,p}, \sigma^2_{sp})$ with $C(i) = s$.
3. Model 3: Bayesian spatio-temporal model with only unstructured random effects u_i and γ_t: $\log(\theta_{it}) = \beta_0 + \sum_p X_{ip}\beta_{itp} + \sum_p X_{i(t-lag)p}\beta_{itp} + u_i + \gamma_t$, where $u_i \sim N(0, \sigma^2_u)$, $\gamma_t \sim N(0, \sigma^2_\gamma)$.
4. Model 4: Bayesian spatio-temporal model with only structured random effects v_i and ζ_t: $\log(\theta_{it}) = \beta_0 + \sum_p X_{ip}\beta_{itp} + \sum_p X_{i(t-lag)p}\beta_{itp} + v_i + \zeta_t$, where $v_i \sim \text{ICAR}(\sigma^2_v)$, $\zeta_t \sim N(\rho_\zeta \zeta_{t-1}, \sigma^2_\zeta)$ with $0 < \rho_\zeta < 1$.
5. Model 5 (the proposed model): Bayesian spatio-temporal model with mixed random effects ($\varepsilon_{it} = u_i + v_i + \gamma_t + \zeta_t$), which simultaneously allows for variable selection and spatial clustering.

Table 1 presents the comparison results for the considered models in terms of the model assessment criteria mentioned earlier. It is seen that Models 1 and 2 have much larger values of DIC_3 and MSPE than the other models, and their MPL values are the smallest, which indicates that only taking into account variable selection (Model 1) or spatial clustering (Model 2) is inadequate for the data. In particular, it seems that leaving spatial clustering out of consideration has more serious impact on the model performance. Models 3 and 4 behave a bit better, but are still inferior to Model 5 (the proposed model),

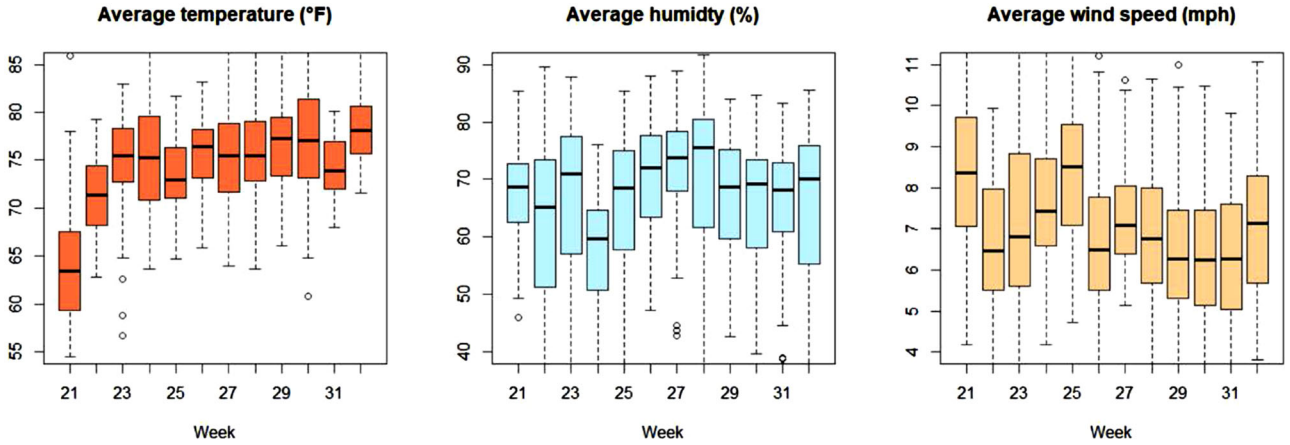

FIGURE 4 Boxplots of the three climate variables in the states of America from week 21 to week 32

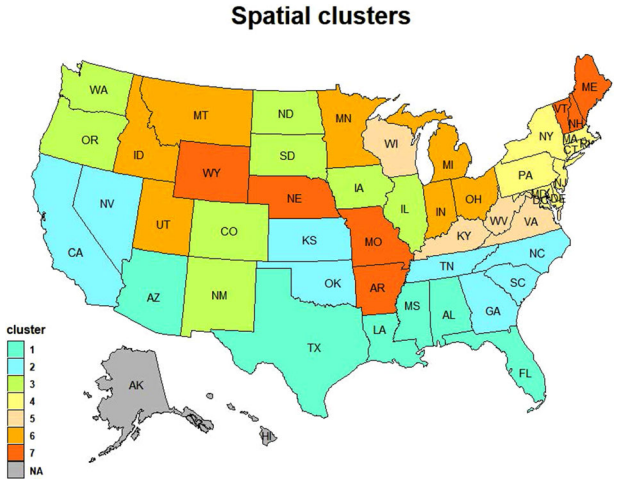

FIGURE 5 The map of spatial clusters for the covariate effects on the Delta variant cases with the number of clusters S = 7. The gray regions, including states of Alaska and Hawaii, are not considered in the model due to their non-adjacency with the mainland

showing that both structured and unstructured random effects provide enhanced results for the model. Thus, the proposed model is the best fit one, based on which the estimation results are displayed in the following sections.

3.3.2 | Spatial clustering of covariate effects

The estimated spatial cluster indicators are shown in the map of Figure 5. There are 6 states in the first group, located on the southern border of the United States. Both the second and the third groups contain 8 states scattered among 3 separated districts. Including 9 states, the fourth group is concentrated on the Atlantic coast and New England region. The fifth group is relatively small, with only 4 states: Wisconsin, Kentucky, West Virginia and Virginia. The states in the sixth group are situated in two districts: the Great Lakes region and the western areas. The last group contains 3 states in the Northeast and 4 states in the Midwest.

It is seen from Figure 5 that the distribution of spatial clusters is well proportioned, with states in the same cluster tending to be geographically contiguous, which makes sense to share common risk factors in the neighbourhood. Meanwhile, spatial discontinuity is also allowed in each cluster. For example, although Oregon, South Dakota and Colorado are separated from each other, they can still be clustered together in group 3. These features verify the flexibility of the proposed model in terms of the clustering patterns for the states.

3.3.3 | Covariate effects estimation

Figure 6 shows the variable selection results with the maps of estimated binary selector $\hat{\delta}_{sp}, p = 1, \ldots, 8$. The colored regions represent that the covariate is selected in the corresponding states. Evidently, the covariate effects on COVID-19 Delta cases are spatially varying. In combination with the spatial clusters in Figure 5, we could see that average temperature is widely selected in 5 clusters, including 35 states in the mainland. Average humidity mainly affects the west coast and central regions of America, covering 4 clusters: (2,3,4,7), while average wind speed has a significant impact on clusters 3 and 6 in the north. Population density plays an important part in clusters 1,3,5 and 7, where cluster 5 is the most densely populated region in the United States. The proportion of elderly people is connected with the Delta variant cases in 4 clusters, which are concentrated in the western and southeastern regions. Personal income is selected in clusters (2,4,5), which contains 21 states. Vaccination coverage is a significant factor in most states, including clusters 2,3,4,5 and 7. The per cent change in transit station mobility also has a broad effect on the spread of Delta variant, specifically in clusters (1,2,4,5,7). The spatial variation of significant contributing factors confirms the importance of variable selection in each cluster for COVID-19 Delta variant.

The coefficient estimation results for $\beta_{stp}, s = 1, \ldots, 7, p = 1, \ldots, 8$ are displayed in Figure 7. Detailed values of the estimates and 95% credible intervals for the coefficients could be found in the Tables S1–S8 of the supplement. Obviously, the covariate effects are dynamically changing with space and time. The unselected covariates have much smaller coefficients close to zero, which is coincident with the results in Figure 6. It is known that the three environmental variables, personal income and vaccination coverage in the selected clusters are all negatively associated with the disease relative ratios, while higher population density, proportion of elderly people (>65-years-old) and per cent change in transit station mobility are connected with increased risk of Delta variant infection in the domain of selected states.

As the COVID-19 epidemic evolves, it is critically important for the whole society to understand the changes of the associations between the contributing factors and the relative risks over time. From the coefficient estimation results in Figure 7, we could see that similar to the concurrent regression model, the covariate effects vary over the process of the pandemic. During the study period, the states in America are between spring and summer. Exploratory analysis in Section 3.2 shows that the three meteorological variables have continuous fluctuations up and down, and correspondingly, Figure 7 indicates that there are marked variations in the meteorological covariate effects for most selected clusters.

It is worth noting that although the demographics and socioeconomic factors are measured at baseline, their effects on the disease risk are not constant over the follow-up time, which is similar to the results of Carroll et al. (2020) and Briz-Redón and Serrano-Aroca (2022). This is because the spread of Delta variant is quite rapid, and due to the pressure of reopening, the prevention measures adopted by state governments are constantly changing, such as the magnitude of mandated lockdowns and vaccine promotion. These result in different patterns at different time points, and then the effects of these baseline variables may change over weeks. In particular, according to Gallagher et al. (2021), in areas with high vaccination coverage and low community mobility, the transmission of COVID-19 could be significantly

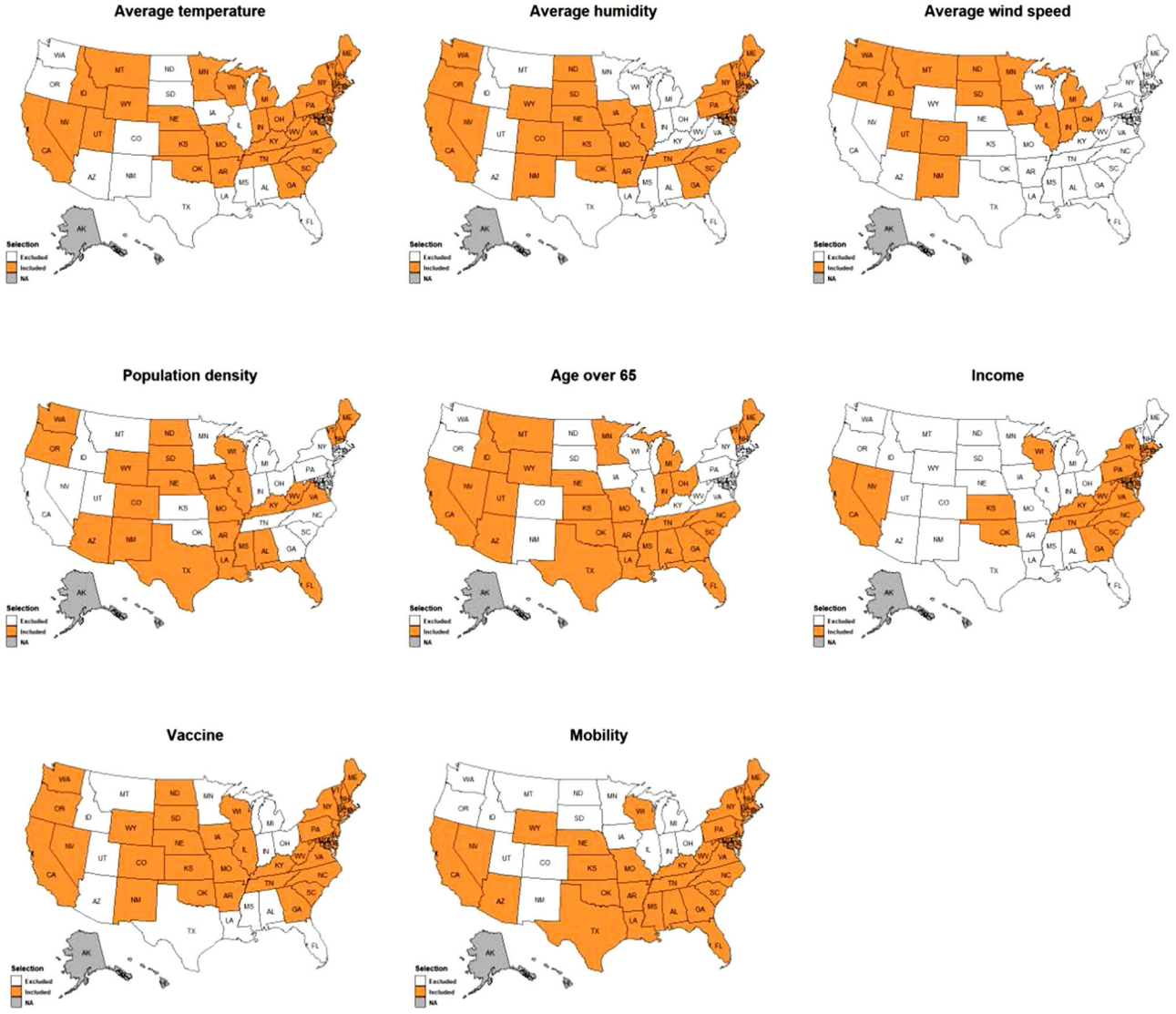

FIGURE 6 Spatial variable selection results for the eight factors in the analysis of COVID-19 Delta variant cases

interrupted, and thus the effect of population density will reduce in a certain degree. In other cases, the effects of these baseline variables are likely to fluctuate week by week due to the complex changes of the environments.

For the intervention variables, vaccination coverage and per cent change in mobility also have varying effects over time. In particular, vaccination coverage shows gradually increasing negative associations in the selected states. According to the study of Briz-Redón and Serrano-Aroca (2022), the social immunity effect is significant only when the vaccination coverage reaches a certain level. Therefore, in the states with low coverage rate, vaccination only provides relatively low protection efficacy, while in other states, with the continuous improvement of vaccination coverage, the protection mechanism for susceptible people is gradually enhanced.

The overall effects of covariates in the clusters are presented in Table 2, which are calculated by averaging the time-varying covariate effects in the period of study. It is seen from Figure 7 and Table 2 that both temporal trends and overall sizes of the coefficients in different clusters are distinct from each other. Therefore, it is necessary to consider spatial clustering determined by the heterogeneous temporal patterns of covariate effects. The spatio-temporally varying coefficients could accurately capture the factor effects on the disease relative risks of COVID-19 Delta variant.

3.3.4 | COVID-19 Delta cases estimation

The relative risks of COVID-19 Delta variant can be estimated for the time period of study. Figure 8 presents the posterior estimates and the corresponding standard deviations for the relative risks θ_{it} in the spatial domain of interest in Week 24, that is, $i = 1, \ldots, 49$ and $t = 3$, while the detailed estimation results for all the time points ($t = 1, \ldots, 12$) can be found in Tables S9–S14 of the supplement. The state names are sorted in alphabetical order, with the corresponding ID numbers

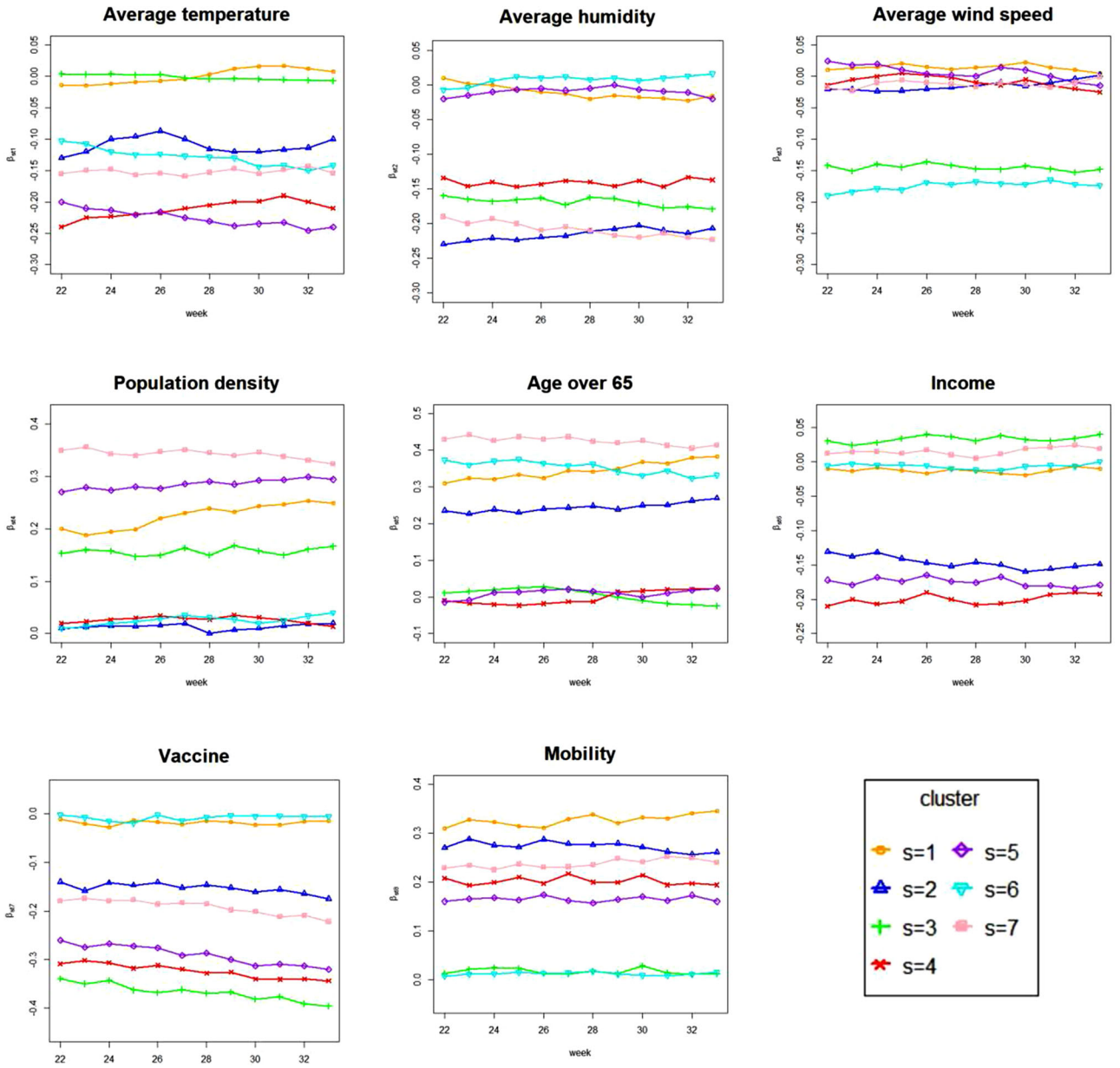

FIGURE 7 The temporal trends of the estimated coefficients for the eight potential factors in the clusters. There are S = 7 clusters marked in different colours

TABLE 2 The overall effects of covariates in the clusters for the period of study

	s = 1	s = 2	s = 3	s = 4	s = 5	s = 6	s = 7
Average temperature	−0.006	−0.120	0.004	−0.213	−0.211	−0.123	−0.155
Average humidity	−0.016	−0.219	−0.170	−0.136	−0.015	0.008	−0.204
Average wind speed	0.014	−0.015	−0.145	−0.009	0.006	−0.175	−0.012
Population density	0.224	0.013	0.157	0.026	0.285	0.023	0.343
Age over 65	0.345	0.244	0.005	−0.002	0.010	0.353	0.425
Income	−0.013	−0.146	0.033	−0.200	−0.175	−0.007	0.015
Vaccine	−0.018	−0.153	−0.368	−0.324	−0.291	−0.007	−0.192
Mobility	0.017	0.273	0.327	0.202	0.165	0.012	0.238

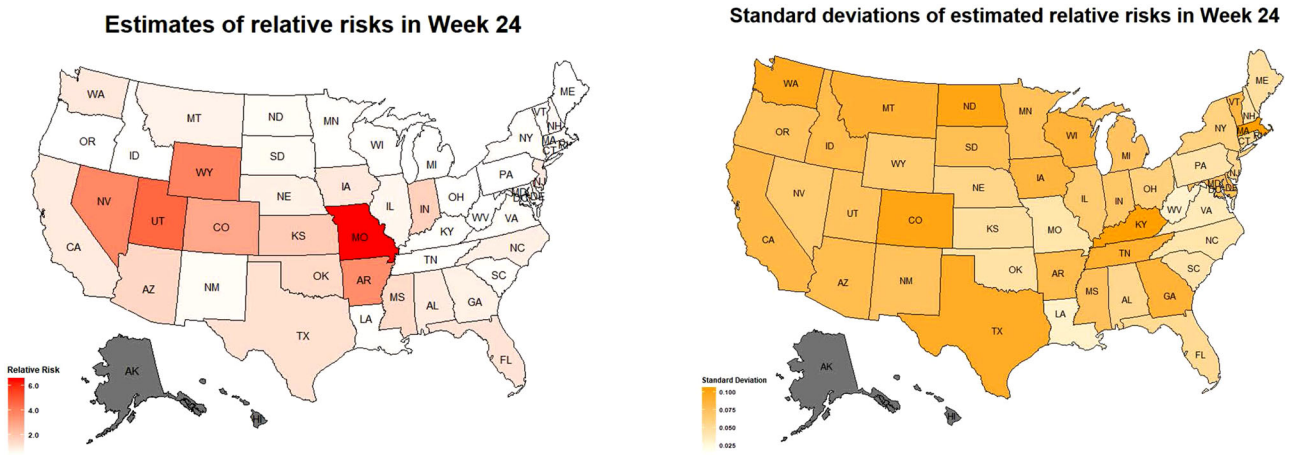

FIGURE 8 Maps of the posterior estimates (left panel) and the corresponding standard deviations (right panel) for the relative risks in Week 24

also detailed in Table S15 of the supplement. Based on the estimation of relative risks, Figure 9 displays the maps of comparison of the observed number of COVID-19 Delta cases and the estimated cases from the proposed model in weeks 24, 28 and 32. The number of cases is counted per 100,000 population. It is shown that the disease cases estimated by the proposed model have recovered the major spatial distributions in the states of interest. Specifically, the Delta variant cases are concentrated in the central regions at the beginning and things are getting worse in the south afterward. Overall, the Delta variant cases are rapidly increasing during the study period. The proposed method correctly captures the evolution of the epidemic across the states of the United States.

4 | DISCUSSION

The paper first illustrates the spatio-temporal distributions of COVID-19 Delta variant cases at the state level in the United States during a 12-week period. The temporal trends and spatial heterogeneity of the data suggest to us to construct a hierarchical model with spatio-temporally varying coefficients and further to cluster the states based on the temporal profiles of the covariate effects. Our findings provide a detailed insight into the significance of the contributing factors on the SMRs of Delta variant. It is shown that the covariate effects are spatially varying, with evidently distinct temporal trends in different clusters.

The study finds dynamic effects of the contributing factors from environmental, socio-economic and demographic aspects as well as public health interventions. Average temperature and average humidity have significant effects in most states of America. Their negative association with the disease morbidity is consistent with the previous findings of Liu et al. (2020) and Mecenas, et al. (2020), where cases sprout up with low temperature and humidity, probably because such condition provides a more suitable environment for Delta variant survival. Average wind speed plays an important role mainly in the north, in agreement with Islam et al. (2021), which also indicates an inverse relationship with COVID-19 cases due to the shorter suspending time and lower concentration of the virus in the case of higher wind speed.

Population density and proportion of elderly people (> 65-years-old) are found positively connected with the risk of infection, congenial with common sense that dense population facilitates the transmission of virus (Poole, 2020) and that elderly people are more susceptible because of weaker immune systems. The effect of personal income is also significant in some states, with high susceptibility to Delta virus for low-income people as a result of deficient self-protection consciousness as well as limited access to health isolation measures (Pequeno et al., 2020). Vaccination promotion and mobility reduction are still efficient prevention actions to protect people from infection, though the implementation of these measures needs to be stepped up in some states.

It can be seen that the significant variables were spatially varying and the covariate effects changed over time, which had potentially influenced the spatial pattern of COVID-19 Delta variate cases. In the early stage of the study period, hotspots were mainly distributed in two regions: Wyoming, Colorado, Utah and Nevada in the west and Missouri and Arkansas in the central south. It is known that there are many mountains in Wyoming, Colorado and Utah, leading to a cold and dry climate, while Nevada is near the Pacific Ocean with cooler temperatures in the spring. Such objective condition is beneficial to the spread of Delta variant in the environment. There are also relatively large proportions of elderly people in Wyoming, Nevada, Missouri and Arkansas (17.1%, 16.1%, 17.3% and 17.4%, respectively). And the community mobility in Wyoming, Missouri and Arkansas was still growing during this period. All these factors contributed to the spread and outbreak of Delta variant in the hotspots in the first few weeks.

In the following weeks, vaccination coverage increased gradually in each state, with significantly higher rate in Washington, Oregon, California, Colorado and New Mexico in the west; so, it is seen that the Delta virus did not continue to spread from the hotspots to other parts of the western region. In addition, due to the significant decline

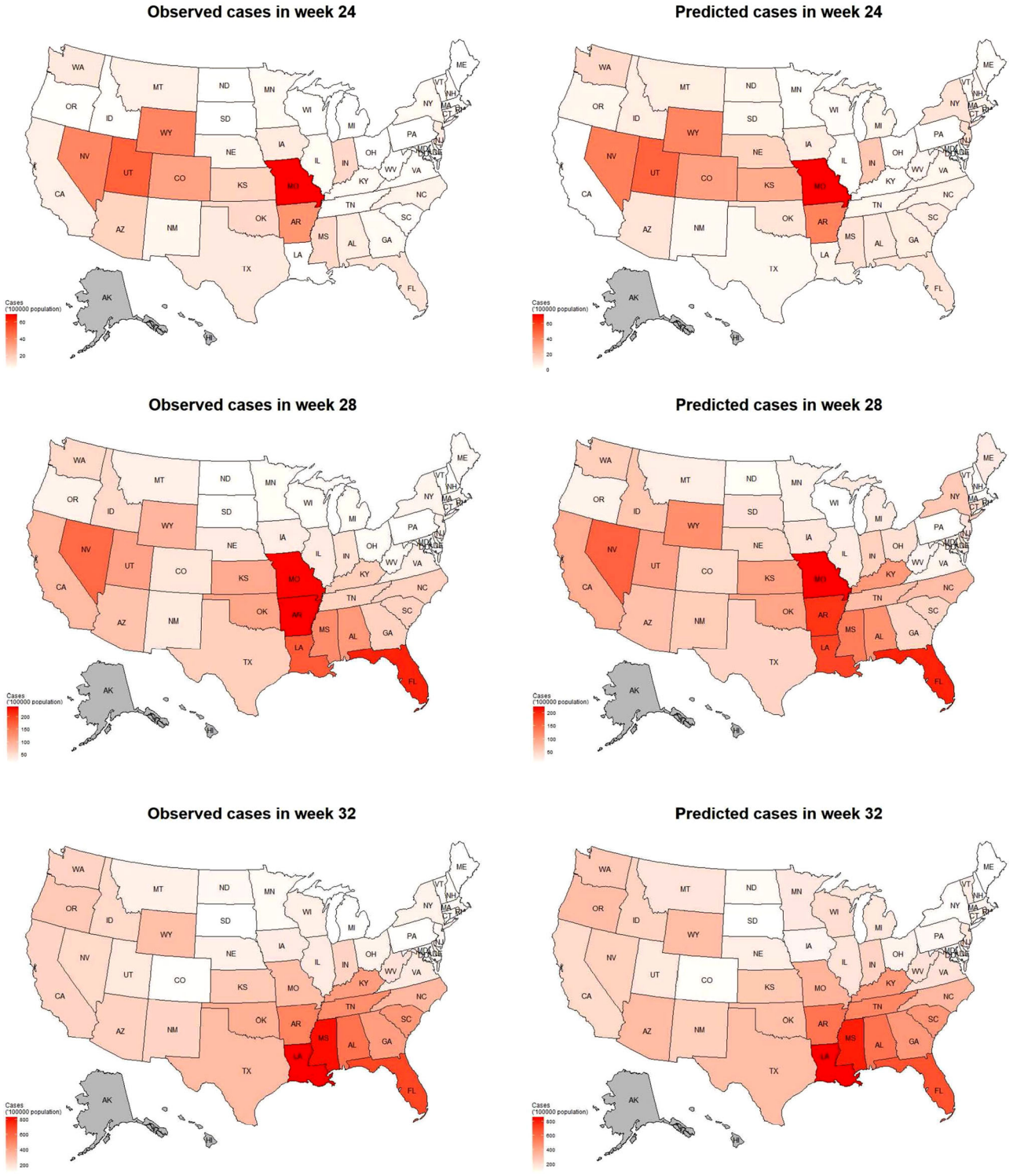

FIGURE 9 Comparison of the observed number of COVID-19 Delta cases and the estimated cases based on the proposed method for 3 weeks within the study period

of mobility in the southwest, including Nevada, Utah and Colorado, the morbidity risk in the western hotspots gradually reduced to the average level.

During this period, there was still relatively higher community mobility but lower vaccination coverage in the southeastern region. Especially in Louisiana, Mississippi and Alabama, which were close to the hotspots, the vaccination coverage was only around 35–40%, providing insufficient protection to the susceptible people. Moreover, these regions also have a large elderly population, facilitating the transmission of the virus to the southeast, which became the new

hotspot in the later stage. Florida, in particular, having a large population density with 20.6% elderly people, developed into the hardest-hit area in the later period, although the mobility there decreased gradually. On the contrary, the states in the northeast maintained a relatively good situation due to the significantly higher vaccination coverage, higher per capita income and reduced community mobility.

Since the beginning of COVID-19 pandemic, a number of major variants, including Alpha, Gamma, Delta and so on, have been identified across the United States. This study focuses on the spread of Delta variant in the states, since it is the most prominent one at the time of writing. Among these lineages, Alpha was the first highly publicized variant and began to spread in the United States by the end of 2020. Based on the data in the website of Cov-spectrum, it is found that from February to March 2021, the incidence of COVID-19 caused by Alpha variant began to increase in the Great Lakes and some southern states such as New Mexico, Texas and Florida. In April, the epidemic situation in these hotspots was further aggravated and the variant diffused to the whole country, gradually becoming the dominant variant in the United States. By May, the spread of Alpha was almost at its peak, with South Dakota and Iowa in the north, as well as Louisiana, Mississippi and Georgia in the south as the worst-hit areas. In June, Alpha gradually receded with the rise of the more invasive Delta variant. On the whole, Delta spread much faster from the central region to the southeast, while Alpha spread from the north and south to the middle.

Carson et al. (2021) showed that Alpha variant was substantially more temperature sensitive, with humidity insignificant in the fitted models. According to Duong (2021) and Zhao et al. (2022), Alpha is 50% more transmissible than the original strain, and thus frequent community mobility and dense population will definitely accelerate the spread of the virus. Many studies (He et al., 2021; Planas et al., 2021) researched the effectiveness of vaccines on the variants and noted that compared with the Alpha strain, Delta is more resistant to antibody, so vaccines appear to be less effective against the Delta variant, but their effects are still significant in the case of high vaccination coverage. Since December 2021, the Omicron variant, which is even more contagious than Delta, has become popular in the United States. Some studies (CDC, 2021; Christensen et al., 2022) have shown that patients infected with Omicron are significantly younger, suggesting that there may be a different age effect, but further research is still needed. Therefore, it is important to maintain SARS-CoV-2 genomic surveillance and build targeted prevention measures for public health.

It should be noted that the number of confirmed cases for COVID-19 Delta variant is subject to some uncertainty, since the detection of Delta variant needs sequencing techniques to scan the viral genome and not every confirmed COVID-19 case is sequenced due to limited laboratory capacity. The paper utilizes the weekly proportion of Delta variant cases relative to all samples sequenced to obtain the approximate counts. Though with some deviations, the counts are the best available data right now and can provide sufficiently reliable basis for analysis.

5 | CONCLUSION

Motivated by the COVID-19 Delta variant cases, this paper develops a flexible Bayesian hierarchical model to identify spatial clusters determined by distinct temporal profiles of coefficients and conduct variable selection for the clusters. It is worth noting that the areas in the same cluster are allowed to be disconnected and that covariate effects vary with space and time. It is shown that for the COVID-19 Delta variant cases, the eight considered environmental, sociodemographic and public intervention factors are selected in different clusters and the factor effects have distinct temporal behaviours depending on the groups. The disease relative risks are well estimated by the proposed method. Based on the results about the associated risk factors, the study could help the government provide early warnings for people at higher risk of infection and develop effective protection suggestions, such as avoiding crowded spaces, keeping well ventilated, taking special care of the elderly and receiving vaccinations timely, which may play an important role in fighting against the Delta variant.

ACKNOWLEDGEMENTS

The work was partially supported by the National Natural Science Foundation of China (No.11861042), and the China Statistical Research Project (No.2020LZ25).

CONFLICT OF INTEREST

The authors declare no conflict of interest with respect to the research, authorship or publication of this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from public sources which have been cited in the references section.

REFERENCES

Adin, A., Lee, D., Goicoa, T., & Ugarte, M. D. (2019). A two-stage approach to estimate spatial and spatio-temporal disease risks in the presence of local discontinuities and clusters. *Statistical Methods in Medical Research*, *28*(9), 2595–2613. https://doi.org/10.1177/0962280218767975

Ahmadi, M., Sharifi, A., Dorosti, S., Ghoushchi, S. J., & Ghanbari, N. (2020). Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Science of the Total Environment*, *729*, 138705. https://doi.org/10.1016/j.scitotenv.2020.138705

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2003). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC. https://doi.org/10.1201/9780203487808

Barbieri, M. M., & Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, *32*(3), 870–897. https://doi.org/10.1214/009053604000000238

Bashir, M. F., Ma, B. J., Bilal, Komal, B., Bashir, M. A., Farooq, T. H., Iqbal, N., & Bashir, M. (2020). Correlation between environmental pollution indicators and COVID-19 pandemic: A brief study in Californian context. *Environmental Research*, *187*, 109652. https://doi.org/10.1016/j.envres.2020.109652

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*, 1–20. https://doi.org/10.1007/BF00116466

Briz-Redón, Á., & Serrano-Aroca, Á. (2020). A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in

Spain. *The Science of the Total Environment*, *728*, 138811. https://doi.org/10.1016/j.scitotenv.2020.138811

Briz-Redón, Á., & Serrano-Aroca, Á. (2022). On the association between COVID-19 vaccination levels and incidence and lethality rates at a regional scale in Spain. *Stochastic Environmental Research and Risk Assessment: Research Journal*, 1–8. https://doi.org/10.1007/s00477-021-02166-y

Carroll, C., Bhattacharjee, S., Chen, Y., Dubey, P., Fan, J., Gajardo, Á., Zhou, X., Müller, H.-G., & Wang, J. -L. (2020). Time dynamics of COVID-19. *Scientific Reports*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-77709-4

Carson, R. T., Carson, S. L., Dye, T. K., Mayfield, S. A., Moyer, D. C., & Yu, C. A. (2021). COVID-19's US temperature response profile. *Environmental and Resource Economics*, *80*(4), 675–704. https://doi.org/10.1007/s10640-021-00603-8

CDC COVID-19 Response Team. (2021). SARS-CoV-2 B.1.1.529 (Omicron) variant – United States, December 1–8, 2021. *Morbidity and Mortality Weekly Report*, *70*(50), 1731–1734. https://doi.org/10.15585/mmwr.mm7050e1

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*(4), 651–673. https://doi.org/10.1214/06-BA122

Choi, J., Lawson, A. B., Cai, B., Hossain, M. M., Kirby, R. S., & Liu, J. (2012). A Bayesian latent model with spatio-temporally varying coefficients in low birth weight incidence data. *Statistical Methods in Medical Research*, *21*(5), 445–456. https://doi.org/10.1177/0962280212446318

Christensen, P. A., Olsen, R. J., Long, S. W., Snehal, R., Davis, J. J., Saavedra, M. O., Reppond, K., Shyer, M. N., Cambric, J., Gadd, R., Thakur, R. M., Batajoo, A., Mangham, R., Pena, S., Trinh, T., Kinskey, J. C., Williams, G., Olson, R., Gollihar, J., & Musser, J. M. (2022). Signals of significantly increased vaccine breakthrough, decreased hospitalization rates, and less severe disease in patients with COVID-19 caused by the Omicron variant of SARS-CoV-2 in Houston, Texas. *The American Journal of Pathology*, *192*, 642–652. In press https://doi.org/10.1016/j.ajpath.2022.01.007

Díaz de León-Martínez, L., de la Sierra-de la Vega, L., Palacios-Ramírez, A., Rodriguez-Aguilar, M., & Flores-Ramírez, R. (2020). Critical review of social, environmental and health risk factors in the Mexican indigenous population and their capacity to respond to the COVID-19. *The Science of the Total Environment*, *733*, 139357. https://doi.org/10.1016/j.scitotenv.2020.139357

Duong, D. (2021). Alpha, Beta, Delta, Gamma: What's important to know about SARS-CoV-2 variants of concern? *CMAJ*, *193*(27), E1059–E1060. https://doi.org/10.1503/cmaj.1095949

Gallagher, M. E., Sieben, A. J., Nelson, K. N., Kraay, A. N. M., Orenstein, W. A., Lopman, B., Handel, A., & Koelle, K. (2021). Indirect benefits are a crucial consideration when evaluating SARS-CoV-2 vaccine candidates. *Nature Medicine*, *27*(1), 4–5. https://doi.org/10.1038/s41591-020-01172-x

Gayawan, E., Awe, O. O., Oseni, B. M., Uzochukwu, I. C., Adekunle, A., Samuel, G., Eisen, D. P., & Adegboye, O. A. (2020). The spatio-temporal epidemic dynamics of COVID-19 outbreak in Africa. *Epidemiology and Infection*, *148*, e212. https://doi.org/10.1017/S0950268820001983

Gelfand, A. E., & Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(3), 501–514. https://doi.org/10.1111/j.2517-6161.1994.tb01996.x

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. *Journal of the American Planning Association*, *86*(4), 495–509. https://doi.org/10.1080/01944363.2020.1777891

He, X., He, C., Hong, W., Zhang, K., & Wei, X. (2021). The challenges of COVID-19 Delta variant: Prevention and vaccine development. *MedComm*, *2*(4), 846–854. https://doi.org/10.1002/mco2.95

Islam, N., Bukhari, Q., Jameel, Y., Shabnam, S., Erzurumluoglu, A. M., Siddique, M. A., Massaro, J. M., & D'Agostino, R. B. Sr. (2021). COVID-19 and climatic factors: A global analysis. *Environmental Research*, *193*, 110355. https://doi.org/10.1016/j.envres.2020.110355

Jamil, T., Alam, I., Gojobori, T., & Duarte, C. M. (2020). No evidence for temperature-dependence of the COVID-19 epidemic. *Frontiers in Public Health*, *8*, 436. https://doi.org/10.3389/fpubh.2020.00436

Lawson, A. B., Choi, J., Cai, B., Hossain, M., Kirby, R. S., & Liu, J. (2012). Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. *Journal of Agricultural, Biological, and Environmental Statistics*, *17*(3), 417–441. https://doi.org/10.1007/s13253-012-0100-3

Lawson, A. B., Choi, J., & Zhang, J. (2014). Prior choice in discrete latent modeling of spatially referenced cancer survival. *Statistical Methods in Medical Research*, *23*(2), 183–200. https://doi.org/10.1177/0962280212447148

Lee, J., Gangnon, R. E., & Zhu, J. (2017). Cluster detection of spatial regression coefficients. *Statistics in Medicine*, *36*(7), 1118–1133. https://doi.org/10.1002/sim.7172

Liu, J., Zhou, J., Yao, J., Zhang, X., Li, L., Xu, X., He, X., Wang, B., Fu, S., Niu, T., Yan, J., Shi, Y., Ren, X., Niu, J., Zhu, W., Li, S., Luo, B., & Zhang, K. (2020). Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. *The Science of the Total Environment*, *726*, 138513. https://doi.org/10.1016/j.scitotenv.2020.138513

Luo, Y., Yan, J., & McClure, S. (2021). Distribution of the environmental and socioeconomic risk factors on COVID-19 death rate across continental USA: A spatial nonlinear analysis. *Environmental Science and Pollution Research International*, *28*(6), 6587–6599. https://doi.org/10.1007/s11356-020-10962-2

Mecenas, P., Bastos, R., Vallinoto, A., & Normando, D. (2020). Effects of temperature and humidity on the spread of COVID-19: A systematic review. *Plos One*, *15*(9), e0238339. https://doi.org/10.1371/journal.pone.0238339

Millett, G. A., Jones, A. T., Benkeser, D., Baral, S., Mercer, L., Beyrer, C., Honermann, B., Lankiewicz, E., Mena, L., Crowley, J. S., Sherwood, J., & Sullivan, P. S. (2020). Assessing differential impacts of COVID-19 on black communities. *Annals of Epidemiology*, *47*, 37–44. https://doi.org/10.1016/j.annepidem.2020.05.003

Ministry of Health. (2021). COVID-19: About the Delta variant. https://www.health.govt.nz/

Mirror. (2021). Delta Covid variant doubles risk of needing hospital treatment https://www.mirror.co.uk/news/uk-news/new-delta-variant-cases-double-25132063

Mlcochova, P., Kemp, S. A., Dhar, M. S., Papa, G., Meng, B., Ferreira, I. A. T. M., Datir, R., Collier, D. A., Albecka, A., Singh, S., Pandey, R., Brown, J., Zhou, J., Goonawardane, N., Mishra, S., Whittaker, C., Mellan, T., Marwal, R., Datta, M., …, & Gupta, R. K. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*, *599*(7883), 114–119. https://doi.org/10.1038/s41586-021-03944-y

Pequeno, P., Mendel, B., Rosa, C., Bosholn, M., Souza, J. L., Baccaro, F., Barbosa, R., & Magnusson, W. (2020). Air transportation, population density and temperature predict the spread of COVID-19 in Brazil. *PeerJ*, *8*, e9322. https://doi.org/10.7717/peerj.9322

Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M. M., Planchais, C., Porrot, F., Robillard, N., Puech, J., Prot, M., Gallais, F., Gantner, P., Velay, A., Le Guen, J., Kassis-Chikhani, N., Edriss, D., Belec, L., Seve, A., …, & Schwartz, O. (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*, *596*(7871), 276–280. https://doi.org/10.1038/s41586-021-03777-9

Poole, L. (2020). Seasonal Influences on the spread of SARS-CoV-2 (COVID19), causality, and forecastabililty. https://ssrn.com/abstract=3554746 or http://doi.org/10.2139/ssrn.3554746

Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S., & Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach.

Sustainable Cities and Society, *62*, 102418. https://doi.org/10.1016/j.scs.2020.102418

Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, O., Lestari, H., Bahar, H., & Asfian, P. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *The Science of the Total Environment*, *725*, 138436. https://doi.org/10.1016/j.scitotenv.2020.138436

World Health Organization. (2020). Coronavirus disease (COVID-19) outbreak situation https://www.who.int/emergencies/diseases/novel-coronavirus-2019

Xie, J., & Zhu, Y. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. *The Science of the Total Environment*, *724*, 138201. https://doi.org/10.1016/j.scitotenv.2020.138201

Xu, H., Yan, C., Fu, Q., Xiao, K., Yu, Y., Han, D., Wang, W., & Cheng, J. (2020). Possible environmental effects on the spread of COVID-19 in China. *Science of the Total Environment*, *731*, 139211. https://doi.org/10.1016/j.scitotenv.2020.139211

Zhao, Y., Huang, J., Zhang, L., Chen, S., Gao, J., & Jiao, H. (2022). The global transmission of new coronavirus variants. *Environmental Research*, *206*, 112240. https://doi.org/10.1016/j.envres.2021.112240

Zheng, R., Xu, Y., Wang, W., Ning, G., & Bi, Y. (2020). Spatial transmission of COVID-19 via public and private transportation in China. *Travel Medicine and Infectious Disease*, *34*, 101626. https://doi.org/10.1016/j.tmaid.2020.101626

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: MA, S., Zhang, X., Wang, K., Zhang, L., Wang, L., Zeng, T., Tang, M.-L., & Tian, M. (2022). Exploring the risk factors of COVID-19 Delta variant in the United States based on Bayesian spatio-temporal analysis. *Transboundary and Emerging Diseases*, 1–14. https://doi.org/10.1111/tbed.14623