Spatiotemporal Associations Between Social Vulnerability, Environmental Measurements, and COVID-19 in the Conterminous United States

Daniel Patrick Johnson[1,1], Niranjan Ravi[1,1], and Christian Braneon[2,2]

[1]Indiana University Purdue University Indianapolis
[2]GISS

December 1, 2022

Abstract

This study introduces the results from fitting a Bayesian hierarchical spatiotemporal model to COVID-19 cases and deaths at the county-level in the United States for the year 2020. Two models were created, one for cases and one for deaths, utilizing a scaled Besag, York, Mollié model with Type I spatial-temporal interaction. Each model accounts for 16 social vulnerability variables and 7 environmental measurements as fixed effects. The spatial structure of COVID-19 infections is heavily focused in the southern U.S. and the states of Indiana, Iowa, and New Mexico. The spatial structure of COVID-19 deaths covers less of the same area but also encompasses a cluster in the Northeast. The spatiotemporal trend of the pandemic in the U.S. illustrates a shift out of many of the major metropolitan areas into the U.S. Southeast and Southwest during the summer months and into the upper Midwest beginning in autumn. Analysis of the major social vulnerability predictors of COVID-19 infection and death found that counties with higher percentages of those not having a high school diploma and having minority status to be significant. Age 65 and over was a significant factor in deaths but not in cases. Among the environmental variables, above ground level (AGL) temperature had the strongest effect on relative risk to both cases and deaths. Hot and cold spots of COVID-19 cases and deaths derived from the convolutional spatial effect show that areas with a high probability of above average relative risk have significantly higher SVI composite scores. Hot and cold spot analysis utilizing the spatiotemporal interaction term exemplifies a more complex relationship between social vulnerability, environmental measurements, and cases/deaths.

Spatiotemporal Associations Between Social Vulnerability, Environmental Measurements, and COVID-19 in the Conterminous United States.

Daniel P. Johnson, Ph.D.*[1], Niranjan Ravi[2], and Christian V. Braneon, Ph.D.[3,4]

*Corresponding Author

[1.] Indiana University - Purdue University at Indianapolis, Department of Geography email: dpjohnso@iu.edu[2.] Indiana University - Purdue University at Indianapolis, Department of Electrical and Computer Engineering[3.] Goddard Institute for Space Studies

[4.] SciSpace, LLC

Key Points

1. Patterns of COVID-19 cases and deaths vary considerably through time and space.
2. COVID-19 cases and deaths concentrated in areas of increased social vulnerability at different times of the year.

3. The relationship between social vulnerability, environmental measurements, and COVID-19 cases and deaths is spatially and temporally variable.

Abstract

This study introduces the results from fitting a Bayesian hierarchical spatiotemporal model to COVID-19 cases and deaths at the county-level in the United States for the year 2020. Two models were created, one for cases and one for deaths, utilizing a scaled Besag, York, Mollié model with Type I spatial-temporal interaction. Each model accounts for 16 social vulnerability variables and 7 environmental measurements as fixed effects. The spatial structure of COVID-19 infections is heavily focused in the southern U.S. and the states of Indiana, Iowa, and New Mexico. The spatial structure of COVID-19 deaths covers less of the same area but also encompasses a cluster in the Northeast. The spatiotemporal trend of the pandemic in the U.S. illustrates a shift out of many of the major metropolitan areas into the U.S. Southeast and Southwest during the summer months and into the upper Midwest beginning in autumn. Analysis of the major social vulnerability predictors of COVID-19 infection and death found that counties with higher percentages of those not having a high school diploma and having non-white status to be significant. Age 65 and over was a significant factor in deaths but not in cases. Among the environmental variables, above ground level (AGL) temperature had the strongest effect on relative risk to both cases and deaths. Hot and cold spots of COVID-19 cases and deaths derived from the convolutional spatial effect show that areas with a high probability of above average relative risk have significantly higher SVI composite scores. Hot and cold spot analysis utilizing the spatiotemporal interaction term exemplifies a more complex relationship between social vulnerability, environmental measurements, and cases/deaths.

Plain Language Summary

COVID-19 affects different locations at different points in time and understanding its impact on communities is an imperative research effort. Communities that are considered socially vulnerable – less resilient to hazards – are disproportionately impacted by pandemics and other environmental stresses. In this study, we utilize a modelling approach that accounts for COVID-19 cases and deaths, social vulnerability, environmental measurements and both space and time domains at the U.S. county level from March 1 – December 31, 2020. Throughout much of the time period, cases were clustered in the U.S. South and the states of Indiana, Iowa and New Mexico. Deaths clustered (although less in extent) in many of these same areas along with the addition of some urbanized counties in the U.S. Northeast. Measurements of social vulnerability were higher in these longer term clusters for both cases and deaths. Examining short-term clusters on a monthly basis, COVID-19 cases and deaths focused more heavily in socially vulnerable areas during the summer and autumn months respectively. The individual social vulnerability variable of not having a high school diploma and non-white status were the most significant contributors to relative risk to both cases and deaths. Age 65 and over contributed significantly to deaths but not to cases. Temperature, with an inverse relationship, had the strongest effect on risk among the environmental measurements. The remaining variables had differing levels of importance in the models. Social vulnerability measures were higher in areas where there was an increased risk of COVID-19 infection and death during the summer and autumn respectively.

Data

Data utilized for the conclusions in this study are available on the Indiana University – Purdue University at Indianapolis Data Repository. https://doi.org/10.7912/D2/23 (Johnson & Ravi, 2021). These data are in CSV format and readily importable into the R statistical package or other platforms.

Index Terms and Keywords

Index Terms: 0240 Public health 0230 Impacts of climate change: human health 0299 General or miscellaneous

Keywords: Spatial epidemiology Social vulnerability COVID-19 pandemic Bayesian spatiotemporal disease mapping Environmental determinants of COVID-19 Remote sensing and COVID-19

Introduction

The Coronavirus Disease 2019 (COVID-19, ICD-10-CM, U07.1, 2019-nCoV acute respiratory disease) pandemic is currently affecting much of the world. As of January 30, 2021, 11 months (325 days) into the pandemic and one year since the WHO declared COVID-19 a Public Health Emergency of International Concern (PHEIC), there are over 100 million confirmed cases of the disease and over 2 million deaths within 223 countries, areas, or territories (WHO 2021). In the United States, as of the same time, there are over 25 million confirmed cases and close to 500,000 deaths; 25.25% of cases and 19.62% of deaths worldwide (U.S. CDC 2021). The U.S. only accounts for 4.23% of the global population, so it is disproportionately affected (*U.S. Census Bureau* , 2021)

Pandemics, as well as other natural and man-made hazards, disproportionately impact socially vulnerable individuals and communities (Freitas & Cidade, 2020; Gaynor & Wilson, 2020; Seddighi, 2020; Usher et al., 2020). The past decade has witnessed an increasing trend in research activity focusing on social and environmental vulnerability as it relates to geophysical and man-made hazards. More recently, there has been vigorous interest in social vulnerability as it relates to the ongoing COVID-19 pandemic (Bilal et al., 2020; Coelho et al., 2020; Dasgupta et al., 2020; Gaynor & Wilson, 2020; Khazanchi et al., 2020; Kim & Bostwick, 2020; Lancet, 2020; Mishra et al., 2020; Mohanty, 2020; Neelon et al., 2020; Snyder & Parks, 2020). Additionally, researchers have attempted to construct COVID-19-specific vulnerability indices, examine spatial relationships or integrate both social and environmental determinants into a complete model, illustrating areas more prone to adverse impacts (Khazanchi et al. 2020; Snyder and Parks 2020; Karaye and Horney 2020).

However, there is a paucity of studies focusing on the spatiotemporal nature of the pandemic and the relationships between social and environmental determinants of COVID-19 vulnerability. This study focuses on a spatiotemporal analysis of the COVID-19 pandemic in the conterminous United States during the year 2020. This investigation not only adds to the growing literature on vulnerability and COVID-19, it also illuminates some of the spatial and temporal underpinnings of the pandemic in the U.S. In order to achieve this, the presented research has three specific aims:

1. Highlight the spatiotemporal associations between social vulnerability, environmental measurements and both cases of and deaths from COVID-19 aggregated by U.S. counties.
2. Model the spatial-temporal dimensions of the pandemic and determine if socially vulnerable counties are more or less impacted at certain times of the year.
3. Create two complementary parsimonious spatiotemporal models - one (1) for COVID-19 cases, and one (1) for COVID-19 deaths - that take into account social vulnerability, environmental measurements, and spatial and temporal random effects.

Background

1.

COVID-19 as a U.S. Health Disparity

The disproportionate impact of COVID-19 on Black, Indigenous and People of Color (BIPOC) in the U.S. is ongoing as of early 2021 (LaVeist 2005; Shi and Stevens 2021). A number of disparities associated with nonwhite status have been examined in the literature. This list includes among others cardiovascular disease, chronic respiratory conditions, hepatitis, and cancer (LaVeist 2005). The fact that COVID-19 is disproportionately represented in U.S. communities of color is not surprising (Singu et al. 2020). The reason(s) for the disparity in representation of COVID-19 cases and deaths within the U.S. population is likely multifaceted

encompassing a variety of cultural, social, environmental and economic contributors. While Persad et al. have noted that "racial identity is not an inherent risk factor", "COVID-19 disparities reflect the health, environmental, and occupational effects of structural racism" (Persad et al., 2020). The popular press and media in the U.S. has highlighted the underfunding of preventative public health infrastructure, an inefficient health-care system, inadequate governmental response, and systemically racist policies that have exacerbated the pandemic's effects. However, one highly probable contributor is the number and extent of socially vulnerable communities within the U.S. that demonstrate reduced resiliency in the face of a hazard (Shi and Stevens 2021; Karaye and Horney 2020; Khazanchi et al. 2020).

Social Vulnerability and COVID-19

Social vulnerability (SV) as a concept refers to a society or communities' impaired ability to respond to an external stressor. These external pressures can be a single incident or the compounding consequences of multiple events leading to deleterious effects on the society or community. Studies highlighting the negative impact COVID-19 has on those considered socially vulnerable have grown exponentially since the beginning of the pandemic. Here we highlight research that focuses on social vulnerability as a covariate in geographic ecological regression studies. Many of these efforts come to similar conclusions; albeit with different variables being more or less related to COVID-19's effects. The studies highlighted utilize the U.S. Centers for Disease Control and Prevention's (U.S. CDC) Social Vulnerability Index (SVI), which we apply in this study (*CDC's Social Vulnerability Index (SVI)* , 2021; Flanagan et al., 2011).

Khazanchi et al., using a quasi-Poisson regression approach, discovered that counties considered vulnerable had a 1.63-fold greater risk for COVID-19 diagnosis and a 1.73-fold greater risk for COVID-19-related death (Khazanchi et al., 2020). When considering only the language and non-white status domain of vulnerability they found a 4.94 and 4.74-fold increase in diagnosis and death respectively. Examining counties broken into the most vulnerable by socioeconomic status, housing and transportation deficiencies resulted in a higher relative risk (i.e. were at a greater risk of COVID-19 infection and death). Further effort by Nayak et al., examining 433 U.S. counties (counties with >= 50 COVID-19 cases as of April 4, 2020), using a generalized linear mixed-effect model, found that higher SV was associated with an increased COVID-19 case fatality rate (CFR) (Aditi Nayak et al. 2020). The relative risk further increased after adjusting for age 65 and over. However, the relationship between the overall SVI score and COVID-19 incidence was not statistically significant. In a study by Neelon et al (2020) utilizing COVID-19 cases and deaths within a Bayesian hierarchical negative binomial model between March 1 and August 31, 2020, counties were classified based on SVI composite percentiles (Neelon et al., 2020). Cases and deaths were examined daily for all U.S. counties after adjusting for percentage rural, percentage poor or in fair health, percentage of adult smokers, county average daily PM2.5 and primary care physicians per 100,000. By March 30, 2020 relative risk became significantly greater than 1.00 in the most vulnerable counties. Upper SVI quartile counties had higher death rates on average beginning on March 30, 2020. By late August the lower quartiles for SVI began to exhibit increasing levels of cases and deaths. Dasgupta et al. (2020), examined COVID-19 cases from June 1 – July 25, 2020 relating them to the CDC's SVI. Areas with a higher proportion of individuals with nonwhite status, housing density, and crowded housing units were more likely to become COVID-19 hot spots; defined as areas where there is a >60% change in the most recent 3-7 day COVID-19 incidence rate (Dasgupta et al., 2020). Among the hot spot counties, those with greater SVI composite scores had higher COVID-19 incidence rates. Karaye and Horney (2020) examined local relationships between COVID-19 case counts and SVI utilizing geographically weighted regression (GWR) (Karaye & Horney, 2020). The study examined data from January 21 through May 12, 2020 and found, (after adjusting for population size, population density, number of persons tested, average daily sunlight, precipitation, air temperature, heat index, and PM2.5) that non-white status, limited English, household composition, transportation, housing and disability effectively predicted case counts in the U.S. Snyder and Parks (2020), in another spatial analysis (utilizing GWR), which did not utilize the CDC SVI, found that socio-ecological vulnerability to COVID-19 varied across the contiguous U.S., with higher levels of vulnerability in the Southeast and low vulnerability in the Upper Midwest, Great Plains and Mountain West (Snyder & Parks, 2020).

Environmental Determinants of COVID-19

Even though we are only 11 months into the pandemic, there is growing evidence regarding environmental determinants of COVID-19 infection and mortality. Initially, it was hypothesized that COVID-19 may behave like many other respiratory infections and cases would subside in the summer months in the Northern Hemisphere due to increases in temperature (Hassan et al., 2020; Jamil et al., 2020; Prata et al., 2020). Therefore, many researchers have concentrated on temperature and less on other meteorological measurements as a factor in spread of SARS-CoV-2. A study conducted in Bangladesh, found that high temperature and high humidity significantly reduce the transmission of COVID-19 when analyzed using ordinary least squares regression (Haque & Rahman, 2020). Another analysis utilizing log-linear generalized additive models across 166 countries, revealed that temperature and relative humidity were also associated with a decrease in COVID-19 cases (Y. Wu et al., 2020). A 1° C increase in temperature was associated with a 3.08% reduction in daily new cases and 1.19% reduction in daily deaths across the studied countries. Relative humidity had a similar effect on cases and deaths. However, this is not surprising given the calculation of relative humidity employs a function of temperature. Rouen et al. , utilizing micro-correlation analysis using a 10-day moving window, found a negative correlation between temperature and outbreak progression (Rouen et al., 2020). Their research was conducted across 4 continents in both hemispheres. Sarkodie and Owusu (2020), using panel estimation techniques, focused their research on the top 20 countries with COVID-19 cases between January 22 and April 27, 2020 and found that high temperature and high relative humidity reduced the transmission of COVID-19 (Sarkodie & Owusu, 2020). However, low temperature, wind speed, dew/frost point, precipitation and surface pressure increased the infectivity of the virus.

At a much finer scale, research in China at 31 different provincial levels revealed a "biphasic" relationship with temperature, using distributed lag nonlinear models (P. Shi et al., 2020). Epidemic intensity was slightly reduced on days following higher temperatures and was associated with a decrease in relative risk. An investigation into temperature and precipitation's relationship with COVID-19 in Oslo, Norway found maximum temperature and average temperature to be positively associate with COVID-19 transmission and precipitation to have a negative relationship, using non-parametric correlation estimaton (Menebo, 2020). Research in the sub-tropical cities of Brazil uncovered a negative relationship between temperature and COVID-19, using generalized additive models and polynomial linear regression (Prata et al., 2020). Bashir et al. (2020) in New York City, New York, USA, between March and April of 2020, found a significant positive correlation between average temperature and minimum temperature on total cases, using Kendall and Spearman rank correlation tests (Bashir et al., 2020). Average temperature was significantly positively related to COVID-19 mortality and the minimum temperature was associated with new cases. Another study, utilizing multivariate regression focusing on all U.S. counties from the beginning of the pandemic to April 14, 2020 found that higher temperatures were associated with a decrease in cases but not deaths (Li et al., 2020). This sample of studies demonstrates – particularly at finer spatial and temporal scales and depending on how temperature is sampled – the relationship between environmental variables and COVID-19 are complex and variable.

Methods

1.

Study Area and Timeframe

This study focuses on the counties (sub-state administrative districts) located in the conterminous United States. We selected this subset due to Alaska and Hawai'i being non-contiguous and the error and complexity these "islands" would introduce into the spatial weighting matrices necessary for the spatiotemporal analysis. Furthermore, we focus on the timeframe from March 1 to December 31, 2020 (10 full months or 307 days); the first calendar year of the pandemic in the United States.

Data Collection

The data described below in steps 3.2.1 – 3.2.4, was used to create the dataset for the analysis (Johnson & Ravi, 2021).

COVID-19 Cases and Deaths

COVID-19 cases and deaths were collected from USAFACTS (*US Coronavirus Daily Cases by County* , 2021; *US Coronavirus Daily Deaths by County* , 2021). These data were retrieved in comma separated values (csv) format and were grouped into monthly cases and deaths for all counties of the contiguous U.S. (n = 3106); two counties were missing data. The expected number of cases and deaths, E , per county area$_i$, was determined by calculating the number of cases and deaths per month and computing the standardized infection rate (SIR) and standardized mortality rate (SMR) for each area for each month; the denominator in the rates $=E_{it}$,. This value is later used as an offset; expected number of cases/deaths at *area* i during*time* t .

Social Vulnerability

In this analysis we utilize the U.S. CDC's Social Vulnerability Index (SVI) (*CDC's Social Vulnerability Index (SVI)* , 2021; Flanagan et al., 2011). The SVI is composed of 18 variables that are related to social vulnerability and the local socio-ecology at the county or census tract-level for the entire U.S. We chose the SVI because it is highly cited in the literature and while there are inherent limitations in all vulnerability indices it demonstrates greater accuracy and relevancy in many studies (Bakkensen et al., 2017; Rufat et al., 2019; Spielman et al., 2020; Tate, 2012). The SVI variables are listed below in Table 1.

<<<Insert Table 1>>>

We utilize the percentage ((variable/total population) * 100) of each variable (except for per capita income where we used U.S. Dollars $) for all the selected counties standardized by their respective z-scores. The SVI also includes 4 themes, based on vulnerability domains, and a composite score of vulnerability. We utilize the composite SVI score for comparison of counties after modeling.

Land Surface Temperature

Daily land surface temperature (LST) measurements were collected from the Moderate Resolution Imaging Spectroradiometer (MODIS) TERRA satellite system; **MOD11a1.006** (Thome, 2020; Wan, Zhengming et al., 2015). MODIS data has a low spatial resolution (1 km) but a high (daily) temporal resolution. This remotely sensed data set, an emissivity corrected land surface temperature image for both daytime and nighttime, was collected from Google Earth Engine using geemap (Gorelick et al., 2017; Q. Wu, 2020). After collection, the daily values were averaged per month for each county in the conterminous U.S. resulting in a monthly average for daytime and nighttime LST. Areas where cloud cover interfered with image acquisition were assigned a NA value. These monthly averaged data were standardized by z-score.

Meteorological Measurements

For additional environmental variables, we utilized the North American Land Data Assimilation System (NLDAS). NLDAS contains land surface model datasets available hourly at 1km spatial resolution and is also accessible in Google Earth Engine (Cosgrove et al., 2003; Gorelick et al., 2017). All but one of the environmental variables listed in Table 2 were averaged by day and then by month for each county and standardized by their respective z-scores. However, for precipitation, we calculated a daily sum and then a monthly average before standardizing. After adjusting for multi-collinearity, the measurements for specific humidity, longwave radiation, shortwave radiation and potential evaporation were removed. Specific humidity is the ratio of the mass of $H_20_{(v)}$ per total mass of the air parcel (kg/kg). This measurement is not a function of temperature and water content like relative humidity, but we still found a greater than 80% correlation across all counties throughout the time period with 2m AGL temperature. The other extraneous environmental variables had correlation coefficients above .80 with 2m AGL temperature.

Modelling and Specification

1.

Bayesian Spatial-Temporal Framework

This study utilizes a Bayesian hierarchical spatiotemporal modeling approach initialized within the freely available R Statistical platform and the R-INLA package (*R: The R Project for Statistical Computing* , 2021; Rue et al., 2009). Furthermore, all models developed in this study were executed within Indiana University's High Performance Computing (HPC) environment (*Research and High Performance Computing* , 2020). Bayesian hierarchical modeling provides a flexible and robust framework where space-time components can be modeled in a straightforward manner. There are numerous introductions to Bayesian disease mapping to which we direct the novice (Best et al., 2005; Blangiardo et al., 2013; A. B. Lawson, 2013; A. Lawson & Lee, 2017; Moraga, 2020).

The Bayesian hierarchical methodology offers many benefits. For example, when creating disease models and relating counts to covariates, it is unreasonable to assume that one can collect all the necessary variables that account for a given response. The approach utilized here, allows for the inclusion of these "unknown" covariates as random effects within the model (Bernardinelli et al., 1995; Best et al., 2005; Congdon, 2019). These effects, in the spatial-temporal domain, account not only for spatial structure (spatial autocorrelation) and noise (overdispersion), but for temporal correlation and interaction between space and time (Besag et al., 1991; N. A. Samat & Pei Zhen, 2017; N.A Samat & Mey, 2017; Ugarte et al., 2014). Also, it is appropriate to utilize the standardized incidence rate (SIR), number of cases/expected number of cases, and standardized mortality rate (SMR), number of deaths/expected number of deaths, for country-level disease modelling. However, when examining disease measurements at a finer level (i.e. county-level or smaller), the SIR/SMR, a surrogate for relative risk, can be unstable and suffer large fluctuations due to some areas possessing a small population relative to the incidence of disease. The Bayesian modelling approach utilized "smooths" values of relative risk through space and time, by "borrowing" information both locally and globally, thereby reducing the impact of these instabilities (Bernardinelli et al., 1995; Besag et al., 1991).

In order to model relative risk, observed counts - in this study, COVID-19 cases of infection and deaths - Y_i are modelled using a Poisson distribution with mean$E_i\ \vartheta_i$; E = expected counts, ϑ_i is the relative risk (RR) of area$i$. The logarithm of $\mathrm{RR_i}$ is the sum of an intercept α and random effects accounting for extra-Poisson variability.

$$Y_i \ \sim \ Poisson\left(E_i\theta_i\right),\ i = 1, \ldots, 3106$$

$$\log\left(\theta_i\right) = \alpha + S_i + U_i$$

α is the overall risk in the study area and S and U are spatial random effects for area i modelling the spatial dependency structure (S) and the unstructured uncorrelated noise (U). Along with the inclusion of covariates, that determine risk (i.e. social vulnerability, environmental measurements), and/or other random effects the overall spatial model can be represented as

$$\log\left(\theta_i\right) = \ d_i\beta + S_i + U_i$$

d_i is a vector consisting of the intercept (α) and β is a coefficient vector; the fixed effects of the model. The parameters of the 27 fixed covariates included in this study are each assigned $\beta_{1:27} \ \sim \ Normal(\mu, \sigma)$ prior distributions.

A widely cited specification for the random spatial effects S and$U$, the Besag, York, Mollié (BYM) model, is regularly utilized in disease mapping studies (Besag et al., 1991). In the BYM model, the spatial random

effect S is assigned a conditional autoregressive (CAR) distribution; smoothing the associated data based on a specified neighborhood structure, where neighbors are defined as areas sharing a common border.

$$S_i|S_{-i} \sim N\left(S_{\delta_i}, \quad \frac{\sigma_S^2}{n_{\delta_i}}\right)$$

$$S_{\delta_i = \, n_{\delta_i}^{-1} \sum_{j \in \delta_i} S_j}$$

$$\delta_i = set\ of\ neighbors$$

$$n_{\delta_i} = number\ of\ neighbors\ of\ area\ i$$

The unstructured component U is modeled as independent and identically distributed (IID) with mean of zero and variance $=\sigma_U^2$. Therefore, data is shared both locally through the$S$ component and globally through the U component.

In this study we follow the parameterization of the Besag, York, Mollié (BYM) model proposed by Simpson at al. (2015) that enables assigning penalized complexity (PC) priors (Simpson et al., 2015). This so-called "BYM2" model, incorporates a scaled spatially structured and unstructured component (S$*$ and U$*$) and is defined as:

$$\log(\theta_i) = \ d_i\beta + \frac{1}{\sqrt{\tau}}(\sqrt{1 - \ \varphi}\ S_* + \ \sqrt{\varphi}U_*)$$

The mixing - between $S\ *$ and$U\ *$ - parameter $\varphi(0 \leq \varphi \leq 1)$ measures the proportion of variance explained by $S\ *$. This scales the BYM2 model making it equal to the spatial model when $\varphi = 1$ and equal to only unstructured spatial noise when $\varphi = 0$ (Riebler et al., 2016). We set priors for these parameters following suggestions by Simpson et al., 2015. PC priors as their name suggest penalize model complexity. In this case they penalize based on the degree with which a given model deviates from a foundational assumption of no spatial dependency ($\varphi = 0$). Conjoining the random spatial effects for each area ($S_* + U\ *$) is termed the convolutional spatial component. The exponential factor for the convolutional spatial effect $e^{(S_* + U_*)}$provides one with RR contribution of the random spatial effects additively. Repeating this procedure for either S$*$ or U$*$ will provide the relative contribution of each and allow the determination of the comparative contribution to variance (spatial fraction). This scaling parameterization makes the BYM2 representation more interpretable between models than the unscaled BYM model.

The above specification for $\log(\vartheta_{\mathrm{i}})$ can be extended into the spatial-temporal domain by the addition of further random effects.

$$\log(\theta_{\mathrm{it}}) = \ d_i\beta + \frac{1}{\sqrt{\tau}}(\sqrt{1 - \ \varphi}\ S_* + \ \sqrt{\varphi}U_*) + \ \gamma_t + \omega_t + \delta_{\mathrm{it}}$$

Here, γ_t & ω_t, correspondingly represent the temporally structured and temporally unstructured random effect. Typically, γ_t, is modeled as a conditional autoregressive random walk of either order one or two (RW1, RW2), but there can be additional specifications (i.e. seasonal). In the present study, we model the temporally structured effect as RW1, and specifyω_t as a Gaussian exchangeable IID ($\omega_t \ \sim \ Normal(0, \frac{1}{\tau_\omega})$). The space-time interaction component δ_{it}, represents a parameter vector that varies jointly through space and time. This vector allows for deviations from the space and time structure that expresses both dynamic spatial changes from one time frame to another and active temporal patterns from one area to another (Knorr-Held, 2000). Therefore, mapping δ_{it}, characterizes short-term clusters of disease activity that deviate from the space-time average over the study area at time t .

Ecological Regression

The covariates representing SVI and environmental measurements (after correcting for multi-collinearity) were included in the models as fixed effects, to examine which measurements are intricately linked to the spatiotemporal processes of the pandemic in the U.S. This study opted to include all variables that logically fit into the framework of vulnerability regardless of their statistical significance provided there is limited issues with multi-collinearity. In addition to accounting for the variables, much of this decision is based on the potentially poor inference generated by utilizing a stepwise framework and the Deviance Information Criteria (DIC) not significantly decreasing when the variables were removed (Greenland et al., 2016; Huberty, 1989). Even though a particular variable might not be statistically significant it is nonetheless important to see its effect in the model and to compare between COVID-19 cases and deaths when an alternative approach, aimed at reducing variables based on their significance, might result in less comparable models. Furthermore, since the response is logarithmic we calculate the exponential of the mean of the β coefficients and subtract from 1.00 to determine each variables effect on relative risk.

Model Selection Criteria

Deviance Information Criteria (DIC), was employed to select the most parsimonious model (Spiegelhalter et al., 2014). During exploratory data analysis, we examined two different prior specifications on the covariates:

$$\beta_{1:27} \sim \ Uniform\,(-\infty, +\infty)$$

$$\beta_{1:27} \sim Normal\,(\mu, \sigma)$$

These resulted in minimal changes to the models and we selected the specification for the covariates which produced the lowest DIC score; in this case the normal prior which produced a DIC score of -10 less than the uniform specification (DIC Cases: Normal 243958.04 vs. Uniform 243969.28 / DIC Deaths: Normal 90824.00 vs. Uniform 90835.76); a somewhat significant reduction (Spiegelhalter et al., 2014). Furthermore, we utilized all 4 types of spatial-temporal interactions suggested by Knorr-Held (2000) and found that Type I best fit our data, temporal stratification, and modeled process. Therefore, in our modeling approach we imposed no restrictions on when or where a space-time anomaly could occur (Type I spatial-temporal interaction).

Disease Mapping

Another key benefit of Bayesian inference is the creation of the posterior distribution where one can generate the probability of exceeding a certain threshold; the so-called exceedance probability. For this purpose, we mapped the convolutional spatial effect,$e^{(S_*+U_*)}$, and the spatially structured effect, $e^{(S_*)}$, at the county level. Also, we plotted the modeled space time interaction, $e^{\delta_{\mathrm{it}}}$, which represents short-term clusters of activity relative to the study-area average at time t . In these maps we followed the classification rule followed by Richardson (2004); areas where$\Pr\left(\theta_{\mathrm{it}} \geq 1\right)\ is\ \geq\ .8$, is considered a hot spot,$\Pr\left(\theta_{\mathrm{it}} \geq 1\right)\ is\ .8\ \geq\ .2$, is considered areas statistically similar to the national average, and$\Pr\left(\theta_{\mathrm{it}} \geq 1\right)\ is\ \leq\ .2$, signifies a cold spot; areas that represent infection/mortality rates below the national average (Richardson et al., 2004). Counties that are considered hot spots through the convolutional spatial effect, the spatially structured effect, and the space time interaction component are compared to the SVI composite score of the combined average and cold spot areas; areas where$\Pr\left(\theta_{\mathrm{it}} \geq 1\right)\ is\ <\ .8$ (non-hot spot areas). This assessment, utilizing notched box-plots, is employed to examine differences in the SVI composite score between the affected areas and in the case of $e^{\delta_{\mathrm{it}}}$, different times.

Limitations and Caveats

A potential limitation of this study - likely not a significant constraint - is the lack of greater temporal resolution with regard to the social vulnerability index. In the case of the CDC's SVI, the index is calculated either yearly or every other year. This study focuses on COVID-19 on a monthly basis and there is no

available capability of measuring changes in social vulnerability at that temporal resolution. That considered, it is likely that many of the individual variables that are used to define social vulnerablity do not change dramatically from one month to the next (Flanagan et al., 2011; Neelon et al., 2020; L. Shi & Stevens, 2021). Also, the internal limitations present in the SVI are extend to our analysis (Bakkensen et al., 2017; Rufat et al., 2019; Tate, 2012)

Another potential limitation is the number of zeros the dataset for COVID-19 fatalities contains; at least initially. We could have opted for a Zero-Inflated Poisson model for deaths but decided to keep the hierarchical specifications and structure as consistent as possible between COVID-19 cases and deaths. By doing so, we eliminate a level of complexity within the model and maintain their comparability.

Finally, we decided not to place temporal lags on the environmental variables in relation to COVID-19 cases and deaths. Estimates for a latency in exposure to COVID-19 and onset of symptoms ranges from 2-24 days (CDC, 2020, p. 19; Grant et al., 2020). The WHO estimates that there is a temporal lag of 2-8 weeks from onset of symptoms to death in the most severe cases (Baud et al., 2020; Woolf et al., 2021). Given these large ranges of temporal associations and the aggregation of the data by month, we opted to compare SVI and environmental measurements on the date where a case or death was reported. Therefore, the coefficients should be interpreted in the proper context and with this consideration in mind.

Results

1.

Temporal Trends in COVID-19 Cases and Deaths

COVID-19 cases by month per 100,000 people is presented in Figure 1A. Cases steadily increased until October, with an exponential increase through October until the end of 2020. Deaths from COVID-19, presented in Figure 1B, increased exponentially between March and April, then decreased and remained fairly stable through October, with another exponential increase in November and December.

<<<Insert Figure 1>>>

Spatiotemporal Ecological Regression

The coefficients resulting from the spatiotemporal ecological regression model for COVID-19 infections are presented in Table 2. The variables grayed out are considered not statistically significant since 0 falls within the 95% credibility interval (Wang et al., 2018). The variables unemployed, age 65 and up, disabled, crowded living, no vehicle, limited English, LST nighttime, AGL temperature, and precipitation have a negative relationship with COVID-19 infection risk. The strongest effect on cases is from the variable "no high school diploma" with a 20.60% increase in risk from a one standard deviation (6.34%) increase. AGL temperature is the second strongest with a 20.70% decrease in risk resulting from an 8.58k (1 standard deviation) increase. The remaining effects of each variable are presented in Table 2.

<<<Insert Table 2>>>

Table 3 contains the coefficient results from the posterior distributions of the ecological regression model for COVID-19 fatalities. The variables with the strongest impact (percent increase) on risk to COVID-19 deaths are non-white status (+40.19%), no high school diploma (35.21%), AGL temperature (-26.68%), age 65 and over (23.93%), and age 17 and under (19.42%) after a standard deviation increase in each variable respectively. Unemployed, single parent, mobile home, group quarters, poverty, uninsured, LST daytime were not statistically significant.

<<<Insert Table 3>>>

Modelled Temporal Trend

Figure 2 exhibits the temporally structured $\gamma_\tau$and unstructured ω_τ effects for both cases and deaths. The left panel shows the modeled structured temporal effect for cases with all covariates, following the random walk order-1 and IID specification for unstructured temporal effects. The structured component shows an increase in relative risk until August with a decrease for the remainder of the year. The unstructured effect tends to fluctuate between being slightly above 1.00 to slightly below 1.00; with its 95% credibility envelope easily encompassing 1.00. In the right panel for deaths, γ_τ, increases until June, drops in July, increases until September, and drops until the end of the study period. Similar to the cases panel, ω_τ , fluctuates between being slightly above 1.00 to slightly below 1.00.

<<<Insert Figure 2>>>

Spatial Effects

The convolutional spatial effect and the spatially structured effect for COVID-19 cases are mapped in Figure 3A. These figures show the probability that the relative risk exceeds 1.00; the national average. There is a strong clustering of high probability for the convolutional spatial effect in Florida, Alabama, Mississippi, Louisiana, Arkansas, Tennessee, Iowa, and Arizona. There is sporadic clustering of high probability most prominent of which is in Indiana, Kansas, and Colorado. There is significant clustering of low probability areas in the Northeast, Pacific Northwest, Upper Atlantic Coast, Upper Midwest, Michigan, and West Virginia. The spatially structured effect for cases follow a similar pattern to the convolutional effect as it explains 82.7% of the variance in the overall spatial effects. Key differences include some higher probabilities in Connecticut and Iowa. Probability is lower in Southern California, Michigan, and New Mexico.

<<<Insert Figure 3>>>

Figure 3B shows exceedance probabilities for the convolutional spatial effect and the spatially structured effect for deaths. There is a strong clustering of high probabilities in New Mexico, Indiana, Louisiana, Eastern Pennsylvania, and the Northeast megalopolis. Higher probabilities are scattered throughout the Southeast and portions of the Midwest into Montana, Idaho and Eastern Oregon. The spatially structured effect accounts for 60.9% of the variance for (S_*+U_*) so similarities are expected, although not as high of a degree as in the spatial effect for cases. The most notable difference is the increase in clustering in the Southeast, the increase in Indiana, and less sporadic dispersion of counties in the Midwest into the Mountain West.

Spatiotemporal Interaction

1.

Cases

Exceedance probabilities using 1.00 as a threshold for the spatiotemporal interaction term are presented in Figure 4, for cases, and Figure 7, for deaths. Initial clustering of high probabilities in March are noted in the Northeast, especially New York, Florida, Louisiana and counties containing some of the major metropolitan areas around the country (i.e., Atlanta, Denver, Detroit, Chicago, Indianapolis, San Diego, Los Angeles, San Francisco, Portland, and Seattle). Lower probabilities are less stable than in April and this is likely due to the average risk being so low at this point in time. In April, the areas noted previously in March have expanded in what appears to be a diffusion pattern; in many cases doubling in extent. Lower probability areas are more prominent and are focused in the Upper Midwest south through the Great Plains into Texas. There is another notable area of low probabilities in Ohio south through West Virginia, west into Kentucky and further south into Tennessee. In May, areas previously noted at high probability have remained fairly stable, with a notable decrease in probability in upper New York, the Upper Northeast and Michigan. There

is a crescent of lower probability extending from Western Pennsylvania, through West Virginia and west to Kansas and Oklahoma. There is a notable increase in cases in Minnesota and upper Iowa.

By June there are some significant changes to areas of high probability. The pandemic seems to have settled much more into the southern U.S. focusing again in Florida, Alabama, Mississippi, Louisiana, eastern Texas, South and North Carolina. Probabilities have decreased in the Northeast and throughout much of the Midwest apart from much of Iowa and southern Minnesota. High probabilities remain in Arizona and have expanded into Utah, Nevada, and much of California. July shows a further solidification of the pandemic in the southern U.S. extending from the Atlantic to the Pacific coast. Arizona northward into Utah and Idaho has joined this high probability area. Metropolitan areas in Minnesota, Wisconsin, Ohio, and Pennsylvania are showing renewed higher probabilities. The Northeast and the middle Midwest are the lowest probability areas, with Illinois and Indiana continuing a trend of decreasing activity. By August, the pandemic is shifting out of the southern U.S. and into the Midwest with increases from Tennessee into Minnesota, North Dakota, and South Dakota. The Mountain West is exhibiting an increase in activity as is much of California. Arizona and New Mexico are showing a decrease in probability and the Northeast remains firmly in the lower category.

September, witnesses the pandemic lessening in the southern U.S., but the increases previously noted in the Upper Midwest have become even more pronounced, with Missouri, Illinois, Iowa, Wisconsin, Minnesota, Kansas, Nebraska, and North and South Dakota heavily burdened. The pandemic continues to lessen in Arizona and California. By October the pandemic continues to rage in the Upper Midwest affecting much of the counties in the states from Wisconsin to Idaho. There is a notable lessening in southern Minnesota and Iowa. The pandemic continues to subside in the southern U.S. and remains stable in the Northeast. November, expresses a further strengthening in the upper Midwest with areas previously showing a lessening pattern overrun by cases. Much of the U.S. is affected apart from the southern U.S., California, Arizona, Washington, and the Northeast. Through December, the pandemic has diminished in the Upper Midwest and shifted with higher probabilities into the Northeast and Texas and has further reasserted itself on the Pacific Coast. The upper Midwest and extreme South continue a waning effect apart from Florida which displays a resurgence.

<<<Insert Figure 4>>>

4.5.2. Deaths

The spatiotemporal interaction term probability exceedances for deaths are shown in Figure 5. As expected there is not much activity in March apart from a deaths in some major cities. By April, deaths begin to show in many of the major metropolitan areas of the U.S. with a clustering of high probabilities in the New York City area, Chicago, Detroit, Indianapolis, Atlanta, San Diego, Los Angeles, San Francisco, Portland, and Seattle. Through May, many of these areas of high probability have expanded in a similar apparent diffusion pattern to cases a month or so earlier. Much of the Northeast megalopolis is affecting along with Detroit, Cleveland, Pittsburg, Chicago, Indianapolis, Nashville, Birmingham, New Orleans, and counties in New Mexico, Arizona, and Southern California. In June, many of the counties around these same cities have become even more heavily burdened, with notable clusters in the Northeast, Ohio, eastern Michigan, northern Illinois, central Indiana, central Mississippi, Arizona and southern California. Through July many of these areas are showing a decrease in deaths apart from the northeast megalopolis, Chicago, central Indiana, Arizona and southern California.

Through August, probabilities of high deaths have shifted to the southern U.S., while lessening in the Northeast and Midwest. The probabilities have strengthened in Arizona and much of California. September witnesses a further strengthening of probabilities in the southern U.S. affecting much of South Carolina, southern Georgia and much of Florida. The pandemic continues to strengthen in Arizona and California. The waning trend has continued throughout much of the upper Northeast and the Ohio Valley. By October, these shifts continue with sporadic higher probabilities throughout the southern U.S. The lessening continues in the Ohio Valley and upper Northeast as Arizona begins a trend of diminishing deaths. November witnesses

a shift in the probability of deaths into the upper Midwest, as suspected considering cases witnessed a similar trend a few months prior. The pandemic begins to lessen in the southern U.S.; but contains some sporadic high probabilities. However, the shift to the north is apparent. Lessening continues in Arizona and California with the same effect stable in the upper Northeast. Through December, the shift in to the upper Midwest is even more evident with the Mountain West now included. December further witnesses a resurgent trend in the upper Northeast, especially northern New York, Vermont, New Hampshire and Maine. The continued lessening in the southern U.S., Arizona and California is noteworthy.

<<<Insert Figure 5>>>

Comparisons of Composite Vulnerability in Hot and Cold Spots

Figure 6A displays the boxplots comparing hot and cold spots for the convolutional spatial effect ($e^{(S_*+U_*)}$) for COVID-19 infections. The light gray distribution is for areas having a probability less than .80 of exceeding 1.00 (cold spots) and the red distribution for areas where the probability is greater than or equal to .80 of exceeding 1.00 (hot spots). Hot spot areas have a significantly higher SVI composite score compared to the low probability areas. Figure 6B illustrates the boxplot comparing areas delineated in the same way to the composite SVI score for COVID-19 fatalities. Likewise, area with a higher probability of COVID-19 deaths have a statistically significant higher SVI composite score. The composite SVI score for the spatially structured effect is similarly higher in areas of higher probability, which is expected based on the percent of variance explained in the convolutional effect by that component.

<<<Insert Figure 6>>>

Taken in the spatiotemporal context, the relationship between higher SVI composite scores and the hot and cold spots are not as straightforward. Figure 7A and 7B displays boxplots comparing the SVI composite score for the areas by month that have a high probability of exceeding 1.00; within the spatiotemporal interaction component$\left(e^{\delta_{it}}\right)$. The individual plots are delineated in the same way as above. Comparing the distributions for SVI composite scores to cases (9A) from April through August the score is higher and statistically significant in hot spot counties. Similarly, for deaths (9B) the SVI score is higher for the counties involved for the months July – October. The mean SVI score for the hot spots during these month nears or exceeds the 3rd quartile value in the cold spot counties. November also witnesses a higher but not statistically significant SVI score interpreted via the notches in the boxplots.

<<<Insert Figure 7>>>

Discussion

1.

Relationships Between Social Vulnerability Variables and COVID-19 Infections and Deaths

Dividing the SVI index into its separate variables and including them in the hierarchical model offered some noteworthy results. Counties that have a high percentage of those without a high school diploma and non-white status have the strongest positive effect on risk; raising risk for COVID-19 infections by 21.44% and 20.60% respectively. Counties with greater percentages of those aged 17 and under are at greater risk; increasing risk by 11.78% for each standard deviation increase (3.43%). Living in a multi-unit dwelling increased risk by 11.18% which is expected due to the density of living spaces. However, living in a home with more than 10 people (crowded living) lowered risk by 5.13%. Additionally, percent unemployed, disabled, no vehicle, limited English, and uninsured negated risk to varying degrees (see Table 2). These findings tend to support other studies that show COVID-19 having a disparate effect in non-white communities (LaVeist 2005; Shi and Stevens 2021; Singu et al. 2020; Karaye and Horney 2020; Khazanchi et al. 2020). They especially support Karaye and Horney (2020) with conclusions related to non-white status, but not as significantly as their findings related to limited English, household composition, transportation, housing and

disability (Karaye & Horney, 2020). This is likely due to their study focusing on cases through May 12, 2020, so the comparison in the amount of data and the timeframe of investigation is different. Dasgupta et al. (2020) found that non-white status and crowded housing were more likely to become COVID-19 hot spots during June and July 2020. Our study supports these findings for non-white status, multi-unit dwelling and group quarters throughout the year 2020.

The SVI index variables and the relationship with death from COVID-19 possesses some dissimilarities compared to cases. Non-white status had the strongest impact on risk of death; increasing relative risk by 40.19% when raised by one standard deviation (19.84%). No high school diploma is again ranked second with a 35.31% increase. Furthermore, age 65 and over demonstrates an increase in relative risk of death by 23.93%; this variable was statistically insignificant in the model for cases. Age 65 and above is a highly recognized individual risk factor for COVID-19 severe disease and mortality (Woolf et al., 2021). Age 17 and under is again relatively high in its impact on risk with a 23.93% increase. Limited English lowered risk by 13.35% (compared to 5.10% for cases) and Multi-unit dwelling increased risk by 11.29%. The result for limited English potentially implies that decreased social connections due to the perceived language barrier may have a preventative effect. The remaining variables are either not statistically significant or offer minimal (less than 10%) impact on deaths. Although there are not as many previous studies that examined deaths as opposed to cases, these findings especially support Khazanchi et al. (2020) where non-white status resulted in a 4.74-fold increase in death.

Relationships Between Environmental Variables and COVID-19 Infections and Deaths

Regarding environmental variables in relation to COVID-19 cases, above ground level (AGL) temperature had the strongest relationship. A one standard deviation increase (8.58K) in temperature results in a 20.70% reduction in relative risk for cases when examined within the context of the entire study. Nighttime LST results in a 11.82% decrease in risk when increased by a single standard deviation (14.61K). Atmospheric pressure produces a 12.34% increase in relative risk for cases when increased by 5465.36 Pa. Precipitation increases lower risk by 4.07% and wind speed (2.84%) and direction (11.79%) have a positive effect on risk of infection. These findings support research which points to increases in temperature lowering the risk of COVID-19 infections (Haque & Rahman, 2020; Prata et al., 2020; Rouen et al., 2020; Sarkodie & Owusu, 2020; P. Shi et al., 2020). Our precipitation finding contradicts Sarkodie and Owusu (2020) where they found a positive association between temperature and COVID-19 cases. However, it does support Menebo (2020) where both variables had a negative association. In relationship to winds, prior research has found a negative association with wind speed and COVID-19 incidence (Islam et al., 2020; Şahin, 2020). However, in our study average monthly wind direction, when the azimuth is increased by 35 degrees, increased risk by 11.79% and average monthly wind speed, when increased by .69m/s, raises risk by 2.84%. While this relationship is difficult to explain it is worth noting and we opted to keep wind data in the analysis to account for its potential effects.

AGL temperature has the strongest effect on risk from death of the environmental variables with a 26.68% decrease in risk for every 8.58K increase. LST nighttime also lowers risk by 11.14% when increased by 14.61K. As atmospheric pressure increases by 5465.36 Pa, it raises the relative risk of death by 14.66%. Precipitation increases by a standard deviation lower risk by 5.62%. Wind direction is significant in the model for deaths. As the azimuth of wind increases by 35° the relative risk lowers by 6.99%. Wind speed is not statistically significant. Again, these findings are for wind 10m above ground level and support the notion that wind may have some impact on COVID-19, although it's difficult to infer. More research is needed on wind's relationship, especially at a finer temporal stratification (i.e. daily) where it should be easier to infer the relationship (Islam et al., 2020; Şahin, 2020).

Temporal Structure of the Pandemic in the United States

Cases and deaths have clearly increased throughout the year 2020 as evidenced by Figure 1. However, the modeled relative risks have fluctuated throughout the time period (Figure 2). These relative risks as modeled through random walk order – 1 show rapid increases in cases and deaths from March through much of the

summer of 2020. Decreases are then evident for the remainder of the year. Temporal relative risk from death shows more fluctuation than cases but also presents a steady decline in average relative risk for the last few months of 2020. On the surface, these results (Figure 1 vs Figure 2) may seem contradictory. Apart from one chart showing per capita COVID-19 cases/deaths and the other modeled relative risk, closer examination of the maps of the spatiotemporal interactions (Figures 4 and 5) show there are more counties affected in the latter months of 2020. This observation suggests the pandemic has broadened in spatial extent, especially in regard to cases during October and November, but has become less intense overall as measured by relative risk. This finding also implies the pandemic may be starting to decline in average intensity – in the U.S. – as we head into 2021. The average non-modelled standardized infection rate (SIR) and standardized mortality rate (SMR) across all counties, support this and demonstrate a similar relative risk trend and corresponding decrease for the latter months of 2020.

Spatial Structure of the Pandemic in the United States

After adjusting for the fixed effects of the covariates and the temporally structured and unstructured random effects, the convolutional spatial effect risk map and the spatially structured effect risk map identified counties at increased risk of COVID-19 infection and death throughout the study period. The most prominent spatial aspect of relative risk to COVID-19 infection were the clusters most heavily focused in the southern U.S., and the states of Indiana, Iowa, and New Mexico. Somewhat supported by Snyder and Parks (2020), with their geographically weighted regression (GWR) approach, this result was not completely unanticipated (Snyder & Parks, 2020). There were additional sporadic areas of increased risk scattered throughout the Midwest, U.S. South and Great Plains. Strong degrees of spatial autocorrelation, which supports the clustering, as modelled with the spatially structured effect, were present in many of those same areas.

Convolutional risk from COVID-19 deaths were less focused than cases but were still prevalent in the U.S. South, especially Louisiana and Tennessee. Indiana and the megalopolis region of the Northeast were also part of this cluster. Interestingly, the latter region did not present as an increased risk area relative to cases throughout the study period. This is a potentially alarming finding suggesting this region has a higher rate of death relative to the number of cases throughout the year. Further west, nearly every county in New Mexico was at an elevated risk. A large degree of spatial autocorrelation (as the spatially structured effect accounts for nearly 61% of the variance between the two components) is also present in many of these identical areas based on the spatially structured risk from death.

Spatiotemporal Structure of the Pandemic in the United States

The spatiotemporal interaction term is a random effect and can be interpreted as the modelled residual risk after accounting for the fixed effects, spatially structured and unstructured and temporally structured and unstructured effects. This represents short-term (month long in our study) sporadic clusters of COVID-19 cases and deaths. During the time period of the study, it is notable that areas impacted by the pandemic shift drastically throughout the country. Cases shift from major metropolitan areas and the Northeast, into the Southern and Southwest U.S. in the summer months. This trend supports evidence found by Snyder and Parks (2020) that the pandemic was focusing in the highly vulnerable U.S. South by mid-summer. By late summer and early autumn, elevated cases have moved into the Upper Midwest with a second shift into the Southwest and a re-emergence in the Northeast by the end of 2020.

The short-term patterns in deaths follow a similar trend but are delayed by approximately a month to month-and-a-half and are not as large in extent or as contiguous. This implies there is a temporal delay from cases to deaths that falls within the WHO suggestion of 2-8 weeks; tending toward the higher end (Baud et al., 2020). Elevated deaths shift into the Southwest by early summer and into the Southern U.S. by late summer. Increased probabilities of above average deaths have moved from the Northeast megalopolis by August as they refocus in the South. By November, deaths have moved into the Upper Midwest and are beginning to shift out of the U.S. South (apart from Florida); into less socially vulnerable areas of the country. December shows the U.S. South to be below average risk along with the Southwest. A further

broadening is evident in the Upper Midwest and Mountain West.

Composite Social Vulnerability in Hot and Cold Spots

The convolutional spatial effect for COVID-19 cases was classified into hot and cold spots based on the criteria of Richardson et al. (2004). For COVID-19 cases of infection, areas that were considered hot spots throughout the course of the study period had higher SVI composite scores (.57) than those that were considered cold spots (.48). In relation to deaths, there was a similar trend with higher SVI composite scores for hot spot areas (.55) versus cold spot areas (.49). These differences were statistically significant at .05 confidence interval. These relationships provide further evidence to studies finding locations that have higher vulnerability scores to be more at-risk of COVID-19 infection and death (Dasgupta et al., 2020; Karaye & Horney, 2020; Khazanchi et al., 2020).

When comparing SVI composite scores to hot and cold spots for cases and deaths based on the spatiotemporal interaction term, the relationship is not as straightforward. In respect to cases, our study supports the findings of Nayak et al., where the relationship between composite SVI score and COVID-19 cases is not statistically significant in the first month of the pandemic in the U.S.(Nayak et al. 2020). The SVI composite score is higher from April to August and again in December in hot spot counties. This observation supports Neelon et al., where counties with higher SVI scores contained higher COVID-19 cases through August. Furthermore, this result verifies Dasgupta et al., where higher SVI scores in June and July 2020 supported the probability of becoming a COVID-19 hotspot. Neelon et al. also discovered that in August, counties with lower composite SVI scores were beginning to be affected. This is also supported by our study, but the composite index is lower in hot spot areas extending into the months after August. For deaths, the index is higher from July through October, with the remaining months either lower or not statistically significant. In the case of August, September and October, the SVI composite score is much higher in hot spots than in cold spots. The mean SVI score in hot spot areas is approaching the 3rd quartile value in cold spot areas.

Much of this observed shift in the relationship between COVID-19 and vulnerability is due to the spatial-temporal nature of the pandemic in the U.S. Cases shift into the southern and Southwest U.S. in the summer months. These areas are known to have significantly larger numbers and percentages of vulnerable than most other areas of the country (L. Shi & Stevens, 2021). Likewise, in the instance of deaths in August, September, and October, they are likewise focused in some of the more vulnerable locations of the American Southeast and Southwest. These results support a more complex or nuanced relationship between temperature and COVID-19 cases and deaths; especially when examined in a smaller spatial and temporal context (Bashir et al., 2020; P. Shi et al., 2020). The overall time period relationships are not stable across all counties for all time periods. By utilizing the spatiotemporal modelling approach used in this study we are able to uncover this more elusive space-time relationship. Alternatively, in areas where the SVI composite score in hot spots is lower than in cold spots, the pandemic has shifted into less socially vulnerable areas of the country; the Upper Midwest and Northeast. However, a key finding is that during the summer months (for cases) and autumn (for deaths) the pandemic seemed to shift into warmer and more socially vulnerable areas of the country. Future spatiotemporal analyses, after the behavior of the pandemic in 2021, are needed to determine if this is likely to be a trend the virus follows or if this is simply how the pandemic initially diffused in the U.S.

Conclusion

Bayesian hierarchical modeling provides a flexible and robust framework in which to model complex spatiotemporal systems. This study presents the findings of fitting a Bayesian hierarchical spatiotemporal model to COVID-19 cases and deaths in the United States for the year 2020. The data collection and modeling framework are explained in detail for straightforward replication. This hopefully will foster continued effort in modeling the spatiotemporal nature of the pandemic in the U.S. and abroad.

A key finding of this study is the spatiotemporal character of the pandemic in the U.S. after accounting for

social vulnerability, environmental measurements and spatial and temporal random effects. In terms of cases, the pandemic shifted into the U.S. South and Southwest during the hotter months of the year, which are the most vulnerable regions of the U.S. Deaths followed the same pattern only with a 1-2-month temporal lag. This demonstrates a potentially alarming trend if this pattern or a similar one is repeated in other years. These highly vulnerable areas are already under significant stress at this time of the year and the introduction of another stressor on a regular basis could be catastrophic in some areas. Further alarming, is the cluster of deaths in the Northeast Megalopolis region.

Studies such as the one presented here provide insight into the complicated mix of social and environmental factors relating to vulnerability. We demonstrate that relationships between COVID-19 cases/deaths, social vulnerability, and environmental measurements are spatially and temporally variable. Even though many of the findings presented here are supportive of other studies, more work is needed in the spatial-temporal domain of the pandemic. One primary effort should be modelling the spatiotemporal structure at a finer temporal scale (i.e. weekly). This could elucidate other relationships with the examined variables and allow for more elaborate spatial temporal interaction specifications (Type IV spatial temporal interaction). Such examinations, perhaps at finer spatial scales (i.e. census tract), could also allow us to focus on some of this study's more alarming results.

Acknowledgments This research was funded by an internal Indiana University grant sponsored by the Indiana University Vice President of Research, IU COVID-19 Rapid Research OVPR and the Indiana Space Grant Consortium.

References

Bakkensen, L. A., Fox-Lent, C., Read, L. K., & Linkov, I. (2017). Validating Resilience and Vulnerability Indices in the Context of Natural Disasters. *Risk Analysis* , *37* (5), 982–1004. https://doi.org/10.1111/risa.12677

Bashir, M. F., Ma, B., Bilal, Komal, B., Bashir, M. A., Tan, D., & Bashir, M. (2020). Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Science of The Total Environment* ,*728* , 138835. https://doi.org/10.1016/j.scitotenv.2020.138835

Baud, D., Qi, X., Nielsen-Saines, K., Musso, D., Pomar, L., & Favre, G. (2020). Real estimates of mortality following COVID-19 infection. *The Lancet Infectious Diseases* , *20* (7), 773. https://doi.org/10.1016/S1473-3099(20)30195-X

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space—Time variation in disease risk. *Statistics in Medicine* , *14* (21–22), 2433–2443. https://doi.org/10.1002/sim.4780142112

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), 1–20.

Best, N., Richardson, S., & Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research* , *14* (1), 35–59.

Bilal, U., Barber, S., Tabb, L., & Diez-Roux, A. V. (2020). Spatial Inequities in COVID-19 Testing, Positivity, Incidence and Mortality in 3 US Cities: A Longitudinal Ecological Study. *MedRxiv* .

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology* , *7* , 39–55. https://doi.org/10.1016/j.sste.2013.07.003

CDC. (2020, February 11). *Coronavirus Disease 2019 (COVID-19)* . Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/index.html

CDC's Social Vulnerability Index (SVI) . (2021, January 19). https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

Coelho, F. C., Lana, R. M., Cruz, O. G., Villela, D. A., Bastos, L. S., Pastore y Piontti, A., Davis, J. T., Vespignani, A., Codeço, C. T., & Gomes, M. F. (2020). Assessing the spread of COVID-19 in Brazil: Mobility, morbidity and social vulnerability. *PloS One* ,*15* (9), e0238214.

Congdon, P. (2019). Spatial heterogeneity in Bayesian disease mapping. *GeoJournal* , *84* (5), 1303–1316.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., & Meng, J. (2003). Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres* , *108* (D22), 2002JD003118. https://doi.org/10.1029/2002JD003118

Dasgupta, S., Bowen, V. B., Leidner, A., Fletcher, K., Rose, C., Cha, A., Kang, G., Dirlikov, E., Pevzner, E., & Rose, D. (2020). Association Between Social Vulnerability and a County's Risk for Becoming a COVID-19 Hotspot—United States, June 1–July 25, 2020. *Morbidity and Mortality Weekly Report* , *69* (42), 1535.

DeCaprio, D., Gartner, J., Burgess, T., Kothari, S., & Sayed, S. (2020). Building a COVID-19 vulnerability index. *ArXiv Preprint ArXiv:2003.07347* .

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management* ,*8* (1). https://doi.org/10.2202/1547-7355.1792

Freitas, C. M. de, & Cidade, N. da C. (2020). COVID-19 AS A GLOBAL DISASTER: Challenges to risk governance and social vulnerability in Brazil. *Ambiente & Sociedade* , *23* .

Gaynor, T. S., & Wilson, M. E. (2020). Social Vulnerability and Equity: The Disproportionate Impact of COVID-19. *Public Administration Review* , *80* (5), 832–838.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* , *202* , 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Grant, M. C., Geoghegan, L., Arbyn, M., Mohammed, Z., McGuinness, L., Clarke, E. L., & Wade, R. G. (2020). The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PLoS ONE* , *15* (6). https://doi.org/10.1371/journal.pone.0234765

Greenland, S., Daniel, R., & Pearce, N. (2016). Outcome modelling strategies in epidemiology: Traditional methods and basic alternatives. *International Journal of Epidemiology* , *45* (2), 565–575. https://doi.org/10.1093/ije/dyw040

Haque, S. E., & Rahman, M. (2020). Association between temperature, humidity, and COVID-19 outbreaks in Bangladesh. *Environmental Science & Policy* , *114* , 253–255. https://doi.org/10.1016/j.envsci.2020.08.012

Hassan, M. M., Zowalaty, M. E. E., Khan, S. A., Islam, A., Nayem, M. R. K., & Jarhult, J. D. (2020). Role of Environmental Temperature on the Attack rate and Case fatality rate of Coronavirus Disease 2019 (COVID-19) Pandemic. *Infection Ecology & Epidemiology* ,*10* (1), 1792620. https://doi.org/10.1080/20008686.2020.1792620

Huberty, C. (1989). Problems with Stepwise Methods—Better Alternatives. *Advances in Social Science Methodology* , *1* , 43–70.

Islam, N., Shabnam, S., & Erzurumluoglu, A. M. (2020). Temperature, humidity, and wind speed are associated with lower Covid-19 incidence. *MedRxiv* , 2020.03.27.20045658. https://doi.org/10.1101/2020.03.27.20045658

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* . https://doi.org/10.3389/fpubh.2020.00436

Johnson, D. P., & Ravi, N. (2021). *Spatiotemporal Associations Between Social Vulnerability, Environmental Measurements, and COVID-19 in the Conterminous United States* [Data set]. IUPUI University Library. https://doi.org/10.7912/D2/23

Karaye, I. M., & Horney, J. A. (2020). The impact of social vulnerability on COVID-19 in the US: An analysis of spatially varying relationships. *American Journal of Preventive Medicine* ,*59* (3), 317–325.

Khazanchi, R., Beiter, E. R., Gondi, S., Beckman, A. L., Bilinski, A., & Ganguli, I. (2020). County-level association of social vulnerability with COVID-19 cases and deaths in the USA. *Journal of General Internal Medicine* , *35* (9), 2784–2787.

Kim, S. J., & Bostwick, W. (2020). Social Vulnerability and Racial Inequality in COVID-19 Deaths in Chicago. *Health Education & Behavior* , 1090198120929677.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* ,*19* (17–18), 2555–2567.

Lancet, T. (2020). Redefining vulnerability in the era of COVID-19.*Lancet (London, England)* , *395* (10230), 1089.

LaVeist, T. A. (2005). *Minority Populations and Health: An Introduction to Health Disparities in the United States* . John Wiley & Sons.

Lawson, A. B. (2013). *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology* . CRC press.

Lawson, A., & Lee, D. (2017). Bayesian disease mapping for public health. In *Handbook of statistics* (Vol. 36, pp. 443–481). Elsevier.

Li, A. Y., Hannah, T. C., Durbin, J. R., Dreher, N., McAuley, F. M., Marayati, N. F., Spiera, Z., Ali, M., Gometz, A., Kostman, J., & Choudhri, T. F. (2020). Multivariate Analysis of Black Race and Environmental Temperature on COVID-19 in the US. *The American Journal of the Medical Sciences* , *360* (4), 348–356. https://doi.org/10.1016/j.amjms.2020.06.015

Menebo, M. M. (2020). Temperature and precipitation associate with Covid-19 new daily cases: A correlation study between weather and Covid-19 pandemic in Oslo, Norway. *Science of The Total Environment* , *737* , 139659. https://doi.org/10.1016/j.scitotenv.2020.139659

Mishra, S. V., Gayen, A., & Haque, S. M. (2020). COVID-19 and urban vulnerability in India. *Habitat International* , *103* , 102230.

Mohanty, S. K. (2020). Contextualising geographical vulnerability to COVID-19 in India. *The Lancet Global Health* , *8* (9), e1104–e1105.

Moraga, P. (2020). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny* . CRC Press. https://www.paulamoraga.com/book-geospatial/

Nayak, A. (2020). *Impact of Social Vulnerability on COVID-19 Incidence and Outcomes in the United States | medRxiv* . https://www.medrxiv.org/content/10.1101/2020.04.10.20060962v2

Neelon, B., Mutiso, F., Mueller, N. T., Pearce, J. L., & Benjamin-Neelon, S. E. (2020). Spatial and temporal trends in social vulnerability and COVID-19 incidence and death rates in the United States. *MedRxiv* .

Persad, G., Peek, M. E., & Emanuel, E. J. (2020). Fairly Prioritizing Groups for Access to COVID-19 Vaccines. *JAMA* , *324* (16), 1601. https://doi.org/10.1001/jama.2020.18513

Prata, D. N., Rodrigues, W., & Bermejo, P. H. (2020). Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil. *Science of The Total Environment* , *729* , 138862. https://doi.org/10.1016/j.scitotenv.2020.138862

R: The R Project for Statistical Computing . (2021). https://www.r-project.org/

Research and high performance computing . (2020). https://kb.iu.edu/d/apes

Richardson, S., Thomson, A., Best, N., & Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives* , *112* (9), 1016–1025.

Riebler, A., Sorbye, S. H., Simpson, D., & Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling.*Statistical Methods in Medical Research* , *25* (4), 1145–1165. https://doi.org/10.1177/0962280216660421

Rouen, A., Adda, J., Roy, O., Rogers, E., & Levy, P. (2020). COVID-19: Relationship between atmospheric temperature and daily new cases growth rate. *Epidemiology & Infection* , *148* . https://doi.org/10.1017/S0950268820001831

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* , *71* (2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rufat, S., Tate, E., Emrich, C. T., & Antolini, F. (2019). How Valid Are Social Vulnerability Models? *Annals of the American Association of Geographers* , *109* (4), 1131–1153. https://doi.org/10.1080/24694452.2018.1535887

Şahin, M. (2020). Impact of weather on COVID-19 pandemic in Turkey.*Science of The Total Environment* , *728* , 138810. https://doi.org/10.1016/j.scitotenv.2020.138810

Samat, N. A., & Pei Zhen, W. (2017). Relative Risk Estimation for Dengue Disease Mapping in Malaysia based on Besag, York and Mollié Model. *PERTANIKA JOURNAL OF SCIENCE AND TECHNOLOGY* , *25* (3), 759–765.

Samat, N.A, & Mey, L. W. (2017). Malaria disease mapping in Malaysia based on Besag-York-Mollie (BYM) Model. *Journal of Physics: Conference Series* , *890* (1), 012167.

Sarkodie, S. A., & Owusu, P. A. (2020). Impact of meteorological factors on COVID-19 pandemic: Evidence from top 20 countries with confirmed cases. *Environmental Research* , *191* , 110101. https://doi.org/10.1016/j.envres.2020.110101

Seddighi, H. (2020). COVID-19 as a natural disaster: Focusing on exposure and vulnerability for response. *Disaster Medicine and Public Health Preparedness* , 1–2.

Shi, L., & Stevens, G. (2021). *Vulnerable Populations in the United States* . John Wiley & Sons.

Shi, P., Dong, Y., Yan, H., Zhao, C., Li, X., Liu, W., He, M., Tang, S., & Xi, S. (2020). Impact of temperature on the dynamics of the COVID-19 outbreak in China. *Science of The Total Environment* , *728* , 138890. https://doi.org/10.1016/j.scitotenv.2020.138890

Simpson, D. P., Rue, H., Martins, T. G., Riebler, A., & Sørbye, S. H. (2015). Penalising model component complexity: A principled, practical approach to constructing priors. *ArXiv:1403.4630 [Stat]* . http://arxiv.org/abs/1403.4630

Singu, S., Acharya, A., Challagundla, K., & Byrareddy, S. N. (2020). Impact of Social Determinants of Health on the Emerging COVID-19 Pandemic in the United States. *Frontiers in Public Health* ,*8* . https://doi.org/10.3389/fpubh.2020.00406

Snyder, B. F., & Parks, V. (2020). Spatial variation in socio-ecological vulnerability to COVID-19 in the contiguous United States. *Health & Place* , *66* , 102471.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* ,*76* (3), 485–493.

Spielman, S. E., Tuccillo, J., Folch, D. C., Schweikert, A., Davies, R., Wood, N., & Tate, E. (2020). Evaluating social vulnerability indicators: Criteria and their application to the Social Vulnerability Index. *Natural Hazards* , *100* (1), 417–436. https://doi.org/10.1007/s11069-019-03820-z

Tate, E. (2012). Social vulnerability indices: A comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards* ,*63* (2), 325–347. https://doi.org/10.1007/s11069-012-0152-2

Thome, K. (2020). *MODIS | Terra* . 610 WebDev. https://terra.nasa.gov/about/terra-instruments/modis

Ugarte, M. D., Adin, A., Goicoa, T., & Militino, A. F. (2014). On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research* ,*23* (6), 507–530. https://doi.org/10.1177/0962280214527528

U.S. Census Bureau . (2021). https://www.census.gov/popclock/

US Coronavirus Daily Cases by County . (2021, March 3). USAFacts.Org. https://usafactsstatic.blob.core.windows.net/public/data/covid-19/covid_confirmed_usafacts.csv

US Coronavirus Daily Deaths by County . (2021, March 3). USAFacts.Org. https://usafactsstatic.blob.core.windows.net/public/data/covid-19/covid_deaths_usafacts.csv

Usher, K., Bhullar, N., Durkin, J., Gyamfi, N., & Jackson, D. (2020). Family violence and COVID-19: Increased vulnerability and reduced options for support. *International Journal of Mental Health Nursing* .

Wan, Zhengming, Hook, Simon, & Hulley, Glynn. (2015). *MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD11A1.006

Wang, X., Yue, R. Y., & Faraway, J. J. (2018). *Bayesian Regression Modeling with INLA* (1st ed.). CRC press. https://www.routledge.com/Bayesian-Regression-Modeling-with-INLA/Wang-Ryan-Faraway/p/book/9780367572266

Woolf, S. H., Chapman, D. A., & Lee, J. H. (2021). COVID-19 as the Leading Cause of Death in the United States. *JAMA* , *325* (2), 123–124. https://doi.org/10.1001/jama.2020.24865

Wu, Q. (2020). geemap: A Python package for interactive mapping with Google Earth Engine. *Journal of Open Source Software* ,*5* (51), 2305. https://doi.org/10.21105/joss.02305

Wu, Y., Jing, W., Liu, J., Ma, Q., Yuan, J., Wang, Y., Du, M., & Liu, M. (2020). Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Science of The Total Environment* , *729* , 139051. https://doi.org/10.1016/j.scitotenv.2020.139051

Variable	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Unemployed	5.74%	2.78%	0%	26.40%
Per Capita Income	$27,036.00	$6457.00	$0	$72,832.00
No High School Diploma	13.45%	6.34%	1.20%	66.30%
Age 65+	18.43%	4.54%	3.80%	55.60%
Age17-	22.35%	3.43%	7.30%	40.30%
Disabled	15.96%	4.40%	3.80%	33.70%
Single Parent	8.30%	2.71%	0%	25.60%
Non-White Status	23.15%	19.84%	0%	99.30%

Variable	Mean	Standard Deviation	Minimum	Maximum
Mobile Home	13.03%	9.62%	0%	59.30%
Crowded Living(>10 per hh)	2.33%	1.92%	0%	33.80%
No Vehicle	6.18%	3.57%	0%	77.00%
Group Quarters	3.48%	4.74%	0%	55.70%
Poverty	15.63%	6.46%	0%	55.10%
Multi-Unit Dwelling	4.63%	5.64%	0%	89.40%
Limited English	1.70%	2.79%	0%	30.40%
Uninsured	10.00%	4.98%	1.70%	42.40%
Daytime LST*	294.65 K	16.62 K	239.00 K	334.00 K
Nighttime LST*	292.73 K	14.61 K	240.21 K	326.94 K
(AGL) Temperature ** 2 m Above Ground Level	289.11 K	8.58 K	262.81 K	309.40 K
Specific Humidity**	.008 kg/kg	.004 kg/kg	.0015 kg/kg	.020 kg/kg
Atmospheric Pressure**	96628.92 Pa	5465.36 Pa	6669.97 Pa	102268.70 Pa
Longwave Radiation**	327.85 W/m^2	51.84 W/m^2	170.07 W/m^2	438.38 W/m^2
Shortwave Radiation**	208.13 W/m^2	72.30 W/m^2	38.76 W/m^2	359.35 W/m^2
Potential Evaporation**	.22 kg/m^2	.10 kg/m^2	-.0051 kg/m^2	.55 kg/m^2
Precipitation**	.11 kg/m^2	.008 kg/m^2	0 kg/m^2	.66 kg/m^2
Wind Speed 10 m AGL**	3.45 m/s	.69 m/s	1.08 m/s	6.93 m/s
Wind Direction 10 m AGL**	190.52°	35.00°	77.38°	335.61°

Table 1: Descriptive statistics of monthly (year 2020) averages for SVI and environmental variables utilized in the models and aggregated by U.S. county * MODIS ** NLDAS.

Variable	Mean ?	Standard Deviation ?	2.5% Credibility Interval	97.5% Credibility In
Intercept	-0.65	0.01	-0.66	-0.63
Unemployed	-0.09	0.04	-0.17	-0.01
Per Capita Income	0.04	0.01	0.01	0.07
No High School Diploma	0.19	0.02	0.16	0.23
Age 65+	-0.02	0.01	-0.05	0
Age17-	0.11	0.01	0.08	0.14
Disabled	-0.07	0.01	-0.1	-0.04
Single Parent	0.03	0.01	0.01	0.05
Minority	0.19	0.01	0.16	0.21
Mobile Home	-0.01	0.01	-0.04	0.01
Crowded Living(>10 per hh)	-0.05	0.01	-0.07	-0.03
No Vehicle	-0.07	0.01	-0.1	-0.05

Variable	Mean ?	Standard Deviation ?	2.5% Credibility Interval	97.5% Credibility In
Group Quarters	0.03	0.01	0.01	0.05
Poverty	0.09	0.04	0.01	0.18
Multi-Unit Dwelling	0.11	0.01	0.08	0.13
Limited English	-0.05	0.01	-0.08	-0.03
Uninsured	-0.01	0.01	-0.03	0.01
LST Day	0	0.01	-0.02	0.02
LST Night	-0.13	0.01	-0.14	-0.11
AGL Temperature	-0.23	0.02	-0.27	-0.2
Pressure	0.12	0.01	0.09	0.14
Precipitation	-0.04	0.01	-0.06	-0.03
Wind Direction	0.11	0.01	0.09	0.13
Wind Speed	0.03	0.01	0.01	0.04

Table 2: Posterior coefficients and 95% credibility intervals of variables utilized in the model for COVID-19 cases of infection. Grayed out variables are not statistically significant.

Variable	Mean ?	Standard Deviation ?	2.5% Credibility Interval	97.5% Credibility In
Intercept	-1.88	0.05	-1.99	-1.79
Unemployed	-0.06	0.07	-0.2	0.08
Per Capita Income	0.06	0.02	0.02	0.11
No High School Diploma	0.3	0.03	0.25	0.36
Age 65+	0.21	0.02	0.17	0.26
Age17-	0.18	0.03	0.13	0.23
Disabled	-0.09	0.02	-0.13	-0.04
Single Parent	0.07	0.02	0.03	0.11
Minority	0.34	0.02	0.3	0.38
Mobile Home	0.01	0.02	-0.03	0.05
Crowded Living(>10 per hh)	-0.09	0.02	-0.13	-0.05
No Vehicle	-0.02	0.02	-0.05	0.01
Group Quarters	0.03	0.02	-0.01	0.06
Poverty	0.07	0.07	-0.08	0.21
Multi-Unit Dwelling	0.11	0.02	0.07	0.15
Limited English	-0.14	0.02	-0.19	-0.1
Uninsured	-0.01	0.02	-0.05	0.03
LST Day	0	0.02	-0.03	0.04
LST Night	-0.12	0.01	-0.15	-0.09
AGL Temperature	-0.31	0.03	-0.37	-0.25
Pressure	0.14	0.02	0.1	0.17
Precipitation	-0.06	0.01	-0.08	-0.03
Wind Direction	-0.07	0.01	-0.1	-0.04
Wind Speed	-0.02	0.01	-0.04	0.01

Table 3: Posterior coefficients and 95% credibility intervals of variables utilized in the model for COVID-19 deaths. Grayed out variables are not statistically significant.

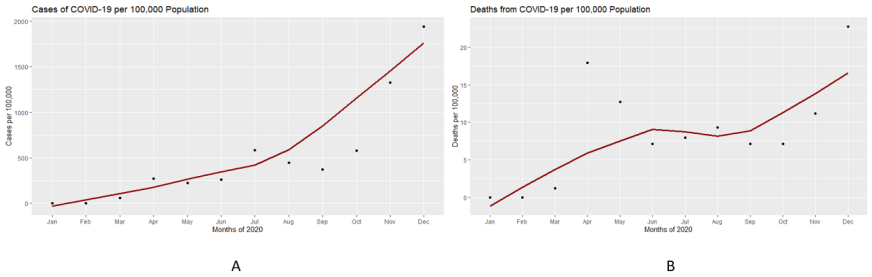


Figure 1: A. Cases of COVID-19 per 100,000 count of population in the United States for the year 2020. (Crimson line = Locally weighted scatterplot smoothing (LOWESS) line of cases). B. Deaths from COVID-19 per 100,000 count of population in the United States for the year 2020. (Crimson line = Locally weighted scatterplot smoothing (LOWESS) line of deaths)

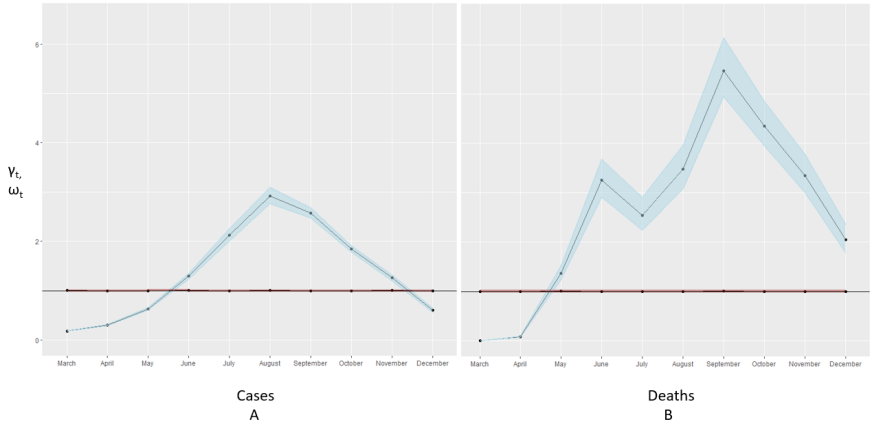
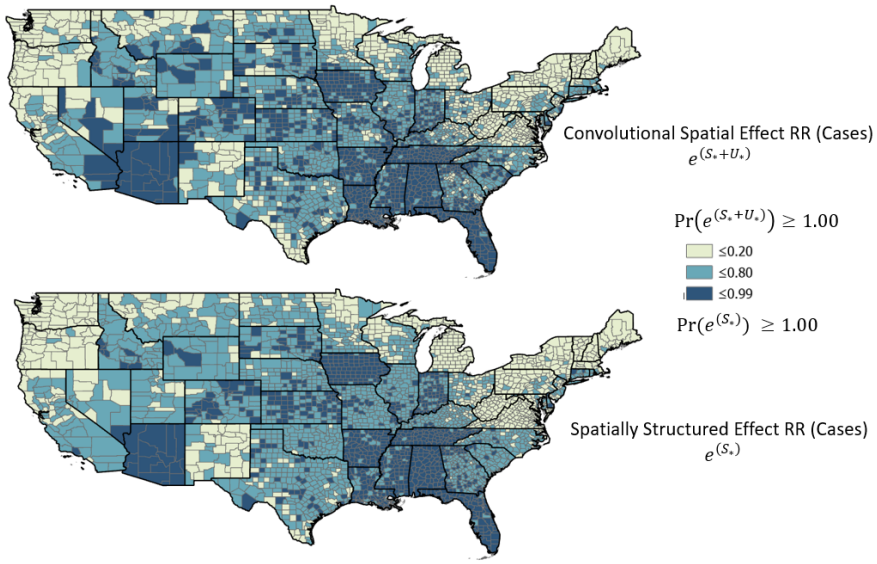
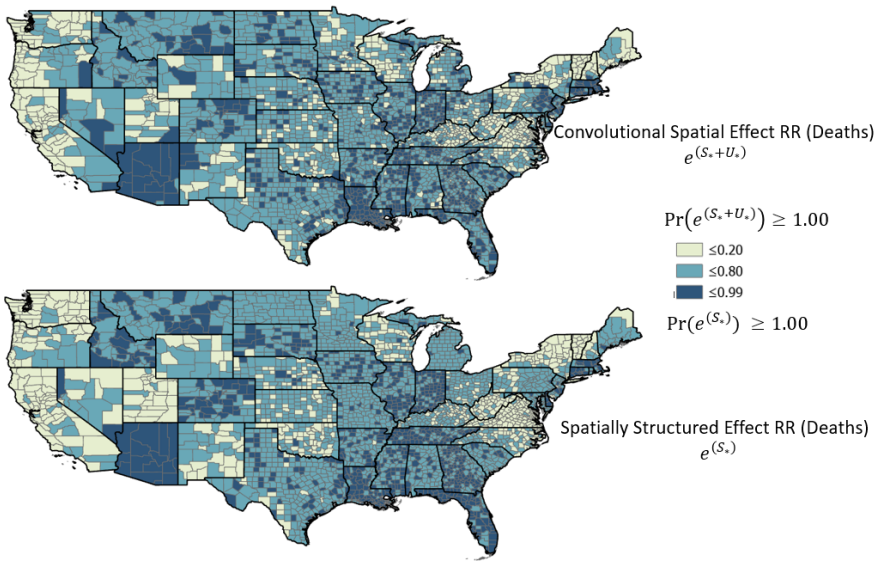


Figure 2: Posterior temporally structured; γ_τ(light blue) and unstructured; ω_τ (light red) effects for COVID-19 cases; panel A, and deaths; panel B (95% credibility envelope) in the U.S. by month for the year 2020.



A



B

Figure 3: Exceedance probabilities of convolutional spatial effect $(e^{(S_*+U_*)})$ and spatially structured effect $(e^{(S_*)})$ relative risk associated with COVID-19 infection; panel A, and COVID-19 mortality; panel B.

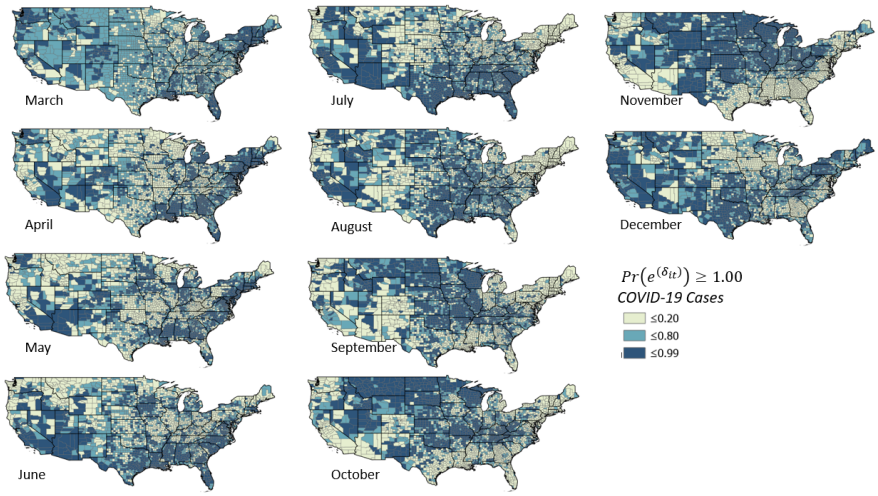


Figure 4: Probabilities of space-time interaction termPr $\left(e^{\delta_{\text{it}}}\right)$ exceeding 1.00 during the year 2020 for U.S. COVID-19 infections stratified monthly.

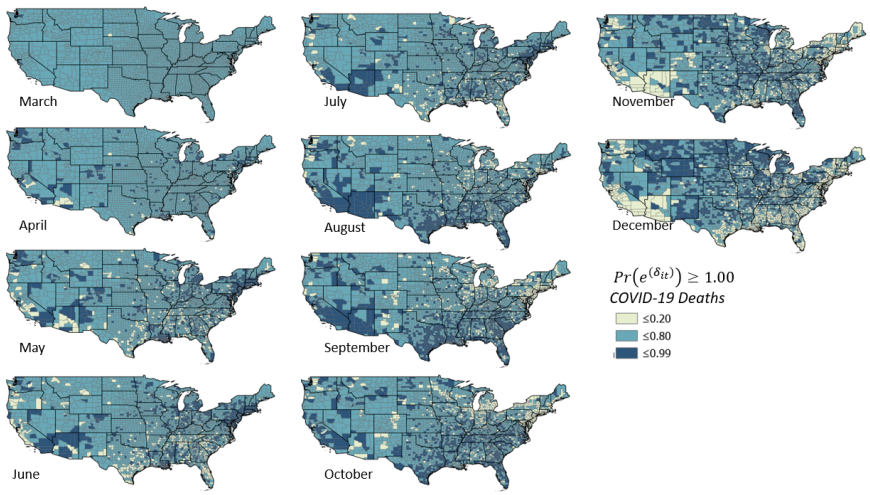
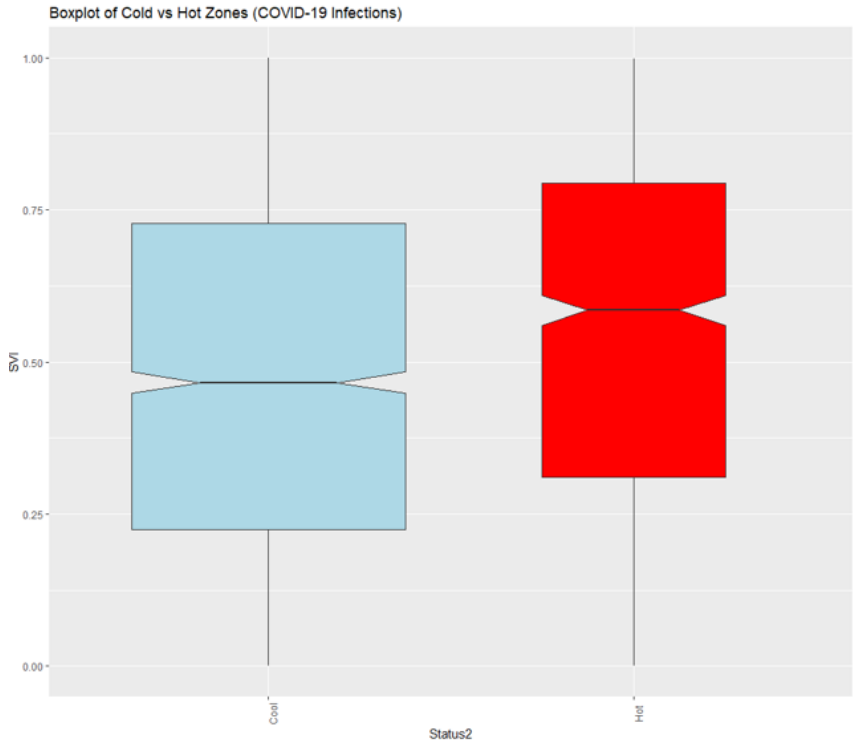
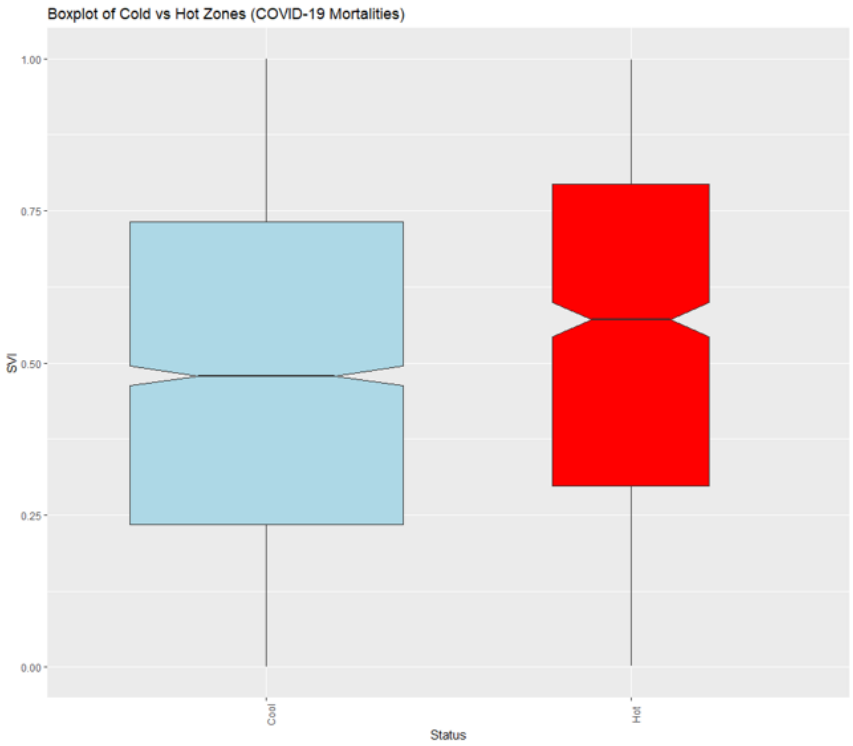


Figure 5: Probabilities of space-time interaction termPr $\left(e^{\delta_{\mathrm{it}}}\right)$ exceeding 1.00 during the year 2020 for U.S. COVID-19 mortalities stratified monthly.



_____A_____



---------- *B* ----------

Figure 6: Boxplot of SVI composite score for hot and cold spots for the convolutional spatial effect for COVID-19 infections; plot A, COVID-19 fatalities; plot B.

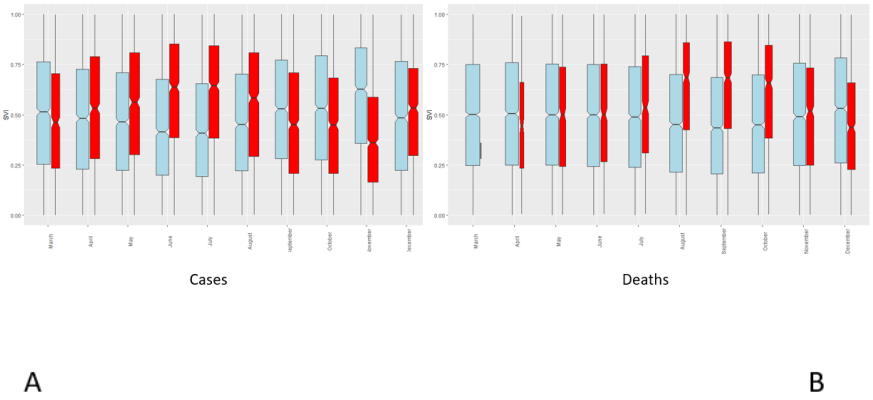


Figure 7: Boxplots organized by month, comparing cold spots and hot spots of COVID-19 cases (A) and deaths (B) to composite SVI score.

Spatiotemporal Associations Between Social Vulnerability, Environmental Measurements, and COVID-19 in the Conterminous United States.

Daniel P. Johnson, Ph.D.*[1], Niranjan Ravi[2], and Christian V. Braneon, Ph.D.[3,4]

*Corresponding Author

[1.] Indiana University - Purdue University at Indianapolis, Department of Geography
email: dpjohnso@iu.edu

[2.] Indiana University - Purdue University at Indianapolis, Department of Electrical and Computer Engineering

[3.] Goddard Institute for Space Studies

[4.] SciSpace, LLC

Key Points

1. Patterns of COVID-19 cases and deaths vary considerably through time and space.
2. COVID-19 cases and deaths concentrated in areas of increased social vulnerability at different times of the year.
3. The relationship between social vulnerability, environmental measurements, and COVID-19 cases and deaths is spatially and temporally variable.

Abstract

This study introduces the results from fitting a Bayesian hierarchical spatiotemporal model to COVID-19 cases and deaths at the county-level in the United States for the year 2020. Two models were created, one for cases and one for deaths, utilizing a scaled Besag, York, Mollié model with Type I spatial-temporal interaction. Each model accounts for 16 social vulnerability variables and 7 environmental measurements as fixed effects. The spatial structure of COVID-19 infections is heavily focused in the southern U.S. and the states of Indiana, Iowa, and New Mexico. The spatial structure of COVID-19 deaths covers less of the same area but also encompasses a cluster in the Northeast. The spatiotemporal trend of the pandemic in the U.S. illustrates a shift out of many of the major metropolitan areas into the U.S. Southeast and Southwest during the summer months and into the upper Midwest beginning in autumn. Analysis of the major social vulnerability predictors of COVID-19 infection and death found that counties with higher percentages of those not having a high school diploma and having non-white status to be significant. Age 65 and over was a significant factor in deaths but not in cases. Among the environmental variables, above ground level (AGL) temperature had the strongest effect on relative risk to both cases and deaths. Hot and cold spots of COVID-19 cases and deaths derived from the convolutional spatial effect show that areas with a high probability of above average relative risk have significantly higher SVI composite scores. Hot and cold spot analysis utilizing the spatiotemporal interaction term exemplifies a more complex relationship between social vulnerability, environmental measurements, and cases/deaths.

Plain Language Summary

COVID-19 affects different locations at different points in time and understanding its impact on communities is an imperative research effort. Communities that are considered socially vulnerable – less resilient to hazards – are disproportionately impacted by pandemics and other environmental stresses. In this study, we utilize a modelling approach that accounts for COVID-19 cases and deaths, social vulnerability, environmental measurements and both space and time domains at the U.S. county level

from March 1 – December 31, 2020. Throughout much of the time period, cases were clustered in the U.S. South and the states of Indiana, Iowa and New Mexico. Deaths clustered (although less in extent) in many of these same areas along with the addition of some urbanized counties in the U.S. Northeast. Measurements of social vulnerability were higher in these longer term clusters for both cases and deaths. Examining short-term clusters on a monthly basis, COVID-19 cases and deaths focused more heavily in socially vulnerable areas during the summer and autumn months respectively. The individual social vulnerability variable of not having a high school diploma and non-white status were the most significant contributors to relative risk to both cases and deaths. Age 65 and over contributed significantly to deaths but not to cases. Temperature, with an inverse relationship, had the strongest effect on risk among the environmental measurements. The remaining variables had differing levels of importance in the models. Social vulnerability measures were higher in areas where there was an increased risk of COVID-19 infection and death during the summer and autumn respectively.

Data

Data utilized for the conclusions in this study are available on the Indiana University – Purdue University at Indianapolis Data Repository. https://doi.org/10.7912/D2/23 (Johnson & Ravi, 2021). These data are in CSV format and readily importable into the R statistical package or other platforms.

Index Terms and Keywords

Index Terms: 0240 Public health
0230 Impacts of climate change: human health
0299 General or miscellaneous

Keywords: Spatial epidemiology
Social vulnerability
COVID-19 pandemic
Bayesian spatiotemporal disease mapping
Environmental determinants of COVID-19
Remote sensing and COVID-19

1. Introduction

The Coronavirus Disease 2019 (COVID-19, ICD-10-CM, U07.1, 2019-nCoV acute respiratory disease) pandemic is currently affecting much of the world. As of January 30, 2021, 11 months (325 days) into the pandemic and one year since the WHO declared COVID-19 a Public Health Emergency of International Concern (PHEIC), there are over 100 million confirmed cases of the disease and over 2 million deaths within 223 countries, areas, or territories (WHO 2021). In the United States, as of the same time, there are over 25 million confirmed cases and close to 500,000 deaths; 25.25% of cases and 19.62% of deaths worldwide (U.S. CDC 2021). The U.S. only accounts for 4.23% of the global population, so it is disproportionately affected (*U.S. Census Bureau*, 2021)

Pandemics, as well as other natural and man-made hazards, disproportionately impact socially vulnerable individuals and communities (Freitas & Cidade, 2020; Gaynor & Wilson, 2020; Seddighi, 2020; Usher et al., 2020). The past decade has witnessed an increasing trend in research activity focusing on social and environmental vulnerability as it relates to geophysical and man-made hazards. More recently, there has been vigorous interest in social vulnerability as it relates to the ongoing COVID-19 pandemic (Bilal et al., 2020; Coelho et al., 2020; Dasgupta et al., 2020; Gaynor & Wilson, 2020; Khazanchi et al., 2020; Kim & Bostwick, 2020; Lancet, 2020; Mishra et al., 2020; Mohanty, 2020; Neelon et al., 2020; Snyder & Parks, 2020). Additionally, researchers have attempted to construct COVID-19-specific vulnerability indices, examine spatial relationships or integrate both social and environmental determinants into a complete model, illustrating areas more prone to adverse impacts (Khazanchi et al. 2020; Snyder and Parks 2020; Karaye and Horney 2020).

However, there is a paucity of studies focusing on the spatiotemporal nature of the pandemic and the relationships between social and environmental determinants of COVID-19 vulnerability. This study focuses on a spatiotemporal analysis of the COVID-19 pandemic in the conterminous United States during the year 2020. This investigation not only adds to the growing literature on vulnerability and COVID-19, it also illuminates some of the spatial and temporal underpinnings of the pandemic in the U.S. In order to achieve this, the presented research has three specific aims:

1. Highlight the spatiotemporal associations between social vulnerability, environmental measurements and both cases of and deaths from COVID-19 aggregated by U.S. counties.
2. Model the spatial-temporal dimensions of the pandemic and determine if socially vulnerable counties are more or less impacted at certain times of the year.

3. Create two complementary parsimonious spatiotemporal models - one (1) for COVID-19 cases, and one (1) for COVID-19 deaths - that take into account social vulnerability, environmental measurements, and spatial and temporal random effects.

2. Background

2.1. COVID-19 as a U.S. Health Disparity

The disproportionate impact of COVID-19 on Black, Indigenous and People of Color (BIPOC) in the U.S. is ongoing as of early 2021 (LaVeist 2005; Shi and Stevens 2021). A number of disparities associated with nonwhite status have been examined in the literature. This list includes among others cardiovascular disease, chronic respiratory conditions, hepatitis, and cancer (LaVeist 2005). The fact that COVID-19 is disproportionately represented in U.S. communities of color is not surprising (Singu et al. 2020). The reason(s) for the disparity in representation of COVID-19 cases and deaths within the U.S. population is likely multifaceted encompassing a variety of cultural, social, environmental and economic contributors. While Persad et al. have noted that "racial identity is not an inherent risk factor", "COVID-19 disparities reflect the health, environmental, and occupational effects of structural racism" (Persad et al., 2020). The popular press and media in the U.S. has highlighted the underfunding of preventative public health infrastructure, an inefficient health-care system, inadequate governmental response, and systemically racist policies that have exacerbated the pandemic's effects. However, one highly probable contributor is the number and extent of socially vulnerable communities within the U.S. that demonstrate reduced resiliency in the face of a hazard (Shi and Stevens 2021; Karaye and Horney 2020; Khazanchi et al. 2020).

2.2. Social Vulnerability and COVID-19

Social vulnerability (SV) as a concept refers to a society or communities' impaired ability to respond to an external stressor. These external pressures can be a single incident or the compounding consequences of multiple events leading to deleterious effects on the society or community. Studies highlighting the negative impact COVID-19 has on those considered socially vulnerable have grown exponentially since the beginning of the pandemic. Here we highlight research that focuses on social vulnerability as a covariate in geographic ecological regression studies. Many of these efforts come to similar conclusions; albeit with different variables being more or less related to COVID-19's effects. The studies highlighted utilize the U.S. Centers for Disease Control and Prevention's (U.S. CDC) Social Vulnerability Index (SVI), which we apply in this study (*CDC's Social Vulnerability Index (SVI)*, 2021; Flanagan et al., 2011).

Khazanchi et al., using a quasi-Poisson regression approach, discovered that counties considered vulnerable had a 1.63-fold greater risk for COVID-19 diagnosis and a 1.73-fold greater risk for COVID-19-related death (Khazanchi et al., 2020). When considering only the language and non-white status domain of vulnerability they found a 4.94 and 4.74-fold increase in diagnosis and death respectively. Examining counties broken into the most vulnerable by socioeconomic status, housing and transportation deficiencies resulted in a higher relative risk (i.e. were at a greater risk of COVID-19 infection and death). Further effort by Nayak et al., examining 433 U.S. counties (counties with >= 50 COVID-19 cases as of April 4, 2020), using a generalized linear mixed-effect model, found that higher SV was associated with an increased COVID-19 case fatality rate (CFR) (Aditi Nayak et al. 2020). The relative risk further increased after adjusting for age 65 and over. However, the relationship between the overall SVI score and COVID-19 incidence was not statistically significant. In a study by Neelon et al (2020) utilizing COVID-19 cases and deaths within a Bayesian hierarchical negative binomial model between March 1 and August 31, 2020, counties were classified based on SVI composite percentiles (Neelon et al., 2020). Cases and deaths were examined daily for all U.S. counties after adjusting for percentage rural, percentage poor or in fair health, percentage of adult smokers, county average daily PM2.5 and primary care physicians per 100,000. By March 30, 2020 relative risk became significantly greater than 1.00 in the most vulnerable counties. Upper SVI quartile counties had higher death rates on average beginning on March 30, 2020. By late August the lower quartiles for SVI began to exhibit increasing levels of cases and deaths. Dasgupta et al. (2020), examined COVID-19 cases from June 1 – July 25, 2020 relating them to the CDC's SVI. Areas with a higher proportion of individuals with nonwhite status, housing density, and crowded housing units were more likely to become COVID-19 hot spots; defined as areas where there is a >60% change in the most recent 3-7 day COVID-19 incidence rate (Dasgupta et al., 2020). Among the hot spot counties, those with greater SVI composite scores had higher COVID-19 incidence rates. Karaye and Horney (2020) examined local relationships between COVID-19 case counts and SVI utilizing geographically weighted regression (GWR) (Karaye & Horney, 2020). The study examined data from January 21 through May 12, 2020 and found, (after adjusting for population size, population density, number of persons tested, average daily sunlight, precipitation, air temperature, heat index, and PM2.5) that non-white status, limited English, household composition, transportation, housing and disability effectively predicted case counts in the U.S. Snyder and Parks (2020), in another spatial analysis (utilizing GWR), which did not utilize the CDC SVI, found that socio-ecological vulnerability to COVID-19 varied across the contiguous U.S., with higher levels of vulnerability in the

Southeast and low vulnerability in the Upper Midwest, Great Plains and Mountain West (Snyder & Parks, 2020).

2.3. Environmental Determinants of COVID-19

Even though we are only 11 months into the pandemic, there is growing evidence regarding environmental determinants of COVID-19 infection and mortality. Initially, it was hypothesized that COVID-19 may behave like many other respiratory infections and cases would subside in the summer months in the Northern Hemisphere due to increases in temperature (Hassan et al., 2020; Jamil et al., 2020; Prata et al., 2020). Therefore, many researchers have concentrated on temperature and less on other meteorological measurements as a factor in spread of SARS-CoV-2. A study conducted in Bangladesh, found that high temperature and high humidity significantly reduce the transmission of COVID-19 when analyzed using ordinary least squares regression (Haque & Rahman, 2020). Another analysis utilizing log-linear generalized additive models across 166 countries, revealed that temperature and relative humidity were also associated with a decrease in COVID-19 cases (Y. Wu et al., 2020). A 1° C increase in temperature was associated with a 3.08% reduction in daily new cases and 1.19% reduction in daily deaths across the studied countries. Relative humidity had a similar effect on cases and deaths. However, this is not surprising given the calculation of relative humidity employs a function of temperature. Rouen et al. , utilizing micro-correlation analysis using a 10-day moving window, found a negative correlation between temperature and outbreak progression (Rouen et al., 2020). Their research was conducted across 4 continents in both hemispheres. Sarkodie and Owusu (2020), using panel estimation techniques, focused their research on the top 20 countries with COVID-19 cases between January 22 and April 27, 2020 and found that high temperature and high relative humidity reduced the transmission of COVID-19 (Sarkodie & Owusu, 2020). However, low temperature, wind speed, dew/frost point, precipitation and surface pressure increased the infectivity of the virus.

At a much finer scale, research in China at 31 different provincial levels revealed a “biphasic” relationship with temperature, using distributed lag nonlinear models (P. Shi et al., 2020). Epidemic intensity was slightly reduced on days following higher temperatures and was associated with a decrease in relative risk. An investigation into temperature and precipitation’s relationship with COVID-19 in Oslo, Norway found maximum temperature and average temperature to be positively associate with COVID-19 transmission and precipitation to have a negative relationship, using non-parametric correlation estimaton (Menebo, 2020). Research in the sub-tropical cities of Brazil uncovered a negative relationship between temperature and COVID-19, using generalized additive models and polynomial

linear regression (Prata et al., 2020). Bashir et al. (2020) in New York City, New York, USA, between March and April of 2020, found a significant positive correlation between average temperature and minimum temperature on total cases, using Kendall and Spearman rank correlation tests (Bashir et al., 2020). Average temperature was significantly positively related to COVID-19 mortality and the minimum temperature was associated with new cases. Another study, utilizing multi-variate regression focusing on all U.S. counties from the beginning of the pandemic to April 14, 2020 found that higher temperatures were associated with a decrease in cases but not deaths (Li et al., 2020). This sample of studies demonstrates – particularly at finer spatial and temporal scales and depending on how temperature is sampled – the relationship between environmental variables and COVID-19 are complex and variable.

3. Methods

3.1. Study Area and Timeframe

This study focuses on the counties (sub-state administrative districts) located in the conterminous United States. We selected this subset due to Alaska and Hawai'i being non-contiguous and the error and complexity these "islands" would introduce into the spatial weighting matrices necessary for the spatiotemporal analysis. Furthermore, we focus on the timeframe from March 1 to December 31, 2020 (10 full months or 307 days); the first calendar year of the pandemic in the United States.

3.2. Data Collection

The data described below in steps 3.2.1 – 3.2.4, was used to create the dataset for the analysis (Johnson & Ravi, 2021).

3.2.1. COVID-19 Cases and Deaths

COVID-19 cases and deaths were collected from USAFACTS (*US Coronavirus Daily Cases by County*, 2021; *US Coronavirus Daily Deaths by County*, 2021). These data were retrieved in comma separated values (csv) format and were grouped into monthly cases and deaths for all counties of the contiguous U.S. (n = 3106); two counties were missing data. The expected number of cases and deaths, *E*, per county $area_i$, was determined by calculating the number of cases and deaths per month and computing the standardized infection rate (SIR) and standardized mortality rate (SMR) for each area for each month; the denominator in the rates = $E_{it,}$. This value is later used as an offset; expected number of cases/deaths at *area i* during *time t* .

3.2.2. *Social Vulnerability*

In this analysis we utilize the U.S. CDC's Social Vulnerability Index (SVI) (*CDC's Social Vulnerability Index (SVI)*, 2021; Flanagan et al., 2011). The SVI is composed of 18 variables that are related to social vulnerability and the local socio-ecology at the county or census tract-level for the entire U.S. We chose the SVI because it is highly cited in the literature and while there are inherent limitations in all vulnerability indices it demonstrates greater accuracy and relevancy in many studies (Bakkensen et al., 2017; Rufat et al., 2019; Spielman et al., 2020; Tate, 2012). The SVI variables are listed below in Table 1.

<<<Insert Table 1>>>

We utilize the percentage ((variable/total population) * 100) of each variable (except for per capita income where we used U.S. Dollars $) for all the selected counties standardized by their respective z-scores. The SVI also includes 4 themes, based on vulnerability domains, and a composite score of vulnerability. We utilize the composite SVI score for comparison of counties after modeling.

3.2.3. *Land Surface Temperature*

Daily land surface temperature (LST) measurements were collected from the Moderate Resolution Imaging Spectroradiometer (MODIS) TERRA satellite system; **MOD11a1.006** (Thome, 2020; Wan, Zhengming et al., 2015). MODIS data has a low spatial resolution (1 km) but a high (daily) temporal resolution. This remotely sensed data set, an emissivity corrected land surface temperature image for both daytime and nighttime, was collected from Google Earth Engine™ using geemap (Gorelick et al., 2017; Q. Wu, 2020). After collection, the daily values were averaged per month for each county in the conterminous U.S. resulting in a monthly average for daytime and nighttime LST. Areas where cloud cover interfered with image acquisition were assigned a NA value. These monthly averaged data were standardized by z-score.

3.2.4. *Meteorological Measurements*

For additional environmental variables, we utilized the North American Land Data Assimilation System (NLDAS). NLDAS contains land surface model datasets available hourly at 1km spatial resolution and is also accessible in Google Earth Engine™ (Cosgrove et al., 2003; Gorelick et al., 2017). All but one of the environmental variables listed in Table 2 were averaged by day and then by month for each county and standardized by their respective z-scores. However, for precipitation, we calculated a daily sum and then a monthly average before standardizing. After adjusting for multi-collinearity, the

measurements for specific humidity, longwave radiation, shortwave radiation and potential evaporation were removed. Specific humidity is the ratio of the mass of $H_20_{(v)}$ per total mass of the air parcel (kg/kg). This measurement is not a function of temperature and water content like relative humidity, but we still found a greater than 80% correlation across all counties throughout the time period with 2m AGL temperature. The other extraneous environmental variables had correlation coefficients above .80 with 2m AGL temperature.

3.3. Modelling and Specification

3.3.1. Bayesian Spatial-Temporal Framework

This study utilizes a Bayesian hierarchical spatiotemporal modeling approach initialized within the freely available R Statistical platform and the R-INLA package (*R: The R Project for Statistical Computing*, 2021; Rue et al., 2009). Furthermore, all models developed in this study were executed within Indiana University's High Performance Computing (HPC) environment (*Research and High Performance Computing*, 2020). Bayesian hierarchical modeling provides a flexible and robust framework where space-time components can be modeled in a straightforward manner. There are numerous introductions to Bayesian disease mapping to which we direct the novice (Best et al., 2005; Blangiardo et al., 2013; A. B. Lawson, 2013; A. Lawson & Lee, 2017; Moraga, 2020).

The Bayesian hierarchical methodology offers many benefits. For example, when creating disease models and relating counts to covariates, it is unreasonable to assume that one can collect all the necessary variables that account for a given response. The approach utilized here, allows for the inclusion of these "unknown" covariates as random effects within the model (Bernardinelli et al., 1995; Best et al., 2005; Congdon, 2019). These effects, in the spatial-temporal domain, account not only for spatial structure (spatial autocorrelation) and noise (overdispersion), but for temporal correlation and interaction between space and time (Besag et al., 1991; N. A. Samat & Pei Zhen, 2017; N.A Samat & Mey, 2017; Ugarte et al., 2014). Also, it is appropriate to utilize the standardized incidence rate (SIR), number of cases/expected number of cases, and standardized mortality rate (SMR), number of deaths/expected number of deaths, for country-level disease modelling. However, when examining disease measurements at a finer level (i.e. county-level or smaller), the SIR/SMR, a surrogate for relative risk, can be unstable and suffer large fluctuations due to some areas possessing a small population relative to the incidence of disease. The Bayesian modelling approach utilized "smooths" values of relative risk through space and time, by "borrowing" information both locally and globally, thereby reducing the impact of these instabilities (Bernardinelli et al., 1995; Besag et al., 1991).

In order to model relative risk, observed counts - in this study, COVID-19 cases of infection and deaths - Y_i are modelled using a Poisson distribution with mean $E_i\theta_i$; E = expected counts, θ_i is the relative risk (RR) of area i. The logarithm of RR_i is the sum of an intercept α and random effects accounting for extra-Poisson variability.

$$Y_i \sim Poisson(E_i\theta_i), i = 1, \ldots, 3106$$

$$\log(\theta_i) = \alpha + S_i + U_i$$

α is the overall risk in the study area and S and U are spatial random effects for area i modelling the spatial dependency structure (S) and the unstructured uncorrelated noise (U). Along with the inclusion of covariates, that determine risk (i.e. social vulnerability, environmental measurements), and/or other random effects the overall spatial model can be represented as

$$\log(\theta_i) = d_i\beta + S_i + U_i$$

$\boldsymbol{d}_i$ is a vector consisting of the intercept (α) and β is a coefficient vector; the fixed effects of the model. The parameters of the 27 fixed covariates included in this study are each assigned $\beta_{1:27} \sim Normal(\mu, \sigma)$ prior distributions.

A widely cited specification for the random spatial effects S and U, the Besag, York, Mollié (BYM) model, is regularly utilized in disease mapping studies (Besag et al., 1991). In the BYM model, the spatial random effect S is assigned a conditional autoregressive (CAR) distribution; smoothing the associated data based on a specified neighborhood structure, where neighbors are defined as areas sharing a common border.

$$S_i \vee S_{-i} \sim N\left(\overline{S}_{\delta_i}, \frac{\sigma_S^2}{n_{\delta_i}}\right)$$

$$\overline{S}_{\delta_i} = n_{\delta_i}^{-1} \sum_{j \in \delta_i} S_j$$

$$\delta_i = set\ of\ neighbors$$

$$n_{\delta_i} = number\ of\ neighbors\ of\ area\ i$$

The unstructured component *U* is modeled as independent and identically distributed (IID) with mean of zero and variance = σ_U^2. Therefore, data is shared both locally through the *S* component and globally through the *U* component.

In this study we follow the parameterization of the Besag, York, Mollié (BYM) model proposed by Simpson at al. (2015) that enables assigning penalized complexity (PC) priors (Simpson et al., 2015). This so-called "BYM2" model, incorporates a scaled spatially structured and unstructured component (S_* and U_*) and is defined as:

$$\log(\theta_i) = d_i\beta + \frac{1}{\sqrt{\tau}}\left(\sqrt{1-\varphi}\,S_{¿} + \sqrt{\varphi}\,U_{¿}\right)$$

The mixing - between S_* and U_* - parameter φ ($0 \le \varphi \le 1$) measures the proportion of variance explained by S_*. This scales the BYM2 model making it equal to the spatial model when φ = 1 and equal to only unstructured spatial noise when φ = 0 (Riebler et al., 2016). We set priors for these parameters following suggestions by Simpson et al., 2015. PC priors as their name suggest penalize model complexity. In this case they penalize based on the degree with which a given model deviates from a foundational assumption of no spatial dependency (φ = 0). Conjoining the random spatial effects for each area (S_* + U_*) is termed the convolutional spatial component. The exponential factor for the convolutional spatial effect $e^{(S_{¿}+U_{¿})}$ provides one with RR contribution of the random spatial effects additively. Repeating this procedure for either S_* or U_* will provide the relative contribution of each and allow the determination of the comparative contribution to variance (spatial fraction). This scaling parameterization makes the BYM2 representation more interpretable between models than the unscaled BYM model.

The above specification for $\log(\theta_i)$ can be extended into the spatial-temporal domain by the addition of further random effects.

$$\log(\theta_{it}) = d_i\beta + \frac{1}{\sqrt{\tau}}\left(\sqrt{1-\varphi}\,S_{¿} + \sqrt{\varphi}\,U_{¿}\right) + \gamma_t + \omega_t + \delta_{it}$$

Here, $\gamma_t \wedge \omega_t$, correspondingly represent the temporally structured and temporally unstructured random effect. Typically, γ_t, is modeled as a conditional autoregressive random walk of either order one or two (RW1, RW2), but there can be additional specifications (i.e. seasonal). In the present study, we model the temporally structured effect as RW1, and specify ω_t as a Gaussian exchangeable IID (

$\omega_t \sim Normal(0, \frac{1}{\tau_\omega})$¿. The space-time interaction component δ_{it}, represents a parameter vector that varies jointly through space and time. This vector allows for deviations from the space and time structure that expresses both dynamic spatial changes from one time frame to another and active temporal patterns from one area to another (Knorr-Held, 2000). Therefore, mapping δ_{it}, characterizes short-term clusters of disease activity that deviate from the space-time average over the study area at time *t*.

3.3.2. Ecological Regression

The covariates representing SVI and environmental measurements (after correcting for multi-collinearity) were included in the models as fixed effects, to examine which measurements are intricately linked to the spatiotemporal processes of the pandemic in the U.S. This study opted to include all variables that logically fit into the framework of vulnerability regardless of their statistical significance provided there is limited issues with multi-collinearity. In addition to accounting for the variables, much of this decision is based on the potentially poor inference generated by utilizing a stepwise framework and the Deviance Information Criteria (DIC) not significantly decreasing when the variables were removed (Greenland et al., 2016; Huberty, 1989). Even though a particular variable might not be statistically significant it is nonetheless important to see its effect in the model and to compare between COVID-19 cases and deaths when an alternative approach, aimed at reducing variables based on their significance, might result in less comparable models. Furthermore, since the response is logarithmic we calculate the exponential of the mean of the β coefficients and subtract from 1.00 to determine each variables effect on relative risk.

3.3.3. Model Selection Criteria

Deviance Information Criteria (DIC), was employed to select the most parsimonious model (Spiegelhalter et al., 2014). During exploratory data analysis, we examined two different prior specifications on the covariates:

$$\beta_{1:27} \sim Uniform(-\infty, +\infty)$$

$$\beta_{1:27} \sim Normal(\mu, \sigma)$$

These resulted in minimal changes to the models and we selected the specification for the covariates which produced the lowest DIC score; in this case the normal prior which produced a DIC score of ~10 less than the uniform specification (DIC Cases: Normal 243958.04 vs. Uniform 243969.28 / DIC Deaths:

Normal 90824.00 vs. Uniform 90835.76); a somewhat significant reduction (Spiegelhalter et al., 2014). Furthermore, we utilized all 4 types of spatial-temporal interactions suggested by Knorr-Held (2000) and found that Type I best fit our data, temporal stratification, and modeled process. Therefore, in our modeling approach we imposed no restrictions on when or where a space-time anomaly could occur (Type I spatial-temporal interaction).

3.4. Disease Mapping

Another key benefit of Bayesian inference is the creation of the posterior distribution where one can generate the probability of exceeding a certain threshold; the so-called exceedance probability. For this purpose, we mapped the convolutional spatial effect, $e^{(S_i+U_i)}$, and the spatially structured effect, $e^{(S_i)}$, at the county level. Also, we plotted the modeled space time interaction, $e^{\delta_{it}}$, which represents short-term clusters of activity relative to the study-area average at time *t*. In these maps we followed the classification rule followed by Richardson (2004); areas where $Pr(\theta_{it} \geq 1)\, is \geq .8$, is considered a hot spot, $Pr(\theta_{it} \geq 1)\, is .8 \geq .2$, is considered areas statistically similar to the national average, and $Pr(\theta_{it} \geq 1)\, is \leq .2$, signifies a cold spot; areas that represent infection/mortality rates below the national average (Richardson et al., 2004). Counties that are considered hot spots through the convolutional spatial effect, the spatially structured effect, and the space time interaction component are compared to the SVI composite score of the combined average and cold spot areas; areas where $Pr(\theta_{it} \geq 1)\, is < .8$ (non-hot spot areas). This assessment, utilizing notched box-plots, is employed to examine differences in the SVI composite score between the affected areas and in the case of $e^{\delta_{it}}$, different times.

3.5. Limitations and Caveats

A potential limitation of this study - likely not a significant constraint - is the lack of greater temporal resolution with regard to the social vulnerability index. In the case of the CDC's SVI, the index is calculated either yearly or every other year. This study focuses on COVID-19 on a monthly basis and there is no available capability of measuring changes in social vulnerability at that temporal resolution. That considered, it is likely that many of the individual variables that are used to define social vulnerablity do not change dramatically from one month to the next (Flanagan et al., 2011; Neelon et al., 2020; L. Shi & Stevens, 2021). Also, the internal limitations present in the SVI are extend to our analysis (Bakkensen et al., 2017; Rufat et al., 2019; Tate, 2012)

Another potential limitation is the number of zeros the dataset for COVID-19 fatalities contains; at least initially. We could have opted for a Zero-Inflated Poisson model for deaths but decided to keep

the hierarchical specifications and structure as consistent as possible between COVID-19 cases and deaths. By doing so, we eliminate a level of complexity within the model and maintain their comparability.

Finally, we decided not to place temporal lags on the environmental variables in relation to COVID-19 cases and deaths. Estimates for a latency in exposure to COVID-19 and onset of symptoms ranges from 2-24 days (CDC, 2020, p. 19; Grant et al., 2020). The WHO estimates that there is a temporal lag of 2-8 weeks from onset of symptoms to death in the most severe cases (Baud et al., 2020; Woolf et al., 2021). Given these large ranges of temporal associations and the aggregation of the data by month, we opted to compare SVI and environmental measurements on the date where a case or death was reported. Therefore, the coefficients should be interpreted in the proper context and with this consideration in mind.

4. Results

4.1. Temporal Trends in COVID-19 Cases and Deaths

COVID-19 cases by month per 100,000 people is presented in Figure 1A. Cases steadily increased until October, with an exponential increase through October until the end of 2020. Deaths from COVID-19, presented in Figure 1B, increased exponentially between March and April, then decreased and remained fairly stable through October, with another exponential increase in November and December.

<<<Insert Figure 1>>>

4.2. Spatiotemporal Ecological Regression

The coefficients resulting from the spatiotemporal ecological regression model for COVID-19 infections are presented in Table 2. The variables grayed out are considered not statistically significant since 0 falls within the 95% credibility interval (Wang et al., 2018). The variables unemployed, age 65 and up, disabled, crowded living, no vehicle, limited English, LST nighttime, AGL temperature, and precipitation have a negative relationship with COVID-19 infection risk. The strongest effect on cases is from the variable “no high school diploma” with a 20.60% increase in risk from a one standard deviation (6.34%) increase. AGL temperature is the second strongest with a 20.70% decrease in risk resulting from an 8.58k (1 standard deviation) increase. The remaining effects of each variable are presented in Table 2.

<<<Insert Table 2>>>

Table 3 contains the coefficient results from the posterior distributions of the ecological regression model for COVID-19 fatalities. The variables with the strongest impact (percent increase) on risk to COVID-19 deaths are non-white status (+40.19%), no high school diploma (35.21%), AGL temperature (-26.68%), age 65 and over (23.93%), and age 17 and under (19.42%) after a standard deviation increase in each variable respectively. Unemployed, single parent, mobile home, group quarters, poverty, uninsured, LST daytime were not statistically significant.

<<<Insert Table 3>>>

4.3. Modelled Temporal Trend

Figure 2 exhibits the temporally structured γ_t and unstructured ω_t effects for both cases and deaths. The left panel shows the modeled structured temporal effect for cases with all covariates, following the random walk order-1 and IID specification for unstructured temporal effects. The structured component shows an increase in relative risk until August with a decrease for the remainder of the year. The unstructured effect tends to fluctuate between being slightly above 1.00 to slightly below 1.00; with its 95% credibility envelope easily encompassing 1.00. In the right panel for deaths, γ_t, increases until June, drops in July, increases until September, and drops until the end of the study period. Similar to the cases panel, ω_t, fluctuates between being slightly above 1.00 to slightly below 1.00.

<<<Insert Figure 2>>>

4.4. Spatial Effects

The convolutional spatial effect and the spatially structured effect for COVID-19 cases are mapped in Figure 3A. These figures show the probability that the relative risk exceeds 1.00; the national average. There is a strong clustering of high probability for the convolutional spatial effect in Florida, Alabama, Mississippi, Louisiana, Arkansas, Tennessee, Iowa, and Arizona. There is sporadic clustering of high probability most prominent of which is in Indiana, Kansas, and Colorado. There is significant clustering of low probability areas in the Northeast, Pacific Northwest, Upper Atlantic Coast, Upper Midwest, Michigan, and West Virginia. The spatially structured effect for cases follow a similar pattern to the convolutional effect as it explains 82.7% of the variance in the overall spatial effects. Key differences include some higher probabilities in Connecticut and Iowa. Probability is lower in Southern California, Michigan, and New Mexico.

<<<Insert Figure 3>>>

Figure 3B shows exceedance probabilities for the convolutional spatial effect and the spatially structured effect for deaths. There is a strong clustering of high probabilities in New Mexico, Indiana, Louisiana, Eastern Pennsylvania, and the Northeast megalopolis. Higher probabilities are scattered throughout the Southeast and portions of the Midwest into Montana, Idaho and Eastern Oregon. The spatially structured effect accounts for 60.9% of the variance for (S_*+U_*) so similarities are expected, although not as high of a degree as in the spatial effect for cases. The most notable difference is the increase in clustering in the Southeast, the increase in Indiana, and less sporadic dispersion of counties in the Midwest into the Mountain West.

4.5. Spatiotemporal Interaction

4.5.1. Cases

Exceedance probabilities using 1.00 as a threshold for the spatiotemporal interaction term are presented in Figure 4, for cases, and Figure 7, for deaths. Initial clustering of high probabilities in March are noted in the Northeast, especially New York, Florida, Louisiana and counties containing some of the major metropolitan areas around the country (i.e., Atlanta, Denver, Detroit, Chicago, Indianapolis, San Diego, Los Angeles, San Francisco, Portland, and Seattle). Lower probabilities are less stable than in April and this is likely due to the average risk being so low at this point in time. In April, the areas noted previously in March have expanded in what appears to be a diffusion pattern; in many cases doubling in extent. Lower probability areas are more prominent and are focused in the Upper Midwest south through the Great Plains into Texas. There is another notable area of low probabilities in Ohio south through West Virginia, west into Kentucky and further south into Tennessee. In May, areas previously noted at high probability have remained fairly stable, with a notable decrease in probability in upper New York, the Upper Northeast and Michigan. There is a crescent of lower probability extending from Western Pennsylvania, through West Virginia and west to Kansas and Oklahoma. There is a notable increase in cases in Minnesota and upper Iowa.

By June there are some significant changes to areas of high probability. The pandemic seems to have settled much more into the southern U.S. focusing again in Florida, Alabama, Mississippi, Louisiana, eastern Texas, South and North Carolina. Probabilities have decreased in the Northeast and throughout much of the Midwest apart from much of Iowa and southern Minnesota. High probabilities remain in Arizona and have expanded into Utah, Nevada, and much of California. July shows a further solidification of the pandemic in the southern U.S. extending from the Atlantic to the Pacific coast. Arizona northward into Utah and Idaho has joined this high probability area. Metropolitan areas in

Minnesota, Wisconsin, Ohio, and Pennsylvania are showing renewed higher probabilities. The Northeast and the middle Midwest are the lowest probability areas, with Illinois and Indiana continuing a trend of decreasing activity. By August, the pandemic is shifting out of the southern U.S. and into the Midwest with increases from Tennessee into Minnesota, North Dakota, and South Dakota. The Mountain West is exhibiting an increase in activity as is much of California. Arizona and New Mexico are showing a decrease in probability and the Northeast remains firmly in the lower category.

September, witnesses the pandemic lessening in the southern U.S., but the increases previously noted in the Upper Midwest have become even more pronounced, with Missouri, Illinois, Iowa, Wisconsin, Minnesota, Kansas, Nebraska, and North and South Dakota heavily burdened. The pandemic continues to lessen in Arizona and California. By October the pandemic continues to rage in the Upper Midwest affecting much of the counties in the states from Wisconsin to Idaho. There is a notable lessening in southern Minnesota and Iowa. The pandemic continues to subside in the southern U.S. and remains stable in the Northeast. November, expresses a further strengthening in the upper Midwest with areas previously showing a lessening pattern overrun by cases. Much of the U.S. is affected apart from the southern U.S., California, Arizona, Washington, and the Northeast. Through December, the pandemic has diminished in the Upper Midwest and shifted with higher probabilities into the Northeast and Texas and has further reasserted itself on the Pacific Coast. The upper Midwest and extreme South continue a waning effect apart from Florida which displays a resurgence.

<<<Insert Figure 4>>>

4.5.2. Deaths

The spatiotemporal interaction term probability exceedances for deaths are shown in Figure 5. As expected there is not much activity in March apart from a deaths in some major cities. By April, deaths begin to show in many of the major metropolitan areas of the U.S. with a clustering of high probabilities in the New York City area, Chicago, Detroit, Indianapolis, Atlanta, San Diego, Los Angeles, San Francisco, Portland, and Seattle. Through May, many of these areas of high probability have expanded in a similar apparent diffusion pattern to cases a month or so earlier. Much of the Northeast megalopolis is affecting along with Detroit, Cleveland, Pittsburg, Chicago, Indianapolis, Nashville, Birmingham, New Orleans, and counties in New Mexico, Arizona, and Southern California. In June, many of the counties around these same cities have become even more heavily burdened, with notable clusters in the Northeast, Ohio, eastern Michigan, northern Illinois, central Indiana, central Mississippi,

Arizona and southern California. Through July many of these areas are showing a decrease in deaths apart from the northeast megalopolis, Chicago, central Indiana, Arizona and southern California.

Through August, probabilities of high deaths have shifted to the southern U.S., while lessening in the Northeast and Midwest. The probabilities have strengthened in Arizona and much of California. September witnesses a further strengthening of probabilities in the southern U.S. affecting much of South Carolina, southern Georgia and much of Florida. The pandemic continues to strengthen in Arizona and California. The waning trend has continued throughout much of the upper Northeast and the Ohio Valley. By October, these shifts continue with sporadic higher probabilities throughout the southern U.S. The lessening continues in the Ohio Valley and upper Northeast as Arizona begins a trend of diminishing deaths. November witnesses a shift in the probability of deaths into the upper Midwest, as suspected considering cases witnessed a similar trend a few months prior. The pandemic begins to lessen in the southern U.S.; but contains some sporadic high probabilities. However, the shift to the north is apparent. Lessening continues in Arizona and California with the same effect stable in the upper Northeast. Through December, the shift in to the upper Midwest is even more evident with the Mountain West now included. December further witnesses a resurgent trend in the upper Northeast, especially northern New York, Vermont, New Hampshire and Maine. The continued lessening in the southern U.S., Arizona and California is noteworthy.

<<<Insert Figure 5>>>

4.6. Comparisons of Composite Vulnerability in Hot and Cold Spots

Figure 6A displays the boxplots comparing hot and cold spots for the convolutional spatial effect ($e^{(S_i+U_i)}$) for COVID-19 infections. The light gray distribution is for areas having a probability less than .80 of exceeding 1.00 (cold spots) and the red distribution for areas where the probability is greater than or equal to .80 of exceeding 1.00 (hot spots). Hot spot areas have a significantly higher SVI composite score compared to the low probability areas. Figure 6B illustrates the boxplot comparing areas delineated in the same way to the composite SVI score for COVID-19 fatalities. Likewise, area with a higher probability of COVID-19 deaths have a statistically significant higher SVI composite score. The composite SVI score for the spatially structured effect is similarly higher in areas of higher probability, which is expected based on the percent of variance explained in the convolutional effect by that component.

<<<Insert Figure 6>>>

Taken in the spatiotemporal context, the relationship between higher SVI composite scores and the hot and cold spots are not as straightforward. Figure 7A and 7B displays boxplots comparing the SVI composite score for the areas by month that have a high probability of exceeding 1.00; within the spatiotemporal interaction component $\left(e^{\delta_{it}}\right)$. The individual plots are delineated in the same way as above. Comparing the distributions for SVI composite scores to cases (9A) from April through August the score is higher and statistically significant in hot spot counties. Similarly, for deaths (9B) the SVI score is higher for the counties involved for the months July – October. The mean SVI score for the hot spots during these month nears or exceeds the 3rd quartile value in the cold spot counties. November also witnesses a higher but not statistically significant SVI score interpreted via the notches in the boxplots.

<<<Insert Figure 7>>>

5. Discussion

5.1. Relationships Between Social Vulnerability Variables and COVID-19 Infections and Deaths

Dividing the SVI index into its separate variables and including them in the hierarchical model offered some noteworthy results. Counties that have a high percentage of those without a high school diploma and non-white status have the strongest positive effect on risk; raising risk for COVID-19 infections by 21.44% and 20.60% respectively. Counties with greater percentages of those aged 17 and under are at greater risk; increasing risk by 11.78% for each standard deviation increase (3.43%). Living in a multi-unit dwelling increased risk by 11.18% which is expected due to the density of living spaces. However, living in a home with more than 10 people (crowded living) lowered risk by 5.13%. Additionally, percent unemployed, disabled, no vehicle, limited English, and uninsured negated risk to varying degrees (see Table 2). These findings tend to support other studies that show COVID-19 having a disparate effect in non-white communities (LaVeist 2005; Shi and Stevens 2021; Singu et al. 2020; Karaye and Horney 2020; Khazanchi et al. 2020). They especially support Karaye and Horney (2020) with conclusions related to non-white status, but not as significantly as their findings related to limited English, household composition, transportation, housing and disability (Karaye & Horney, 2020). This is likely due to their study focusing on cases through May 12, 2020, so the comparison in the amount of data and the timeframe of investigation is different. Dasgupta et al. (2020) found that non-white status and crowded housing were more likely to become COVID-19 hot spots during June and July 2020. Our

study supports these findings for non-white status, multi-unit dwelling and group quarters throughout the year 2020.

The SVI index variables and the relationship with death from COVID-19 possesses some dissimilarities compared to cases. Non-white status had the strongest impact on risk of death; increasing relative risk by 40.19% when raised by one standard deviation (19.84%). No high school diploma is again ranked second with a 35.31% increase. Furthermore, age 65 and over demonstrates an increase in relative risk of death by 23.93%; this variable was statistically insignificant in the model for cases. Age 65 and above is a highly recognized individual risk factor for COVID-19 severe disease and mortality (Woolf et al., 2021). Age 17 and under is again relatively high in its impact on risk with a 23.93% increase. Limited English lowered risk by 13.35% (compared to 5.10% for cases) and Multi-unit dwelling increased risk by 11.29%. The result for limited English potentially implies that decreased social connections due to the perceived language barrier may have a preventative effect. The remaining variables are either not statistically significant or offer minimal (less than 10%) impact on deaths. Although there are not as many previous studies that examined deaths as opposed to cases, these findings especially support Khazanchi et al. (2020) where non-white status resulted in a 4.74-fold increase in death.

5.2. Relationships Between Environmental Variables and COVID-19 Infections and Deaths

Regarding environmental variables in relation to COVID-19 cases, above ground level (AGL) temperature had the strongest relationship. A one standard deviation increase (8.58K) in temperature results in a 20.70% reduction in relative risk for cases when examined within the context of the entire study. Nighttime LST results in a 11.82% decrease in risk when increased by a single standard deviation (14.61K). Atmospheric pressure produces a 12.34% increase in relative risk for cases when increased by 5465.36 Pa. Precipitation increases lower risk by 4.07% and wind speed (2.84%) and direction (11.79%) have a positive effect on risk of infection. These findings support research which points to increases in temperature lowering the risk of COVID-19 infections (Haque & Rahman, 2020; Prata et al., 2020; Rouen et al., 2020; Sarkodie & Owusu, 2020; P. Shi et al., 2020). Our precipitation finding contradicts Sarkodie and Owusu (2020) where they found a positive association between temperature and COVID-19 cases. However, it does support Menebo (2020) where both variables had a negative association. In relationship to winds, prior research has found a negative association with wind speed and COVID-19

incidence (Islam et al., 2020; Şahin, 2020). However, in our study average monthly wind direction, when the azimuth is increased by 35 degrees, increased risk by 11.79% and average monthly wind speed, when increased by .69m/s, raises risk by 2.84%. While this relationship is difficult to explain it is worth noting and we opted to keep wind data in the analysis to account for its potential effects.

AGL temperature has the strongest effect on risk from death of the environmental variables with a 26.68% decrease in risk for every 8.58K increase. LST nighttime also lowers risk by 11.14% when increased by 14.61K. As atmospheric pressure increases by 5465.36 Pa, it raises the relative risk of death by 14.66%. Precipitation increases by a standard deviation lower risk by 5.62%. Wind direction is significant in the model for deaths. As the azimuth of wind increases by 35° the relative risk lowers by 6.99%. Wind speed is not statistically significant. Again, these findings are for wind 10m above ground level and support the notion that wind may have some impact on COVID-19, although it's difficult to infer. More research is needed on wind's relationship, especially at a finer temporal stratification (i.e. daily) where it should be easier to infer the relationship (Islam et al., 2020; Şahin, 2020).

5.3. Temporal Structure of the Pandemic in the United States

Cases and deaths have clearly increased throughout the year 2020 as evidenced by Figure 1. However, the modeled relative risks have fluctuated throughout the time period (Figure 2). These relative risks as modeled through random walk order – 1 show rapid increases in cases and deaths from March through much of the summer of 2020. Decreases are then evident for the remainder of the year. Temporal relative risk from death shows more fluctuation than cases but also presents a steady decline in average relative risk for the last few months of 2020. On the surface, these results (Figure 1 vs Figure 2) may seem contradictory. Apart from one chart showing per capita COVID-19 cases/deaths and the other modeled relative risk, closer examination of the maps of the spatiotemporal interactions (Figures 4 and 5) show there are more counties affected in the latter months of 2020. This observation suggests the pandemic has broadened in spatial extent, especially in regard to cases during October and November, but has become less intense overall as measured by relative risk. This finding also implies the pandemic may be starting to decline in average intensity – in the U.S. – as we head into 2021. The average non-modelled standardized infection rate (SIR) and standardized mortality rate (SMR) across all counties, support this and demonstrate a similar relative risk trend and corresponding decrease for the latter months of 2020.

5.4. Spatial Structure of the Pandemic in the United States

After adjusting for the fixed effects of the covariates and the temporally structured and unstructured random effects, the convolutional spatial effect risk map and the spatially structured effect risk map identified counties at increased risk of COVID-19 infection and death throughout the study period. The most prominent spatial aspect of relative risk to COVID-19 infection were the clusters most heavily focused in the southern U.S., and the states of Indiana, Iowa, and New Mexico. Somewhat supported by Snyder and Parks (2020), with their geographically weighted regression (GWR) approach, this result was not completely unanticipated (Snyder & Parks, 2020). There were additional sporadic areas of increased risk scattered throughout the Midwest, U.S. South and Great Plains. Strong degrees of spatial autocorrelation, which supports the clustering, as modelled with the spatially structured effect, were present in many of those same areas.

Convolutional risk from COVID-19 deaths were less focused than cases but were still prevalent in the U.S. South, especially Louisiana and Tennessee. Indiana and the megalopolis region of the Northeast were also part of this cluster. Interestingly, the latter region did not present as an increased risk area relative to cases throughout the study period. This is a potentially alarming finding suggesting this region has a higher rate of death relative to the number of cases throughout the year. Further west, nearly every county in New Mexico was at an elevated risk. A large degree of spatial autocorrelation (as the spatially structured effect accounts for nearly 61% of the variance between the two components) is also present in many of these identical areas based on the spatially structured risk from death.

5.5. Spatiotemporal Structure of the Pandemic in the United States

The spatiotemporal interaction term is a random effect and can be interpreted as the modelled residual risk after accounting for the fixed effects, spatially structured and unstructured and temporally structured and unstructured effects. This represents short-term (month long in our study) sporadic clusters of COVID-19 cases and deaths. During the time period of the study, it is notable that areas impacted by the pandemic shift drastically throughout the country. Cases shift from major metropolitan areas and the Northeast, into the Southern and Southwest U.S. in the summer months. This trend supports evidence found by Snyder and Parks (2020) that the pandemic was focusing in the highly vulnerable U.S. South by mid-summer. By late summer and early autumn, elevated cases have moved into the Upper Midwest with a second shift into the Southwest and a re-emergence in the Northeast by the end of 2020.

The short-term patterns in deaths follow a similar trend but are delayed by approximately a month to month-and-a-half and are not as large in extent or as contiguous. This implies there is a temporal delay from cases to deaths that falls within the WHO suggestion of 2-8 weeks; tending toward the higher end (Baud et al., 2020). Elevated deaths shift into the Southwest by early summer and into the Southern U.S. by late summer. Increased probabilities of above average deaths have moved from the Northeast megalopolis by August as they refocus in the South. By November, deaths have moved into the Upper Midwest and are beginning to shift out of the U.S. South (apart from Florida); into less socially vulnerable areas of the country. December shows the U.S. South to be below average risk along with the Southwest. A further broadening is evident in the Upper Midwest and Mountain West.

5.6. Composite Social Vulnerability in Hot and Cold Spots

The convolutional spatial effect for COVID-19 cases was classified into hot and cold spots based on the criteria of Richardson et al. (2004). For COVID-19 cases of infection, areas that were considered hot spots throughout the course of the study period had higher SVI composite scores (.57) than those that were considered cold spots (.48). In relation to deaths, there was a similar trend with higher SVI composite scores for hot spot areas (.55) versus cold spot areas (.49). These differences were statistically significant at .05 confidence interval. These relationships provide further evidence to studies finding locations that have higher vulnerability scores to be more at-risk of COVID-19 infection and death (Dasgupta et al., 2020; Karaye & Horney, 2020; Khazanchi et al., 2020).

When comparing SVI composite scores to hot and cold spots for cases and deaths based on the spatiotemporal interaction term, the relationship is not as straightforward. In respect to cases, our study supports the findings of Nayak et al., where the relationship between composite SVI score and COVID-19 cases is not statistically significant in the first month of the pandemic in the U.S.(Nayak et al. 2020). The SVI composite score is higher from April to August and again in December in hot spot counties. This observation supports Neelon et al., where counties with higher SVI scores contained higher COVID-19 cases through August. Furthermore, this result verifies Dasgupta et al., where higher SVI scores in June and July 2020 supported the probability of becoming a COVID-19 hotspot. Neelon et al. also discovered that in August, counties with lower composite SVI scores were beginning to be affected. This is also supported by our study, but the composite index is lower in hot spot areas extending into the months after August. For deaths, the index is higher from July through October, with the remaining months either lower or not statistically significant. In the case of August, September and October, the SVI

composite score is much higher in hot spots than in cold spots. The mean SVI score in hot spot areas is approaching the 3rd quartile value in cold spot areas.

Much of this observed shift in the relationship between COVID-19 and vulnerability is due to the spatial-temporal nature of the pandemic in the U.S. Cases shift into the southern and Southwest U.S. in the summer months. These areas are known to have significantly larger numbers and percentages of vulnerable than most other areas of the country (L. Shi & Stevens, 2021). Likewise, in the instance of deaths in August, September, and October, they are likewise focused in some of the more vulnerable locations of the American Southeast and Southwest. These results support a more complex or nuanced relationship between temperature and COVID-19 cases and deaths; especially when examined in a smaller spatial and temporal context (Bashir et al., 2020; P. Shi et al., 2020). The overall time period relationships are not stable across all counties for all time periods. By utilizing the spatiotemporal modelling approach used in this study we are able to uncover this more elusive space-time relationship. Alternatively, in areas where the SVI composite score in hot spots is lower than in cold spots, the pandemic has shifted into less socially vulnerable areas of the country; the Upper Midwest and Northeast. However, a key finding is that during the summer months (for cases) and autumn (for deaths) the pandemic seemed to shift into warmer and more socially vulnerable areas of the country. Future spatiotemporal analyses, after the behavior of the pandemic in 2021, are needed to determine if this is likely to be a trend the virus follows or if this is simply how the pandemic initially diffused in the U.S.

6. Conclusion

Bayesian hierarchical modeling provides a flexible and robust framework in which to model complex spatiotemporal systems. This study presents the findings of fitting a Bayesian hierarchical spatiotemporal model to COVID-19 cases and deaths in the United States for the year 2020. The data collection and modeling framework are explained in detail for straightforward replication. This hopefully will foster continued effort in modeling the spatiotemporal nature of the pandemic in the U.S. and abroad.

A key finding of this study is the spatiotemporal character of the pandemic in the U.S. after accounting for social vulnerability, environmental measurements and spatial and temporal random effects. In terms of cases, the pandemic shifted into the U.S. South and Southwest during the hotter months of the year, which are the most vulnerable regions of the U.S. Deaths followed the same pattern only with a 1-2-month temporal lag. This demonstrates a potentially alarming trend if this pattern or a similar one is repeated in other years. These highly vulnerable areas are already under significant stress

at this time of the year and the introduction of another stressor on a regular basis could be catastrophic in some areas. Further alarming, is the cluster of deaths in the Northeast Megalopolis region.

Studies such as the one presented here provide insight into the complicated mix of social and environmental factors relating to vulnerability. We demonstrate that relationships between COVID-19 cases/deaths, social vulnerability, and environmental measurements are spatially and temporally variable. Even though many of the findings presented here are supportive of other studies, more work is needed in the spatial-temporal domain of the pandemic. One primary effort should be modelling the spatiotemporal structure at a finer temporal scale (i.e. weekly). This could elucidate other relationships with the examined variables and allow for more elaborate spatial temporal interaction specifications (Type IV spatial temporal interaction). Such examinations, perhaps at finer spatial scales (i.e. census tract), could also allow us to focus on some of this study's more alarming results.

Acknowledgments

This research was funded by an internal Indiana University grant sponsored by the Indiana University Vice President of Research, IU COVID-19 Rapid Research OVPR and the Indiana Space Grant Consortium.

References

Bakkensen, L. A., Fox−Lent, C., Read, L. K., & Linkov, I. (2017). Validating Resilience and Vulnerability Indices in the Context of Natural Disasters. *Risk Analysis*, *37*(5), 982–1004. https://doi.org/10.1111/risa.12677

Bashir, M. F., Ma, B., Bilal, Komal, B., Bashir, M. A., Tan, D., & Bashir, M. (2020). Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Science of The Total Environment*, *728*, 138835. https://doi.org/10.1016/j.scitotenv.2020.138835

Baud, D., Qi, X., Nielsen-Saines, K., Musso, D., Pomar, L., & Favre, G. (2020). Real estimates of mortality following COVID-19 infection. *The Lancet Infectious Diseases*, *20*(7), 773. https://doi.org/10.1016/S1473-3099(20)30195-X

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space—Time variation in disease risk. *Statistics in Medicine*, *14*(21–22), 2433–2443. https://doi.org/10.1002/sim.4780142112

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), 1–20.

Best, N., Richardson, S., & Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, *14*(1), 35–59.

Bilal, U., Barber, S., Tabb, L., & Diez-Roux, A. V. (2020). Spatial Inequities in COVID-19 Testing, Positivity, Incidence and Mortality in 3 US Cities: A Longitudinal Ecological Study. *MedRxiv*.

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, *7*, 39–55. https://doi.org/10.1016/j.sste.2013.07.003

CDC. (2020, February 11). *Coronavirus Disease 2019 (COVID-19)*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/index.html

CDC's Social Vulnerability Index (SVI). (2021, January 19). https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

Coelho, F. C., Lana, R. M., Cruz, O. G., Villela, D. A., Bastos, L. S., Pastore y Piontti, A., Davis, J. T., Vespignani, A., Codeço, C. T., & Gomes, M. F. (2020). Assessing the spread of COVID-19 in Brazil: Mobility, morbidity and social vulnerability. *PloS One*, *15*(9), e0238214.

Congdon, P. (2019). Spatial heterogeneity in Bayesian disease mapping. *GeoJournal*, *84*(5), 1303–1316.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., & Meng, J. (2003). Real−time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, *108*(D22), 2002JD003118. https://doi.org/10.1029/2002JD003118

Dasgupta, S., Bowen, V. B., Leidner, A., Fletcher, K., Rose, C., Cha, A., Kang, G., Dirlikov, E., Pevzner, E., & Rose, D. (2020). Association Between Social Vulnerability and a County's Risk for Becoming a COVID-19 Hotspot—United States, June 1–July 25, 2020. *Morbidity and Mortality Weekly Report*, *69*(42), 1535.

DeCaprio, D., Gartner, J., Burgess, T., Kothari, S., & Sayed, S. (2020). Building a COVID-19 vulnerability index. *ArXiv Preprint ArXiv:2003.07347*.

Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management*, *8*(1). https://doi.org/10.2202/1547-7355.1792

Freitas, C. M. de, & Cidade, N. da C. (2020). COVID-19 AS A GLOBAL DISASTER: Challenges to risk governance and social vulnerability in Brazil. *Ambiente & Sociedade*, *23*.

Gaynor, T. S., & Wilson, M. E. (2020). Social Vulnerability and Equity: The Disproportionate Impact of COVID-19. *Public Administration Review*, *80*(5), 832–838.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Grant, M. C., Geoghegan, L., Arbyn, M., Mohammed, Z., McGuinness, L., Clarke, E. L., & Wade, R. G. (2020). The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PLoS ONE*, *15*(6). https://doi.org/10.1371/journal.pone.0234765

Greenland, S., Daniel, R., & Pearce, N. (2016). Outcome modelling strategies in epidemiology: Traditional methods and basic alternatives. *International Journal of Epidemiology*, *45*(2), 565–575. https://doi.org/10.1093/ije/dyw040

Haque, S. E., & Rahman, M. (2020). Association between temperature, humidity, and COVID-19 outbreaks in Bangladesh. *Environmental Science & Policy*, *114*, 253–255. https://doi.org/10.1016/j.envsci.2020.08.012

Hassan, M. M., Zowalaty, M. E. E., Khan, S. A., Islam, A., Nayem, M. R. K., & Järhult, J. D. (2020). Role of Environmental Temperature on the Attack rate and Case fatality rate of Coronavirus Disease 2019 (COVID-19) Pandemic. *Infection Ecology & Epidemiology*, *10*(1), 1792620. https://doi.org/10.1080/20008686.2020.1792620

Huberty, C. (1989). Problems with Stepwise Methods—Better Alternatives. *Advances in Social Science Methodology*, *1*, 43–70.

Islam, N., Shabnam, S., & Erzurumluoglu, A. M. (2020). Temperature, humidity, and wind speed are associated with lower Covid-19 incidence. *MedRxiv*, 2020.03.27.20045658. https://doi.org/10.1101/2020.03.27.20045658

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*. https://doi.org/10.3389/fpubh.2020.00436

Johnson, D. P., & Ravi, N. (2021). *Spatiotemporal Associations Between Social Vulnerability, Environmental Measurements, and COVID-19 in the Conterminous United States* [Data set]. IUPUI University Library. https://doi.org/10.7912/D2/23

Karaye, I. M., & Horney, J. A. (2020). The impact of social vulnerability on COVID-19 in the US: An analysis of spatially varying relationships. *American Journal of Preventive Medicine*, *59*(3), 317–325.

Khazanchi, R., Beiter, E. R., Gondi, S., Beckman, A. L., Bilinski, A., & Ganguli, I. (2020). County-level association of social vulnerability with COVID-19 cases and deaths in the USA. *Journal of General Internal Medicine*, *35*(9), 2784–2787.

Kim, S. J., & Bostwick, W. (2020). Social Vulnerability and Racial Inequality in COVID-19 Deaths in Chicago. *Health Education & Behavior*, 1090198120929677.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, *19*(17–18), 2555–2567.

Lancet, T. (2020). Redefining vulnerability in the era of COVID-19. *Lancet (London, England)*, *395*(10230), 1089.

LaVeist, T. A. (2005). *Minority Populations and Health: An Introduction to Health Disparities in the United States*. John Wiley & Sons.

Lawson, A. B. (2013). *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology*. CRC press.

Lawson, A., & Lee, D. (2017). Bayesian disease mapping for public health. In *Handbook of statistics* (Vol. 36, pp. 443–481). Elsevier.

Li, A. Y., Hannah, T. C., Durbin, J. R., Dreher, N., McAuley, F. M., Marayati, N. F., Spiera, Z., Ali, M., Gometz, A., Kostman, J., & Choudhri, T. F. (2020). Multivariate Analysis of Black Race and Environmental Temperature on COVID-19 in the US. *The American Journal of the Medical Sciences*, *360*(4), 348–356. https://doi.org/10.1016/j.amjms.2020.06.015

Menebo, M. M. (2020). Temperature and precipitation associate with Covid-19 new daily cases: A correlation study between weather and Covid-19 pandemic in Oslo, Norway. *Science of The Total Environment*, *737*, 139659. https://doi.org/10.1016/j.scitotenv.2020.139659

Mishra, S. V., Gayen, A., & Haque, S. M. (2020). COVID-19 and urban vulnerability in India. *Habitat International*, *103*, 102230.

Mohanty, S. K. (2020). Contextualising geographical vulnerability to COVID-19 in India. *The Lancet Global Health*, *8*(9), e1104–e1105.

Moraga, P. (2020). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. CRC Press. https://www.paulamoraga.com/book-geospatial/

Nayak, A. (2020). *Impact of Social Vulnerability on COVID-19 Incidence and Outcomes in the United States | medRxiv*. https://www.medrxiv.org/content/10.1101/2020.04.10.20060962v2

Neelon, B., Mutiso, F., Mueller, N. T., Pearce, J. L., & Benjamin-Neelon, S. E. (2020). Spatial and temporal trends in social vulnerability and COVID-19 incidence and death rates in the United States. *MedRxiv*.

Persad, G., Peek, M. E., & Emanuel, E. J. (2020). Fairly Prioritizing Groups for Access to COVID-19 Vaccines. *JAMA*, *324*(16), 1601. https://doi.org/10.1001/jama.2020.18513

Prata, D. N., Rodrigues, W., & Bermejo, P. H. (2020). Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil. *Science of The Total Environment*, *729*, 138862. https://doi.org/10.1016/j.scitotenv.2020.138862

R: The R Project for Statistical Computing. (2021). https://www.r-project.org/

Research and high performance computing. (2020). https://kb.iu.edu/d/apes

Richardson, S., Thomson, A., Best, N., & Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*, *112*(9), 1016–1025.

Riebler, A., Sørbye, S. H., Simpson, D., & Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, *25*(4), 1145–1165. https://doi.org/10.1177/0962280216660421

Rouen, A., Adda, J., Roy, O., Rogers, E., & Lévy, P. (2020). COVID-19: Relationship between atmospheric temperature and daily new cases growth rate. *Epidemiology & Infection*, *148*. https://doi.org/10.1017/S0950268820001831

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rufat, S., Tate, E., Emrich, C. T., & Antolini, F. (2019). How Valid Are Social Vulnerability Models? *Annals of the American Association of Geographers*, *109*(4), 1131–1153. https://doi.org/10.1080/24694452.2018.1535887

Şahin, M. (2020). Impact of weather on COVID-19 pandemic in Turkey. *Science of The Total Environment*, *728*, 138810. https://doi.org/10.1016/j.scitotenv.2020.138810

Samat, N. A., & Pei Zhen, W. (2017). Relative Risk Estimation for Dengue Disease Mapping in Malaysia based on Besag, York and Mollié Model. *PERTANIKA JOURNAL OF SCIENCE AND TECHNOLOGY*, *25*(3), 759–765.

Samat, N.A, & Mey, L. W. (2017). Malaria disease mapping in Malaysia based on Besag-York-Mollie (BYM) Model. *Journal of Physics: Conference Series*, *890*(1), 012167.

Sarkodie, S. A., & Owusu, P. A. (2020). Impact of meteorological factors on COVID-19 pandemic: Evidence from top 20 countries with confirmed cases. *Environmental Research*, *191*, 110101. https://doi.org/10.1016/j.envres.2020.110101

Seddighi, H. (2020). COVID-19 as a natural disaster: Focusing on exposure and vulnerability for response. *Disaster Medicine and Public Health Preparedness*, 1–2.

Shi, L., & Stevens, G. (2021). *Vulnerable Populations in the United States*. John Wiley & Sons.

Shi, P., Dong, Y., Yan, H., Zhao, C., Li, X., Liu, W., He, M., Tang, S., & Xi, S. (2020). Impact of temperature on the dynamics of the COVID-19 outbreak in China. *Science of The Total Environment*, *728*, 138890. https://doi.org/10.1016/j.scitotenv.2020.138890

Simpson, D. P., Rue, H., Martins, T. G., Riebler, A., & Sørbye, S. H. (2015). Penalising model component complexity: A principled, practical approach to constructing priors. *ArXiv:1403.4630 [Stat]*. http://arxiv.org/abs/1403.4630

Singu, S., Acharya, A., Challagundla, K., & Byrareddy, S. N. (2020). Impact of Social Determinants of Health on the Emerging COVID-19 Pandemic in the United States. *Frontiers in Public Health*, *8*. https://doi.org/10.3389/fpubh.2020.00406

Snyder, B. F., & Parks, V. (2020). Spatial variation in socio-ecological vulnerability to COVID-19 in the contiguous United States. *Health & Place*, *66*, 102471.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *76*(3), 485–493.

Spielman, S. E., Tuccillo, J., Folch, D. C., Schweikert, A., Davies, R., Wood, N., & Tate, E. (2020). Evaluating social vulnerability indicators: Criteria and their application to the Social Vulnerability Index. *Natural Hazards*, *100*(1), 417–436. https://doi.org/10.1007/s11069-019-03820-z

Tate, E. (2012). Social vulnerability indices: A comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards*, *63*(2), 325–347. https://doi.org/10.1007/s11069-012-0152-2

Thome, K. (2020). *MODIS | Terra*. 610 WebDev. https://terra.nasa.gov/about/terra-instruments/modis

Ugarte, M. D., Adin, A., Goicoa, T., & Militino, A. F. (2014). On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research*, *23*(6), 507–530. https://doi.org/10.1177/0962280214527528

U.S. Census Bureau. (2021). https://www.census.gov/popclock/

US Coronavirus Daily Cases by County. (2021, March 3). USAFacts.Org. https://usafactsstatic.blob.core.windows.net/public/data/covid-19/covid_confirmed_usafacts.csv

US Coronavirus Daily Deaths by County. (2021, March 3). USAFacts.Org. https://usafactsstatic.blob.core.windows.net/public/data/covid-19/covid_deaths_usafacts.csv

Usher, K., Bhullar, N., Durkin, J., Gyamfi, N., & Jackson, D. (2020). Family violence and COVID-19: Increased vulnerability and reduced options for support. *International Journal of Mental Health Nursing*.

Wan, Zhengming, Hook, Simon, & Hulley, Glynn. (2015). *MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD11A1.006

Wang, X., Yue, R. Y., & Faraway, J. J. (2018). *Bayesian Regression Modeling with INLA* (1st ed.). CRC press. https://www.routledge.com/Bayesian-Regression-Modeling-with-INLA/Wang-Ryan-Faraway/p/book/9780367572266

Woolf, S. H., Chapman, D. A., & Lee, J. H. (2021). COVID-19 as the Leading Cause of Death in the United States. *JAMA*, *325*(2), 123–124. https://doi.org/10.1001/jama.2020.24865

Wu, Q. (2020). geemap: A Python package for interactive mapping with Google Earth Engine. *Journal of Open Source Software*, *5*(51), 2305. https://doi.org/10.21105/joss.02305

Wu, Y., Jing, W., Liu, J., Ma, Q., Yuan, J., Wang, Y., Du, M., & Liu, M. (2020). Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries. *Science of The Total Environment*, *729*, 139051. https://doi.org/10.1016/j.scitotenv.2020.139051

Variable	Mean	Standard Deviation	Minimum	Maximum
Unemployed	5.74%	2.78%	0%	26.40%
Per Capita Income	$27,036.00	$6457.00	$0	$72,832.00
No High School Diploma	13.45%	6.34%	1.20%	66.30%
Age 65+	18.43%	4.54%	3.80%	55.60%
Age17-	22.35%	3.43%	7.30%	40.30%
Disabled	15.96%	4.40%	3.80%	33.70%
Single Parent	8.30%	2.71%	0%	25.60%
Non-White Status	23.15%	19.84%	0%	99.30%
Mobile Home	13.03%	9.62%	0%	59.30%
Crowded Living(>10 per hh)	2.33%	1.92%	0%	33.80%
No Vehicle	6.18%	3.57%	0%	77.00%
Group Quarters	3.48%	4.74%	0%	55.70%
Poverty	15.63%	6.46%	0%	55.10%
Multi-Unit Dwelling	4.63%	5.64%	0%	89.40%
Limited English	1.70%	2.79%	0%	30.40%
Uninsured	10.00%	4.98%	1.70%	42.40%
Daytime LST*	294.65 K	16.62 K	239.00 K	334.00 K
Nighttime LST*	292.73 K	14.61 K	240.21 K	326.94 K
(AGL) Temperature ** 2 m Above Ground Level	289.11 K	8.58 K	262.81 K	309.40 K
Specific Humidity**	.008 kg/kg	.004 kg/kg	.0015 kg/kg	.020 kg/kg
Atmospheric Pressure**	96628.92 Pa	5465.36 Pa	6669.97 Pa	102268.70 Pa
Longwave Radiation**	327.85 W/m^2	51.84 W/m^2	170.07 W/m^2	438.38 W/m^2
Shortwave Radiation**	208.13 W/m^2	72.30 W/m^2	38.76 W/m^2	359.35 W/m^2
Potential Evaporation**	.22 kg/m^2	.10 kg/m^2	-.0051 kg/m^2	.55 kg/m^2
Precipitation**	.11 kg/m^2	.008 kg/m^2	0 kg/m^2	.66 kg/m^2
Wind Speed 10 m AGL**	3.45 m/s	.69 m/s	1.08 m/s	6.93 m/s
Wind Direction 10 m AGL**	190.52°	35.00°	77.38°	335.61°

Table 1: Descriptive statistics of monthly (year 2020) averages for SVI and environmental variables utilized in the models and aggregated by U.S. county * MODIS ** NLDAS.

Variable	Mean ?	Standard Deviation ?	2.5% Credibility Interval	97.5% Credibility Interval	Effect
Intercept	-0.65	0.01	-0.66	-0.63	
Unemployed	-0.09	0.04	-0.17	-0.01	-
Per Capita Income	0.04	0.01	0.01	0.07	4
No High School Diploma	0.19	0.02	0.16	0.23	2
Age 65+	-0.02	0.01	-0.05	0	-
Age17-	0.11	0.01	0.08	0.14	1
Disabled	-0.07	0.01	-0.1	-0.04	-
Single Parent	0.03	0.01	0.01	0.05	2
Minority	0.19	0.01	0.16	0.21	2
Mobile Home	-0.01	0.01	-0.04	0.01	-
Crowded Living(>10 per hh)	-0.05	0.01	-0.07	-0.03	-
No Vehicle	-0.07	0.01	-0.1	-0.05	-
Group Quarters	0.03	0.01	0.01	0.05	3
Poverty	0.09	0.04	0.01	0.18	
Multi-Unit Dwelling	0.11	0.01	0.08	0.13	1
Limited English	-0.05	0.01	-0.08	-0.03	-
Uninsured	-0.01	0.01	-0.03	0.01	-
LST Day	0	0.01	-0.02	0.02	0
LST Night	-0.13	0.01	-0.14	-0.11	-1
AGL Temperature	-0.23	0.02	-0.27	-0.2	-
Pressure	0.12	0.01	0.09	0.14	1
Precipitation	-0.04	0.01	-0.06	-0.03	-
Wind Direction	0.11	0.01	0.09	0.13	1
Wind Speed	0.03	0.01	0.01	0.04	2

938
939 Table 2: Posterior coefficients and 95% credibility intervals of variables utilized in the model for COVID-19 cases of infection. Grayed out variables are not
940 statistically significant.

941

Variable	Mean ?	Standard Deviation ?	2.5% Credibility Interval	97.5% Credibility Interval	Effect on Risk (%)
Intercept	-1.88	0.05	-1.99	-1.79	
Unemployed	-0.06	0.07	-0.2	0.08	-5.57
Per Capita Income	0.06	0.02	0.02	0.11	6.65
No High School Diploma	0.3	0.03	0.25	0.36	35.21
Age 65+	0.21	0.02	0.17	0.26	23.93
Age17-	0.18	0.03	0.13	0.23	19.42
Disabled	-0.09	0.02	-0.13	-0.04	-8.37
Single Parent	0.07	0.02	0.03	0.11	7.05
Minority	0.34	0.02	0.3	0.38	40.19
Mobile Home	0.01	0.02	-0.03	0.05	1.09
Crowded Living(>10 per hh)	-0.09	0.02	-0.13	-0.05	-8.54
No Vehicle	-0.02	0.02	-0.05	0.01	-1.78
Group Quarters	0.03	0.02	-0.01	0.06	3.02
Poverty	0.07	0.07	-0.08	0.21	7.06
Multi-Unit Dwelling	0.11	0.02	0.07	0.15	11.29
Limited English	-0.14	0.02	-0.19	-0.1	-13.35
Uninsured	-0.01	0.02	-0.05	0.03	-0.73
LST Day	0	0.02	-0.03	0.04	0.48
LST Night	-0.12	0.01	-0.15	-0.09	-11.14
AGL Temperature	-0.31	0.03	-0.37	-0.25	-26.68
Pressure	0.14	0.02	0.1	0.17	14.66
Precipitation	-0.06	0.01	-0.08	-0.03	-5.62
Wind Direction	-0.07	0.01	-0.1	-0.04	-6.99
Wind Speed	-0.02	0.01	-0.04	0.01	-1.73

Table 3: Posterior coefficients and 95% credibility intervals of variables utilized in the model for COVID-19 deaths. Grayed out variables are not statistically significant.

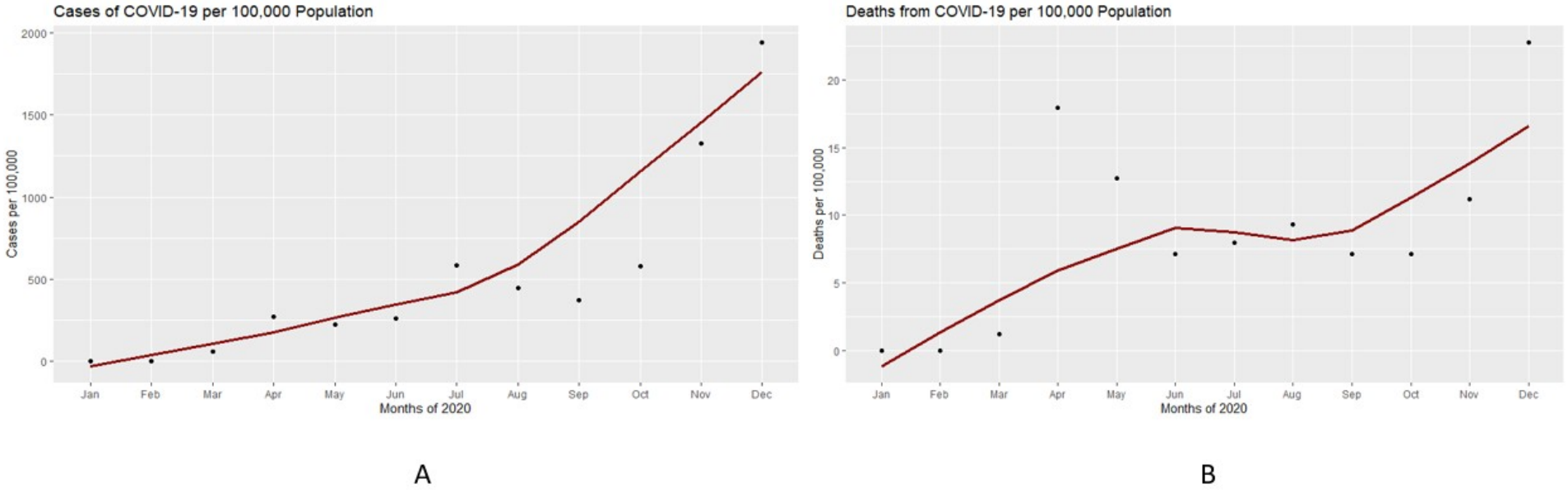


Figure 1: A. Cases of COVID-19 per 100,000 count of population in the United States for the year 2020. (Crimson line = Locally weighted scatterplot smoothing (LOWESS) line of cases). B. Deaths from COVID-19 per 100,000 count of population in the United States for the year 2020. (Crimson line = Locally weighted scatterplot smoothing (LOWESS) line of deaths)

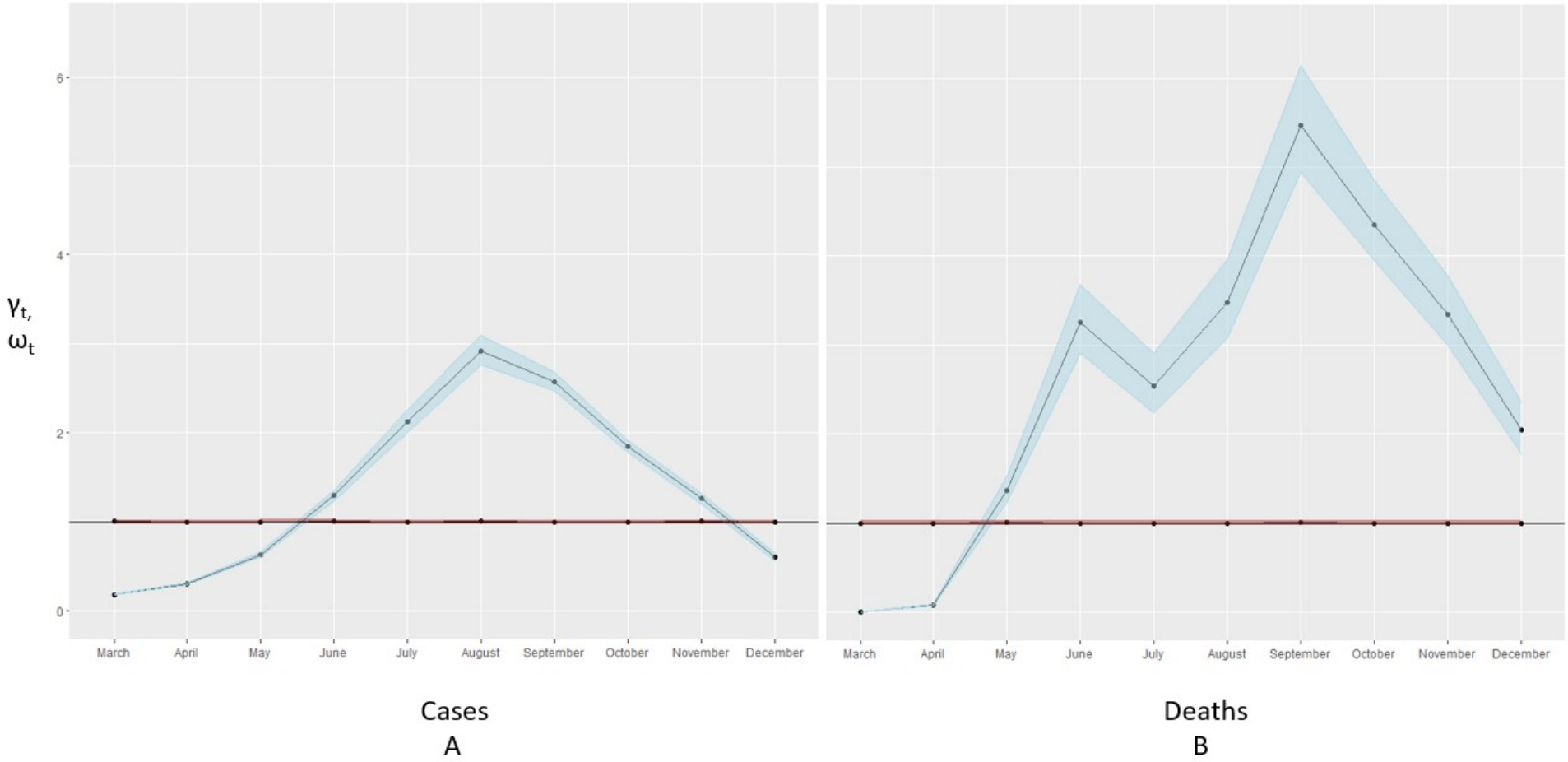
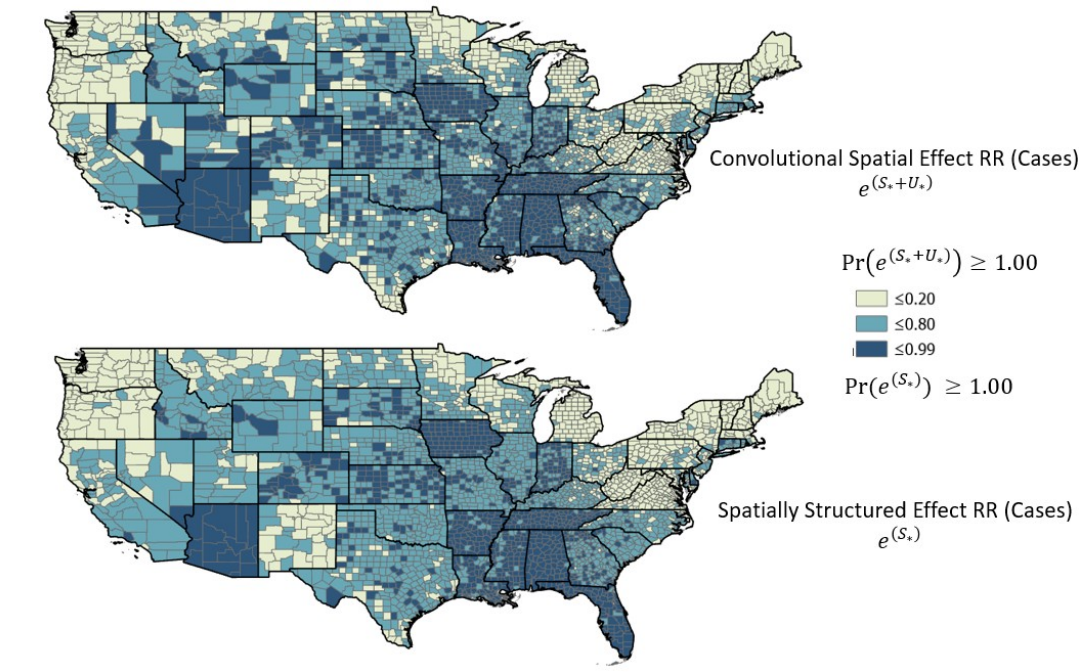
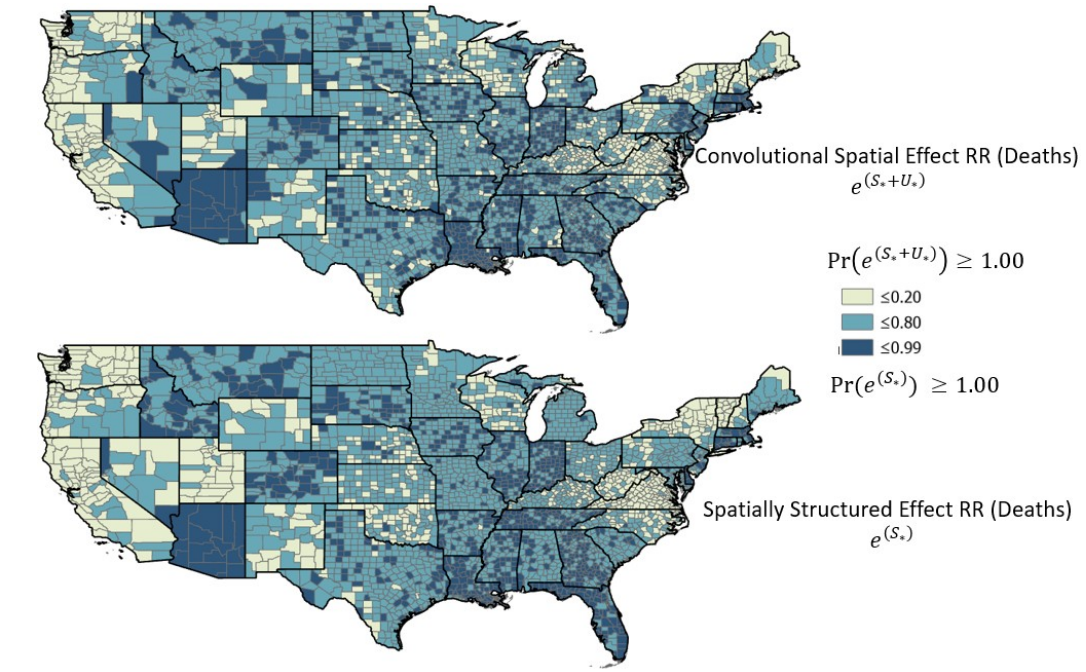


Figure 2: Posterior temporally structured; γ_t (light blue) and unstructured; ω_t (light red) effects for COVID-19 cases; panel A, and deaths; panel B (95% credibility envelope) in the U.S. by month for the year 2020.

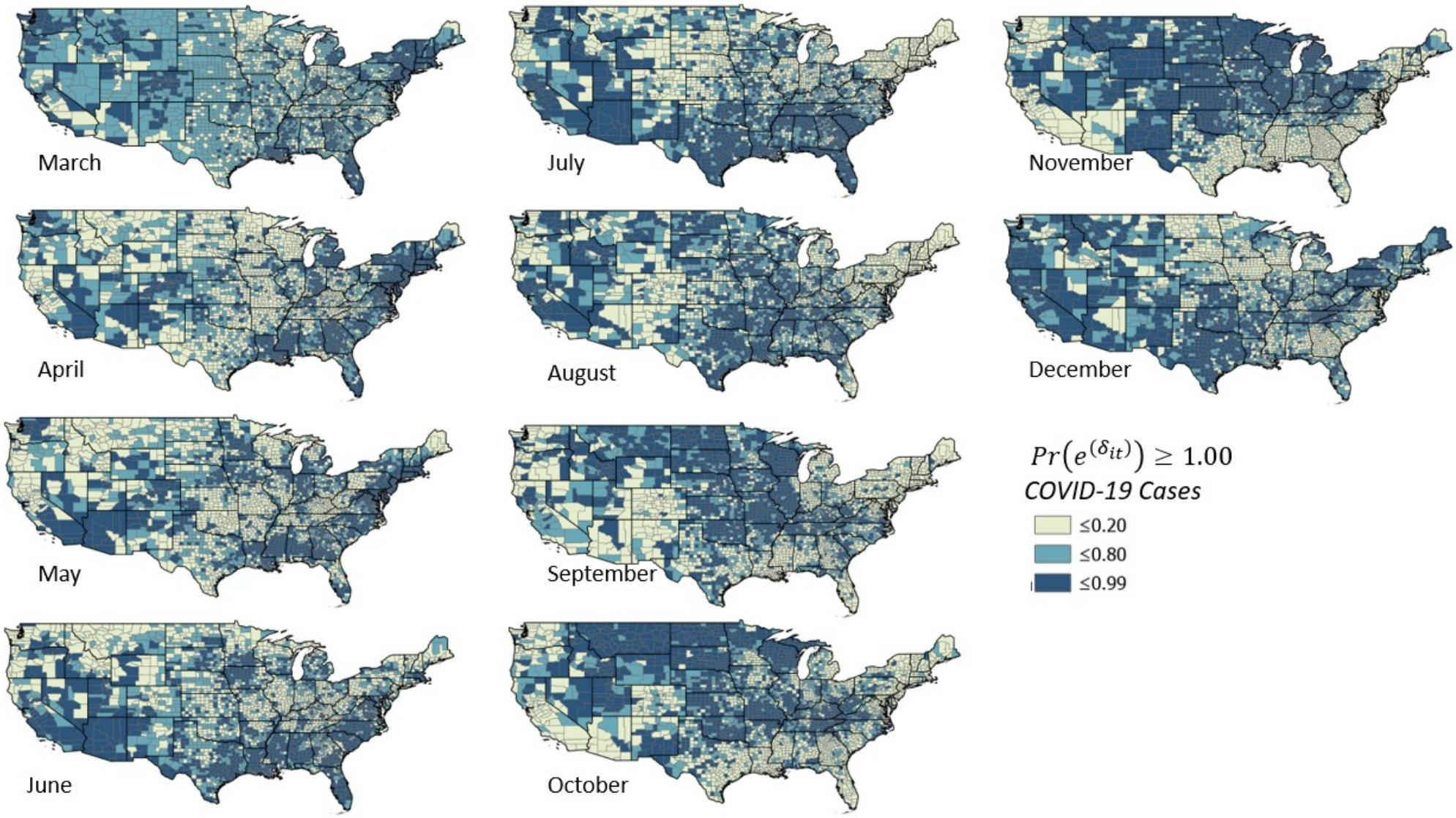


A



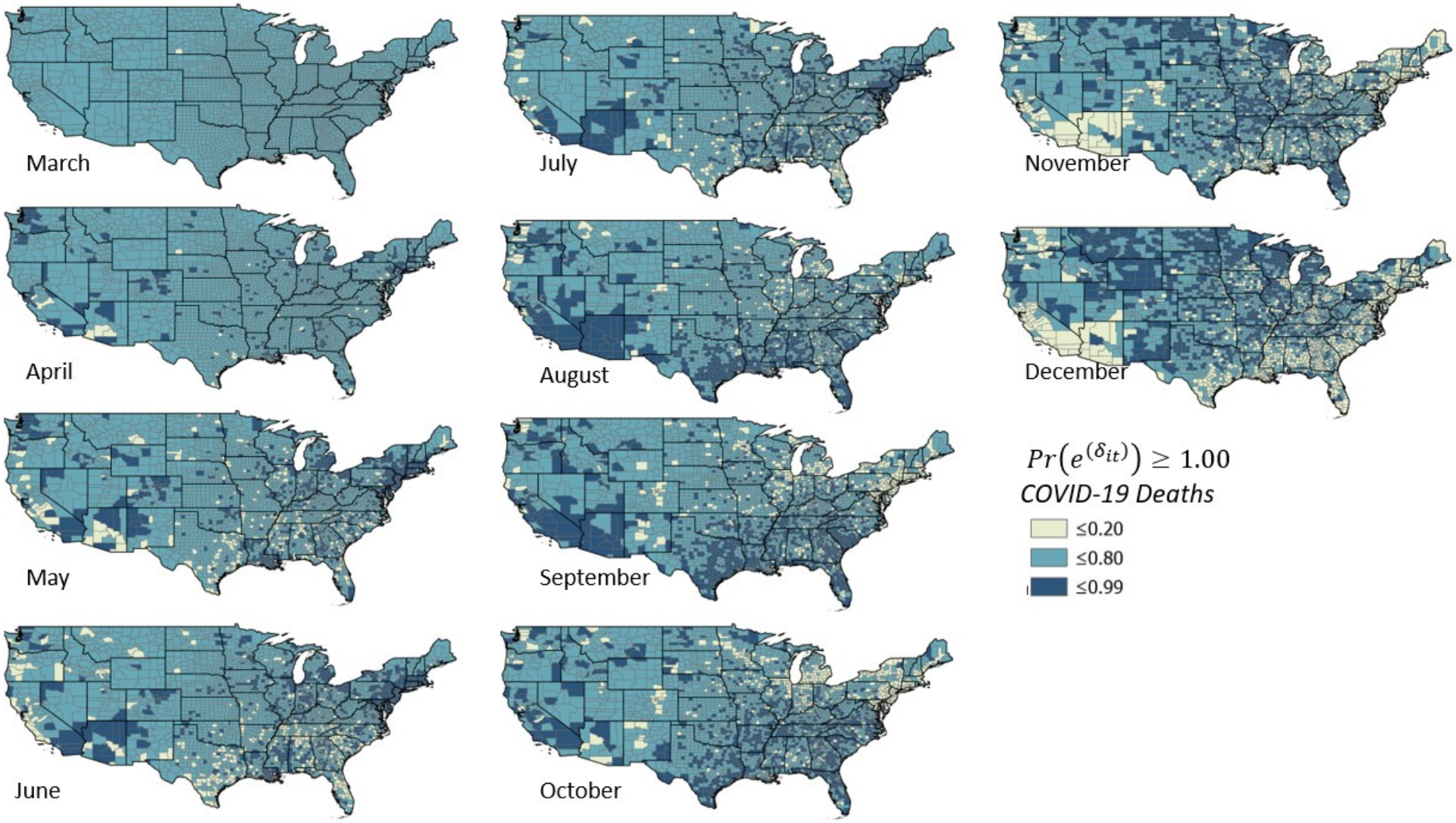
B

Figure 3: Exceedance probabilities of convolutional spatial effect ($e^{(S_i+U_i)}$) and spatially structured effect ($e^{(S_i)}$) relative risk associated with COVID-19 infection; panel A, and COVID-19 mortality; panel B.



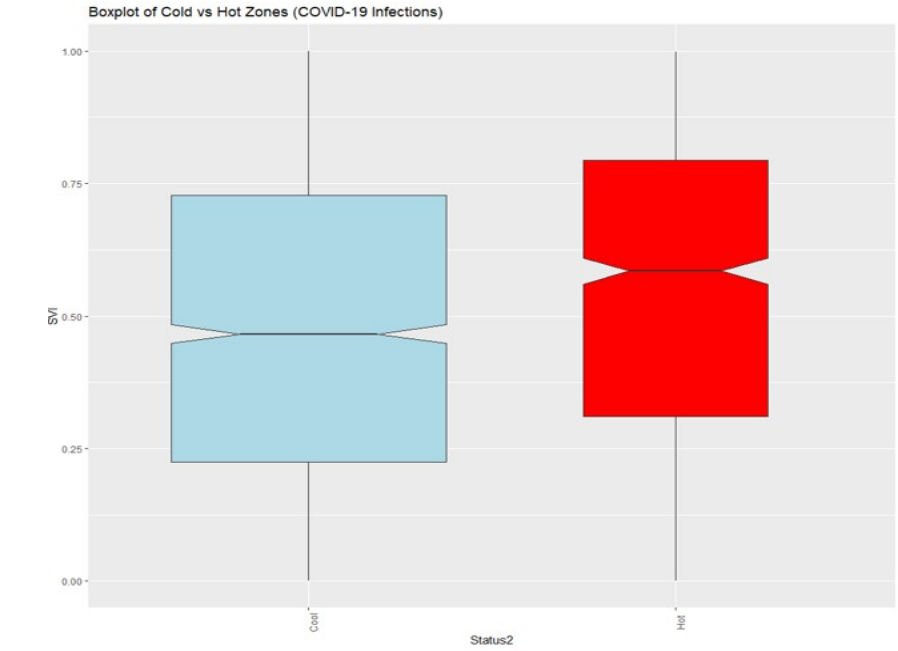
965

966 Figure 4: Probabilities of space-time interaction term $Pr\left(e^{\delta_{it}}\right)$ exceeding 1.00 during the year 2020 for U.S. COVID-19 infections stratified monthly.

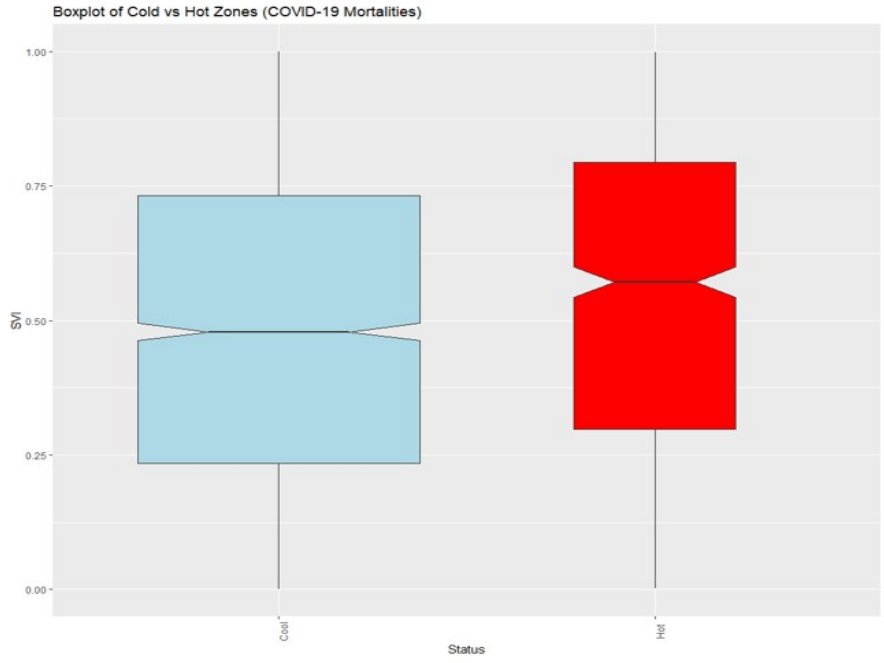


967

968 Figure 5: Probabilities of space-time interaction term $Pr\left(e^{\delta_{it}}\right)$ exceeding 1.00 during the year 2020 for U.S. COVID-19 mortalities stratified monthly.



A



B

Figure 6: Boxplot of SVI composite score for hot and cold spots for the convolutional spatial effect for COVID-19 infections; plot A, COVID-19 fatalities; plot B.

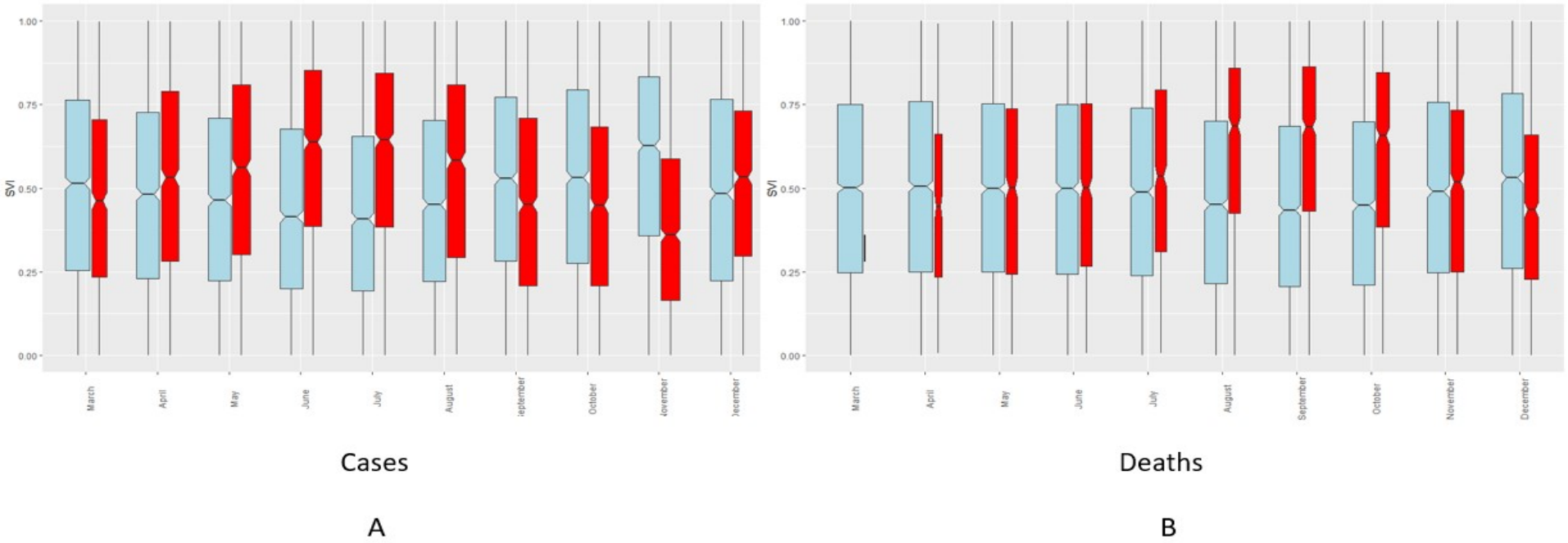


Figure 7: Boxplots organized by month, comparing cold spots and hot spots of COVID-19 cases (A) and deaths (B) to composite SVI score.