

Understanding spatiotemporal variation of social vulnerabilities from longitudinal hurricane-pandemic data: A multilevel model of the Covid-19 pandemic during hurricane Sally in Florida

Seyedreza Abazari^a, O. Arda Vanli^{a, *}, Onur Alisan^b, Eren Erman Ozguven^b

^a Department of Industrial and Manufacturing Engineering, Florida A&M University – Florida State University College of Engineering, Tallahassee, FL, 32310, USA

^b Department of Civil and Environmental Engineering, Florida A&M University – Florida State University College of Engineering, Tallahassee, FL, 32310, USA

ARTICLE INFO

Keywords:

Multilevel modeling
Longitudinal data
Social vulnerability
Compound hazards

ABSTRACT

An important question in the context of compound disasters is the degree to which geophysical disasters amplify the transmission of infectious diseases during pandemics and how this relationship is influenced by the social vulnerability of affected populations. This article proposes a spatiotemporal modeling approach to understand spatially varying social, demographic and health drivers of vulnerability during pandemics co-occurring with geophysical hazards. A multilevel mixed-effects model is developed to investigate the dynamic association between census tract-level Covid-19 case count trajectories co-occurring with a hurricane and demographic, socioeconomic and health factors. A state-level analysis is conducted to identify the distinct geographical regions in which significant changes are seen in the infection count trends due to the hurricane. A subsequent region-level analysis is performed to describe, at a higher spatial resolution, the impact of social vulnerability on the infection count trajectories at a community level. The method provides an approach to systematically study the effects of compound hazards and distinct patterns of infectious disease spread during hurricanes by quantifying (1) dynamic associations between infection counts and social factors and (2) spatial heterogeneities of these associations between communities. A case study for modeling the spatiotemporal variation of social vulnerability with data from Covid-19 pandemic and Hurricane Sally in Florida is presented to illustrate the application of the approach.

1. Introduction

Large-scale disasters, in particular, concurrently occurring geophysical hazards and pandemics, disproportionately affect disadvantaged communities. The coronavirus disease 2019 (Covid-19/SARS-CoV-2) pandemic, which caused more than 97 million cases and 1,072,582 deaths in the United States as of November 6, 2022 [1], has highlighted systemic social and health inequities that disproportionately affected racial minority and underserved groups in the U.S. and the World. It is shown that natural disasters increase the spread of infectious diseases during pandemics. For example, after Hurricane Katrina in 2005, a norovirus outbreak affected over 1000 evacuees residing in the Astrodome megashelter [2]; after the 2011 Japan earthquake and tsunami, overcrowded shelters caused high incidence of respiratory infections [3]; and the Haiti earthquake in 2010 was followed by a cholera outbreak resulting in

* Corresponding author.

E-mail address: oavanli@eng.fsu.edu (O.A. Vanli).

more than 9000 deaths and 650,000 infections [4]. When a pandemic concurrently occurs with a natural disaster, disaster response and support agencies should carefully evaluate the complex health effects of disasters to ensure a safe and successful delivery of evacuation and sheltering services to vulnerable sub-populations such as the elderly, children, and persons with disabilities [5]. To make equitable decisions on allocating health and emergency response resources, caring for vulnerable sub-populations, and implementing local and state-level interventions, access to dynamic pandemic and disaster situation data and data on key vulnerabilities at the community level are essential [6].

The current paper proposes a new multilevel mixed-effects modeling methodology to understand the associations between the census-tract infection-count trajectories observed during a pandemic a co-occurring with a disruptive hurricane and the demographic and socioeconomic characteristics of communities. Mixed-effects models are particularly well-suited for this specific application, with their flexibility in modeling spatial heterogeneities in dynamic associations between the infection-count time trajectories and social factors. Based on the proposed mixed-effects models, we propose a new heterogeneity metric that measures the spatial variations in Covid-19 case rate accelerations between census tracts following a hurricane and show its practical implications for quantifying local impacts of compound disasters. In a case study we demonstrate the application of the proposed method used to assess the association between social vulnerability and the pandemic spread by using publicly available spatial social vulnerability data and daily infection case counts collected during the Covid-19 pandemic co-occurring with a hurricane in Florida.

The proposed approach makes unique contributions to the literature as follows. A critical shortcoming of the existing pandemic vulnerability indicators, which utilize (total) infection counts at a particular point in time, is that they lack the ability to include the temporal dimension in vulnerability analysis. The proposed modeling method enables the study of *dynamic* associations between the temporally observed infection counts and social characteristics of communities and provides a unique ability to investigate the social drivers and patterns of infectious disease spread during pandemics with a particular focus on the co-occurrence of geophysical hazards. Secondly, the mixed-effects model can estimate spatially varying model coefficients which provides the ability to make statistical inferences about the *spatial heterogeneities* of such associations. The abilities to add the temporal dimension in the study of vulnerabilities and to make statistical inferences about spatial heterogeneities are crucial in the understanding of local drivers of vulnerability during compound hazards. For example, it enables to test whether the variation in the speed of the disease progression in different communities is significant and how this variation is related to the social vulnerabilities and the co-occurrence of geophysical hazards. The ability to quantify the vulnerabilities of communities against compound hazards, e.g., hurricane-hits during pandemics, in turn, will be helpful for policy and decision-makers in emergency management and healthcare operations planning to make effective resource allocation and disaster mitigation decisions.

2. Relevant literature

In the event of disasters of all types, including pandemics like Covid-19, socioeconomic factors like the percentage of people living in poverty or without access to transportation can contribute to the degree to which communities withstand disaster impacts and exacerbate human suffering and financial loss. The knowledge of factors that describe a community's social vulnerability against disasters is crucial to enable the identification of populations most likely to need support before, during, and after a disaster and in determining how resources might be allocated for mitigation and response. For example, lack of access to health resources or facilities may be a cause of increased vulnerabilities to hurricane-pandemics in a certain place. Vulnerability is defined as the potential to suffer losses and influences the way people "anticipate, cope with, resist, and recover from the impact" of an event [7]. One of the earliest definitions of social vulnerability using indices was proposed by Cutter et al. [8], to describe the relative social vulnerability of communities to natural hazards.

Factors affecting community vulnerability have been well studied in the context of geophysical disasters like floods and hurricanes and in the context of pandemics. For pandemics, Flanagan et al. [9], working with the U.S. Centers for Disease Control, developed a social vulnerability index, termed CDC SVI. CDC SVI estimates the relative vulnerability of each census tract in the U.S. by ranking 15 variables grouped across four dimensions: The socioeconomic theme that comprises percentage below poverty, percentage unemployed, per capita income, and percentage with no high school diploma; the household composition theme that comprises the percentage aged ≥ 65 years, percentage aged ≤ 17 years, percentage civilian with a disability, and percentage single-parent households; the minority status theme that comprises percentage minority and the percentage who speak English less than well; and the housing and transportation theme that constitutes percentage multiunit structures, percentage mobile homes, percentage crowding, percentage no vehicle, and percentage group quarters. Karaye and Horney [10] used CDC SVI to study Covid-19 prevalence. More recently, Marvel et al. [11], developed a vulnerability index, labelled COVID PVI, incorporating Covid-19 specific covariates including, diabetes, obesity, hospital intensive care unit (ICU) beds per capita, and health insurance coverage. Rahman et al., [12] explored the impact of various demographic, socioeconomic, mobility and health related variables on the incidence of COVID-19 cases and investigated a static pandemic vulnerability index at the city level (PVI-CI), at a specific point in time, to classify cities into five vulnerability classes, ranging from very high to very low. Obesity was identified as a risk factor for progression during the 2009 H1N1 influenza epidemic and with Covid-19 in the U.S. by researchers [13]. Diabetes has been found to be a risk factor in early studies of Covid-19 [14]. Fewer ICU beds were found to be positively associated with increased cases and deaths; health insurance coverage, more prevalent in rural counties, was found to be positively associated with increased cases and deaths [15]. Long-term exposure to particulate matter pollution was shown to correlate with Covid-19 prevalence [16,17]. Other studies focusing on socio-economic determinants of the prevalence of Covid-19 include [18,19].

The most basic and well-studied model in epidemiology is the Susceptible-Infected-Removed (SIR) compartmental model that describes the dynamics of an epidemic in which an individual is in one of three disjoint compartments at any given time: susceptible to

disease, infected and recovered [20]. The compartmental models are based on systems of differential equations for the movement of the population through discrete states, including entry into and exit from the population. The SIR models have been successfully applied to estimate the infection and recovery rates from observed data and assess the effectiveness of various interventions, such as school closures and mask wearing, on pandemic spread [21]. However, it is usually difficult to accommodate spatial heterogeneity and risk factor covariates in such models. A more natural and flexible way to model the time trajectory of infectious disease counts is to model them from growth curve or longitudinal data, which consists of repeated measurements of some characteristic obtained from a single subject or geographical unit (e.g., census tract) at different occasions or times (e.g., infection counts in consecutive time periods). Longitudinal data have a temporal order, due to this ordering observations for different geographical units are clustered and data within a cluster will typically exhibit positive correlation which must be accounted for in the analysis. Mixed-effects models are powerful modeling tools that allow for the analysis of infection data with complex associations with socioeconomic covariates, hierarchical structures [22,23], in particular, infectious disease data, to account for unobserved spatial heterogeneity by means of random effects. The hierarchical data sets of interest often consist multiple levels of nesting, for example, longitudinal infection counts from census tracts, nested under regions of census tracts, and regions nested under a state. Fritz and Kauermann [24] used region-specific random effects in their Covid-19 spread prediction models to account for unobserved heterogeneities due to super spreader events or local outbreaks in meat processing plants.

Time trajectories of epidemic curves often exhibit nonlinearities or inflection points that require the use of flexible, non-parametric functions for modeling. Boschi et al. [25], investigated the patterns of Covid-19 mortality curves in Italy and their association with mobility, positivity, and socio-demographic, infrastructural and environmental covariates by using functional data analysis techniques. Logistic regression with a smooth temporal effect was used to identify adequate risk factors for severe Covid-19 cases in a matched case-control study in Scotland [26]. Jiang et al. [27], proposed modeling the Covid 19 infection trajectories using a piecewise linear quantile trend model to naturally capture the phase transitions of the epidemic growth rate via change-points. Basellini and Camarda [28] investigated the regional differences in the epidemic mortality time trends during the first Covid 19 wave in Italy and associated risk factors through regression splines and random effects that account for regional heterogeneities. Approaches based on recurrent neural networks from the field of deep learning, specifically Long Short-Term Memory (LSTM) neural networks, have been utilized for analyzing and forecasting the dynamics of Covid-19. Nikparvar et al. [29], demonstrated the efficacy of a multivariate LSTM network incorporating mobility data, resulting in improved accuracy in forecasting spatial patterns of Covid-19 spread. Shahid et al. [30], highlighted the superiority of bidirectional LSTM in forecasting Covid-19 cases, deaths, and recoveries compared to time series and support vector machine methods. While the primary focus of LSTM models is modeling complex temporal dependencies in sequential data, their main drawback is that they are not specifically tailored to handle spatial dependencies or heterogeneities. Compared to the approach proposed in the current paper, from the modeling perspective the approach of [28] is quite similar in the sense that both approaches consider modeling epidemic curves to study their associations with socioeconomic risk factors. However, the latter does not consider the effect of another disruptive event, such as a hurricane, which our approach account for explicitly.

The relationship between explanatory variables and an outcome variable in statistical modeling is not always constant across a study area. Variations in relationships over space are referred to as spatial heterogeneity (LeSage and Pace [31], p.29) and spatial non-stationarity (Fotheringham et al. [32], p. 9). Two competing models to address this issue are geographically weighted regression or GWR [32] and multilevel mixed-effects (random coefficient) models [22] with spatially varying coefficient processes (SVCP). While GWR assumes local relationships vary smoothly, the SVCP allows coefficients to vary randomly ([32], p.20). GWR and spatial random effects models are considered to be the two main approaches to estimate spatially varying associations between outcomes and covariates. However, mixed-effects models have several advantages over GWR [33]. Unlike GWR, in mixed-effects models, the data do not need to be independent to define inference; the model can explicitly incorporate spatial correlation (or dependence) in the observations and associations through the spatial random effect (prior) distributions. To account for possible spatial dependence among adjacent regions, an error term following a spatial autoregressive (SAR) model and an adjacency matrix can be incorporated in the mixed-effects models (Waller and Gotway [34], pp.363). As an extension of SAR models, which focus solely on the spatial dependence aspect, spatial panel data models have been developed to analyze not only the spatial dependence but also the time dependence in the data by incorporating temporally lagged covariates and error terms in the model (Ehorst, [35]). The basic reference for spatial correlation structures used with linear models with no random effects is Cressie [36] and spatial correlation structures in the context of mixed-effects models are described at length in Diggle et al. [37]. Note that in this study our model formulation accounts for spatial heterogeneity but not spatial dependency. However, framework is general and can be extended to incorporate spatial correlation, through for example a SAR structure [34]. In addition, mixed-effects models provide the ability to conduct model-based estimation and inference on whether the spatial heterogeneity due to some factor is significant, while GWR only allows relatively ad hoc inference.

Spatial differences or heterogeneities in community vulnerability have been previously studied within the context of disasters. The focus was limited to studying the impacts of a single disaster, such as, geophysical disasters [8] or pandemics [9,38], separately. However, research studying heterogeneities in vulnerabilities against compound disasters is limited. Karaye and Horney [10] studied the variables associated with increased cumulative Covid-19 case counts from the US county-level case count data. To accommodate the fact that vulnerability to Covid-19 may actually be nonstationary and vary between U.S. counties, a GWR model was constructed using social vulnerability indicators. Another regression model with spatially varying coefficients to model cumulative case counts at the county-level was presented by Snyder and Parks [39] with a wider breadth of Covid-19 specific covariates including, diabetes, obesity, hospital beds per capita, and health insurance coverage. A multilevel Bayesian local regression approach presented by Song et al. [40], aimed at modeling spatial heterogeneities of vulnerability against Covid-19.

3. Study region, time period and data set: the first Covid-19 wave in Florida

To motivate the proposed methodology, we describe the region and time period of focus in our study and the data set that will be used in the illustrations of the paper. We consider data from the first wave of the Covid-19 pandemic, covering July 22 to October 17, 2020 as the study period, and focusing on the state of Florida. Covid-19 was declared by the World Health Organization as a pandemic on March 11, 2020 [41]. The study time period coincides with the first phase of the Covid-19 pandemic during which no vaccines were yet available (the first vaccines only became publicly available starting in December 2020). In addition, during this time period a concurrent geophysical hazard, hurricane Sally impacted the study region, which made landfall in Alabama and Pensacola, FL as a category 3 hurricane on Sep 16, 2020 [42].

The study period covers 88 days, and the study region of the state of Florida covers more than 4200 census tracts, which makes a mixed-effects modeling of the Covid-19 infection curves at the census-tract level for the entire state computationally intractable. To ease the computational burden, Florida census tracts are grouped into 7 non-overlapping regions, as shown in Fig. 1, and the analysis is conducted for the regions separately. The regions are constructed, as an aggregation of between 300 and 900 contiguous census tracts, following the Florida Department of Health, health care coalitions district definitions [43] and numbered according to the proximity to the hurricane landfall location. Fig. 1 shows the counties of the regions and the Hurricane Sally's 5-day ahead forecast issued by National Hurricane Center (NHC) on Sep 15, 2020. Table 1 provides the number of census tracts and the respective populations of the regions.

Confirmed daily Covid-19 case count data for Florida zip codes are accessed at Florida Covid-19 Hub [44]. The United States Census Bureau provides zip code tabulation area (ZCTA) files to convert mail route-defined zip codes into areal units. The zip codes are matched with census tracts using the ZCTA to Census Tract Relationship File [45]. The Covid-19 case counts at the census tract level are obtained as a population-weighted average of the zip code level case counts using populations of the zip codes as weights. The census tract populations required to find the daily infection rates (i.e., daily infection count divided by population) were obtained from 2020 U.S. Census [46], which aligns with the infection count timeframe. The data and documentation on demographic and socioeconomic variables were accessed at the Centers for Disease Control and Prevention Social Vulnerability Index (SVI) web site [47,48]. In addition, the number of hospital beds per 1000 people, air pollution, prevalence of diabetes and obesity and percent uninsured, that were associated with worsening prognosis of Covid-19 are considered as health-related variables [11,39]. The health-related variables measured at the county level were accessed at the COVID Pandemic Vulnerability Index (PVI) public repository [49]. Table 2 shows the definitions and data sources of the 21 variables in addition to the infection case counts used in the study. All data processing and modeling have been conducted¹ in R statistical computing language [50].

4. Proposed method: mixed-effect model for infection count trajectories of geographic regions

In this research we develop a statistical modeling methodology to study the spatially varying relation between the growth of census tract-level infectious disease counts over time and the demographic, socioeconomic and health characteristics of the communities during a pandemic that is concurrently occurring with a hurricane. As shown in Fig. 2 from each census tract daily Covid-19 infection counts are observed. The census tracts are nested under regions and regions are nested under the study region. The nested structure of the data requires a multi-level model be constructed. Since the lowest spatial level of modeling is census tracts, the terms community and census tract will be used interchangeably in this paper. An application to the first Covid-19 wave in Florida and hurricane Sally is presented. A multilevel linear mixed-effects model is constructed to study the statistical dependence of the changes in the infection rates over time and how these changes are associated with the community characteristics and the timing of the hurricane. Using the proposed approach, we show how to determine the demographic, socioeconomic and health variables that are most influential on the significant trends of the infection rate trajectories and to identify the hot spots where community vulnerability and the occurrence of the hurricane were positively associated with increased case rates.

A flexible regression-spline smoothing is used to construct the temporal structure of the model and account for the possible changes in the infection rates of census-tracts that are impacted by the hurricane. The covariates of the mixed-effects model are selected by variable selection from a pool of potentially important demographic, socioeconomic and health factors. Demographic and socioeconomic factors include those that impact the community's capacity to respond to Covid-19 and hurricanes, such as, education, race, age, income, prevalence of disabilities and language, all measured at the census tract level [9]. Health factors include pre-existing conditions and behaviors that are thought to increase disease incidences such as smoking prevalence, diabetes, obesity, health insurance coverage, number of hospital beds per capita, and air pollution.

For each census tract, the outcome variable is determined by dividing the daily counts of confirmed Covid-19 cases by the total population of the respective census tract. To satisfy the parametric requirement of normal probability distribution of linear models, we log-transformed the outcome variable and exponentiated the model coefficients for ease of interpretation. Accordingly, the outcome variable is defined as $y_{ij} = \log((z_{ij} + 1)/m_i)$ where z_{ij} is the confirmed infection count in census-tract i at the j -th measurement occasion and m_i is the total population of census tract i . The outcome variable of interest y_{ij} is, therefore, a log infection rate (with 1 added to each to avoid taking the logarithm of zeros for zero count census tracts). By dividing the counts by the population at risk, the counts from different census tracts are made more comparable, since the sizes of the population at risk differ in different census tracts. Furthermore, the normal approximation is more appropriate with the log rates than applied to counts ([34], p. 347).

¹ The R codes and data of the study can be obtained at <https://github.com/avanli/LongitudinalCovid19MixedModel>.

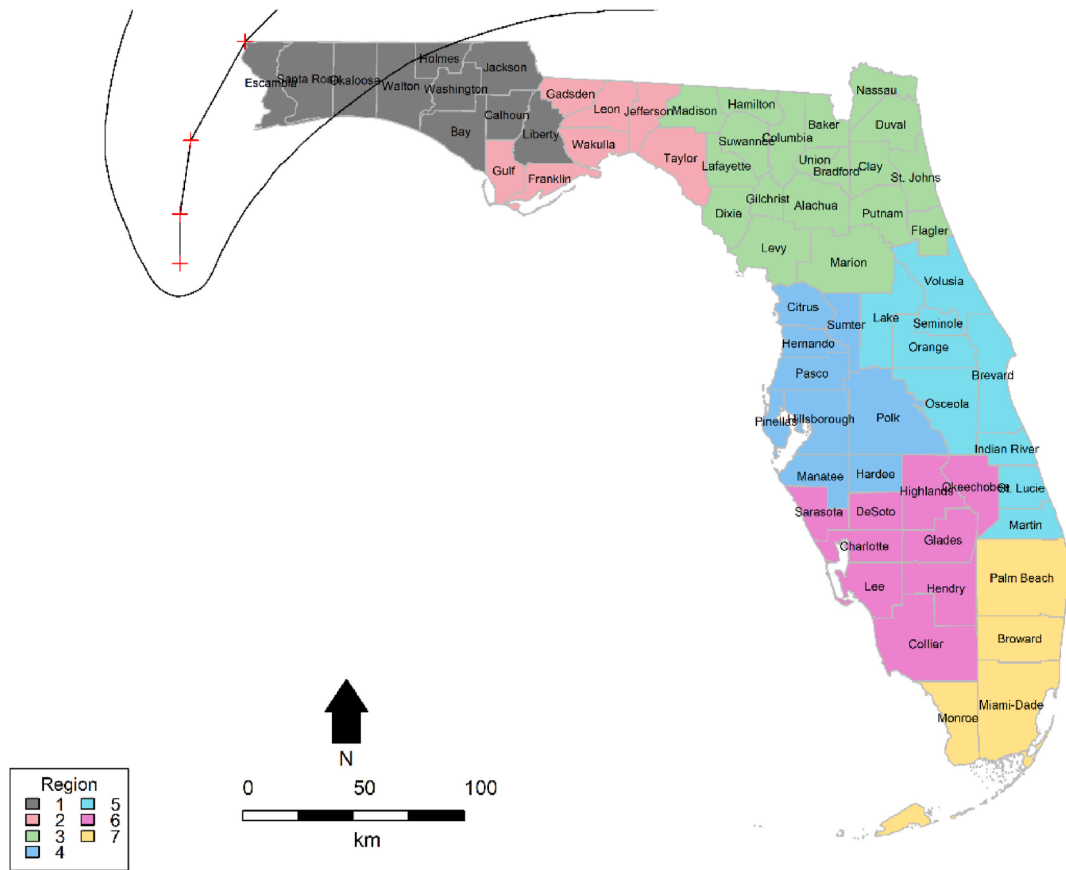


Fig. 1. Florida Department of Health regions [43] showing the counties and the 5-day ahead forecast and uncertainty cone for Sep 15, 2020, 1:00pm advisory [42].

Table 1

Populations and numbers of census tracts of the Florida regions.

Region	1	2	3	4	5	6	7
Population	922,106	386,941	2,287,758	3,891,048	3,660,723	1,672,241	5,558,784
Number of census tracts (N)	218	95	464	1,020	719	425	1,235

The proposed approach arrives at the final multi-level model of each region in two stages. In the first stage, the method identifies the optimal temporal structure of the mixed-effects models for all regions shown in Fig. 1 to describe the temporal changes of the Covid-19 case counts after the hurricane landfall. Based on the temporal trends of the regions, the method also identifies the regions that are most significantly impacted by the occurrence of the hurricane. In the second stage, the method focuses on the most impacted regions and develop more detailed mixed-effects models that contains both temporal trends and the spatially varying effects of demographic, socioeconomic and health covariates on the census tract-level infection count trajectories. Both the temporal structure and the covariates of the model were selected using the Akaike Information Criterion. To demonstrate the spatial heterogeneities in how the covariates predict Covid-19 case counts across the census tracts, coefficient choropleth maps were generated for the random coefficients of the predictor variables using hot-cold rendering color schemes.

A multi-level mixed-effects model summarizes the coefficients in terms of two parts: a “fixed” part which is common across census tracts, and a “random” part, which is allowed to vary from tract-to-tract. A general formula for the model used to describe the changes in the infection rate y_{ij} over time periods t_{ij} for census-tract i and measurement occasion j , to model impacts of compound hurricane-pandemic events is

$$y_{ij} = \theta(t_{ij}) + \gamma(x_i) + f_i(t_{ij}) + g_i(x_i) + \epsilon_{ij}, j = 1, \dots, n, i = 1, 2, \dots, N \quad (1)$$

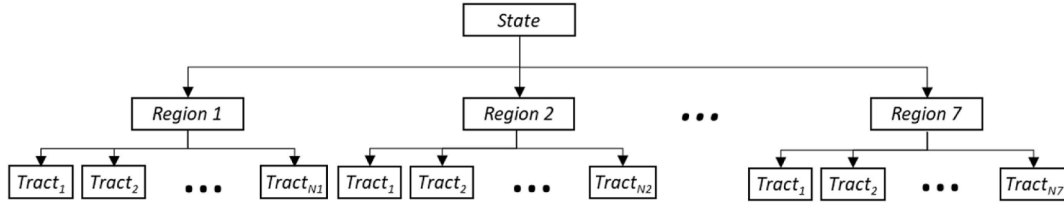
where N is the number of census tracts within the region of interest, n is the number of measurements and $x_i = (x_{i1}, \dots, x_{ip})$ is the collection of p (non-time varying) covariates that comprise the demographic, socioeconomic, and health related variables for census-tract i . Table 1 lists N and populations of different regions and for the study period $n = 88$ days for all regions and census tracts.

The fixed-effects part of the model, $\theta(\cdot) + \gamma(\cdot)$, models the predicted Covid-19 infection rate trajectory for the entire region, while the random-effects part of the model, $f_i(\cdot) + g_i(\cdot)$, models the deviation over time of i -th census tract trajectory from the region trajec-

Table 2

Demographic, socioeconomic and health variables. Data sources are CDC [47] and COVID19 PVI [49].

Name	Source	Description
AirPollution	COVID19PVI	Particulate matter
Diabetes	COVID19PVI	Diabetes prevalence
Obesity	COVID19PVI	Adult obesity prevalence
Smoking	COVID19PVI	Adult smoking prevalence
PctNoIns	COVID19PVI	Percentage of persons with no health insurance
PctBeds	COVID19PVI	Number of hospital beds per population (per 1000 persons)
EP_POV	CDC	Percentage of persons below poverty
EP_UNEMP	CDC	Percentage of civilians (age > 16) unemployed
EP_PCI	CDC	Per capita income
EP_NOHSDP	CDC	Percentage of persons with no high school diploma (age > 25)
EP_AGE65	CDC	Percentage of persons aged 65 and older
EP_AGE17	CDC	Percentage of persons aged 17 and younger
EP_DISABL	CDC	Percentage of population with a disability
EP_SNGPNT	CDC	Percentage of single parent households with children under 18
EP_MINRTY	CDC	Percentage minority (all persons except white, non-Hispanic)
EP_LIMENG	CDC	Percentage of persons (age > 5) who speak English less than well
EP_MUNIT	CDC	Percentage of housing in structures with 10 or more units
EP_MOBILE	CDC	Percentage of mobile homes
EP_CROWD	CDC	Percentage of housing units with more people than rooms
EP_NOVEH	CDC	Percentage of households with no vehicle available
EP_GROUPQ	CDC	Percentage of persons in institutionalized group quarters

**Fig. 2.** Nested nature of the dataset used in the study.

tory and ϵ_{ij} is the random modeling error which describes the within-census-tract deviations. From Fig. 2, a region-level trajectory is the mean of the trajectories of all the census-tract-level trajectories below it, and the state-level trajectory is the mean of all region-level trajectories. The functions $\theta(\cdot)$ and $f_i(\cdot)$ model the association between the changes in the infection rates and time and constitute the “temporal structure” of the model. By contrast, the functions $\gamma(\cdot)$ and $g_i(\cdot)$ model the association between the changes in the infection rates and the covariates and constitute the “covariate structure” of the model. In what follows we will describe our methodology to determine the specific functional definitions of these two structures.

4.1. Selecting the optimal temporal structure by modeling the data of entire study region

In this section we discuss the proposed methodology to determine, from the observed infection count data, the best representation of the fixed-effect term $\theta(\cdot)$ and random-effect term $f_i(\cdot)$ that explain the temporal trends. In this first stage of the analysis no covariates are included in the model (i.e., $\gamma(\cdot) = 0$ and $g_i(\cdot) = 0$). The temporal trajectories of the daily infection count of census tracts are highly non-linear functions of time, in particular, at the early stages of a pandemic where the community spread is very fast. In order to flexibly describe such nonlinear trends, regression-splines temporal smoothing [51] is used to represent the temporal trends

$$y_{ij} = \underbrace{\beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 Q(t_{ij})}_{\theta(t_{ij})} + \underbrace{b_{1i} + b_{2i} t_{ij} + b_{3i} L(t_{ij})}_{f_i(t_{ij})} + \epsilon_{ij} \quad (2)$$

where κ is the breakpoint or knot, set equal to the known day of the hurricane landfall. The linear spline function $L(t_{ij}) = (t_{ij} - \kappa)_+$ defines the linear spline function [51]. The “+” subscript denotes the positive part; that is, for any number u , $u_+ = u$ if u is positive and equals zero otherwise. The quadratic spline function $Q(t_{ij}) = (t_{ij} - \kappa)_+^2$ is defined in a similar way: that is, for any number u , $u_+^2 = u^2$ if u is positive and equals zero otherwise. The spline functions $Q(t_{ij})$ and $L(t_{ij})$ are the unique aspect of our modeling methodology to account for potential changes in the infection rates due to the hurricane landfall depending on the geographic location.

The mixed-effects model (2) consists of the fixed effects $(\beta_1, \beta_2, \beta_3, \beta_4)$ and the random effects (b_{1i}, b_{2i}, b_{3i}) . The fixed effects describe the global patterns (for the entire region) of change in the mean infection rates. The random effects describe how the temporal trend of the i -th census-tract deviates from the region trend, thus accounts for spatial heterogeneities. The random effects, modeled using appropriate probability distribution functions, allow each spatial unit to follow a different time trend. The intercept β_1 is the mean infection rate at time 0, β_2 and β_3 are the linear and quadratic increases per day in the mean infection rate. β_4 is the additional change

(increase or decrease) in the quadratic time trend incurred following the hurricane landfall at time κ (owing to the fact that $Q(\cdot)$ is non-zero only after the hurricane landfall). That is, it measures potential accelerations or decelerations in Covid-19 case rates incurred in the region after the hurricane. The random effects b_{1i} and b_{2i} allow the intercept and the linear trend, respectively, to vary randomly from one census tract to another. Similarly, the random slope b_{3i} of the linear spline measures the additional change (increase or decrease) in the linear time trend, that is, potential acceleration or deceleration in Covid-19 case rates, incurred in each census tract following the hurricane. In summary, the spline terms of the model provide a unique ability to account for possible effects of the hurricane on the infection rates both at the region level and locally: β_4 measures the association of the hurricane event with increased Covid-19 cases at the region-level and b_{3i} measures the association of the hurricane with increased Covid-19 counts to vary between census-tracts. In this sense, our approach provides a framework to treat disrupting events similar to intervention measures, such as school closures, in the SIR epidemiology literature. However, the mixed effect model has the added benefit of the ability to account for spatial heterogeneities by incorporating random effects.

We remark that the temporal trend of the census-tract infection rates may be better described by a more complex model than (2). To be prepared for this possibility and select the most appropriate model, we adopt a model selection approach that evaluated the fit of various potential spline-regression models for the random-effects structures (with the same quadratic-spline for the fixed-effects part). For assessing the quality of the fit of a linear mixed-effects model and comparing several models, Akaike Information Criterion (AIC) will be used as the model selection criteria [52]. Given a set of candidate models fitted to the data, the AIC estimates the quality of each model relative to each of the other models and provides a means for model selection. The AIC is defined as twice the difference between the value of the log-likelihood (a measure of fit) and the total number of parameters (a penalty for model complexity) and the preferred model is the one with the minimum AIC value.

To accommodate more complex temporal trends, a quadratic-spline temporal random effects model is also considered:

$$y_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 Q(t_{ij}) + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + b_{4i} Q(t_{ij}) + \epsilon_{ij} \quad (3)$$

Model (2) and Model (3) are referred to as the linear-spline and quadratic-spline random effects models, respectively. To determine the significance of the spatial heterogeneities in the temporal trends due to the hurricane impact, we compare the linear-spline and quadratic spline models to the following “benchmark” model which does not include a random slope for the linear spline

$$y_{ij} = \beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 Q(t_{ij}) + b_{1i} + b_{2i} t_{ij} + \epsilon_{ij} \quad (4)$$

Model (4), referred to as the non-spline random effects model, can account for region-level change in the time trend of Covid-19 case rates after the disruptive hurricane, but not the heterogeneities of the census tracts. We define a Heterogeneity Metric (HM), computed based on the mixed-effect models, to quantify the spatial heterogeneities of Covid-19 case rate accelerations between census tracts *incurred following the hurricane* as

$$HM = \frac{AIC_0 - AIC_1}{|AIC_0|} \quad (5)$$

which measures the improvement in the fit of the model (reduction of the AIC) of the non-spline random effects model attained by using a spline random effects model. Specifically AIC_0 is the AIC of non-spline model (4) and AIC_1 is the smallest AIC out of the AICs of the linear and quadratic spline models (2) and (3), respectively. The larger the HM , the more heterogeneous the impact of the hurricane on Covid-19 case rate accelerations in a region.

4.2. Selecting the optimal covariate structure by modeling the data of impacted regions

In the second stage of the analysis (once the best functional forms of the temporal trends are identified) we determine the functional forms and the covariates of $\gamma(\mathbf{x}_i)$ and $g_i(\mathbf{x}_i)$ from the observed infection count and demographic data. The analysis consists in optimally selecting the sets of demographic, socioeconomic and health variables that are most strongly associated with increased longitudinal Covid-19 counts in the regions and time periods of interest. The resulting mixed-effects model with covariates will enable us to determine how changes in census tract-level infection counts are associated with the demographic, socioeconomic and health characteristics of the region.

To represent the relation between the covariates $\mathbf{x}_i$ and the infection counts in addition to the temporal trends, we add in Model (2) two new terms, a fixed effect term $\gamma(\cdot)$ and a random effect term $g_i(\cdot)$. Both terms assume a linear functional form, that is, $\gamma(\mathbf{x}_i) = \gamma_1 x_{i1} + \dots + \gamma_p x_{ip}$ and $g_i(\mathbf{x}_i) = g_{i1} x_{i1} + \dots + g_{ip} x_{ip}$, where $(\gamma_1, \dots, \gamma_p)$ is the vector of fixed slopes of the covariates and $(g_{i1}, \dots, g_{ip})$ is the vector of random slopes of the covariates for census tract i . Accordingly, a linear-spline temporal random effects model with covariates is:

$$y_{ij} = \underbrace{\beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4 Q(t_{ij}) + \gamma(\mathbf{x}_i)}_{\theta(t_{ij}) + \gamma(\mathbf{x}_i)} + \underbrace{b_{1i} + b_{2i} t_{ij} + L(t_{ij}) + g_i(\mathbf{x}_i)}_{f_i(t_{ij}) + g_i(\mathbf{x}_i)} + \epsilon_{ij}. \quad (6)$$

The fixed slopes $(\gamma_1, \dots, \gamma_p)$ describe the relation between the region-level mean infection rates and the covariates, while the random slopes $(g_{i1}, \dots, g_{ip})$ describe how the i -th census-tract relation deviates from the region-level relation.

The model parameters are estimated using a restricted maximum likelihood estimation (RMLE) approach ([22,23]). The model parameters to be estimated consist of the vector of fixed effects $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \dots, \gamma_p)$, containing 1 intercept, 3 temporal slopes and p covariate slopes, and the vector of random effects $b_i = (b_{1i}, b_{2i}, b_{3i}, g_{1i}, \dots, g_{pi})$ for each census tract indexed by i , containing 3 temporal slopes and p covariate slopes for each census tract, for a total of $N(p + 3)$ slopes for the entire region. The estimation approach relies on the assumption that the random effects b_i and the random modeling errors ϵ_{ij} from N census tracts and n observation periods follow Gaussian probability density functions:

$$b_i \sim N_r(0, \Psi), \text{ and } \epsilon_{i1}, \dots, \epsilon_{in} \sim N(0, \sigma^2)$$

where Ψ is the $r \times r$ variance-covariance matrix of the $r = p + 3$ dimensional vector of random-effects b_i and σ^2 is the error variance. The terms $N_r(0, \Psi)$ and $N(0, \sigma^2)$ denote Gaussian density functions: the first one being an r -dimensional distribution with zero mean and covariance matrix Ψ and the second one being a univariate distribution with zero mean and with common variance σ^2 . The parameters of the model to be estimated by the RMLE method, therefore, are $\beta, b_1, \dots, b_N, \Psi$ and σ^2 . The heterogeneities in random effects are measured by their standard deviations, which are obtained as the square root of the diagonal elements of the matrix Ψ : the effects that have large spatial variations have large standard deviations.

The demographic, socioeconomic and health-related variables listed in Table 2 were used as potential covariates to describe a community's social vulnerability against high Covid-19 infections. A top-down variable selection strategy, commonly used in the mixed-effects model selection literature ([23], p.135), was utilized to select the covariates of the mixed-effects models. For a given region, the global relations and trends are dictated by the fixed effects, however, if the random effects are poorly chosen, the fixed effects are biased because the random effects work their way into the standard errors of the slopes for the fixed effects. On the other hand, if the variation in the outcome variable is not modeled in terms of fixed effects, it ends up in random effects. Finding the correct random structure, therefore, is crucial not only to adequately estimate the fixed slopes of the model by properly modeling the variance, but also through the estimation of the random slopes which will enable the analysts to assess the spatial heterogeneities of the relations.

The top-down strategy consists of the following steps: (a) Start with a model where the fixed component contains all potential covariates and a random structure consisting of random slopes only for the temporal effects. This is called the beyond optimal model. (b) Using the beyond optimal model, find the optimal random structure. Start with a random intercept, enter a random slope for one covariate at a time to the random structure using the likelihood ratio test until no significant random slopes remain (c) Once the optimal random structure has been found, the optimal fixed structure is found by comparing several models with nested fixed effects, but with the same random structure. A fixed slope for one covariate at a time is removed from the fixed structure using the likelihood ratio test.

Estimation, diagnostics, and visualization of the multilevel mixed-effects models to the spatial infection, demographic and socioeconomic data are conducted using the `lmer` function in the R package `lme4`. In addition, the R package `tigris` is used to read and process the shape files of the census tract polygonal boundaries and their populations and the functions in `sp` package are used for plotting mapping functions.

5. Application to the first COVID-19 wave and hurricane Sally in Florida

In the first stage, we identify the temporal structure of the mixed-effects model based on the census-tract level daily Covid-19 infection count data for the Florida regions shown in Fig. 1. To accommodate the different rates of change in the infection case rates before and after the occurrence of hurricane Sally at the region level, we fitted the benchmark model (4) with a common knot at day $\tau = 52$, that corresponds to Sally's landfall date of Sep 10, 2020. Since the benchmark model does not contain a spline in the random-effects part, it is able to quantify the temporal impact of the hurricane at the region level but not at the census-tract level. To accommodate any census-tract level heterogeneities in the changes in the Covid-19 infection case rates after the hurricane, we fitted the linear-spline and quadratic-spline random-effects models (2) and (3), respectively. We compare these models to model (4) on the basis of AIC and compute the Heterogeneity Metric (*HM*) using Equation (5), to determine which regions impacted the most, in terms of heterogeneities in the Covid-19 case rate accelerations, following the hurricane. All spline models have a common knot at day $\tau = 52$.

To visualize the basic features of the mixed-effect models, Fig. 3 compares the predictions of models (2), (3), and (4) to the observed Covid-19 counts for only one census tract in Volusia County and one census tract in Okaloosa County. Note that fitted models provide predicted case rate y_{ij} of census tract i at measurement time j . Therefore, the predicted case count is obtained using the inverse transformation $z_{ij} = \exp(y_{ij})m_i + 1$ where m_i is the population. There is an evident change in the slope of the case counts near the landfall on day 52 for Volusia but the change in Okaloosa is somewhat before day 52. The linear-spline model (2) has a broken-line turning point at the knot (green line), while the quadratic-spline model (3) has a smooth turning point at the knot (red line). The baseline model (4) (blue line) is not able to capture the turning point and fits the data poorly.

It is important to note that in most of the public health surveillance studies a common practice is to use some type of moving average or weekly sum of infection case counts to smooth the volatility in the reported infection count data [53]. In this application we used 7-day moving-averaged case count data to fit the mixed-effects models in order to reduce the effect of volatility on the estimated parameters. In addition, to see the effects of data volatility on the model results, we repeat the analysis of the temporal model with the raw (un-averaged) case count data.

Supplementary Appendix 1a shows the AIC values of the fitted models for all Florida regions. The last column is the Heterogeneity Metric (*HM*), measuring the improvement of the model fit by the addition of a random slope on the spline with a knot at the hurricane landfall time (AIC_1) over the benchmark, non-spline model (AIC_0). *HM* of regions are summarized in Fig. 4. *HM* explicitly quan-

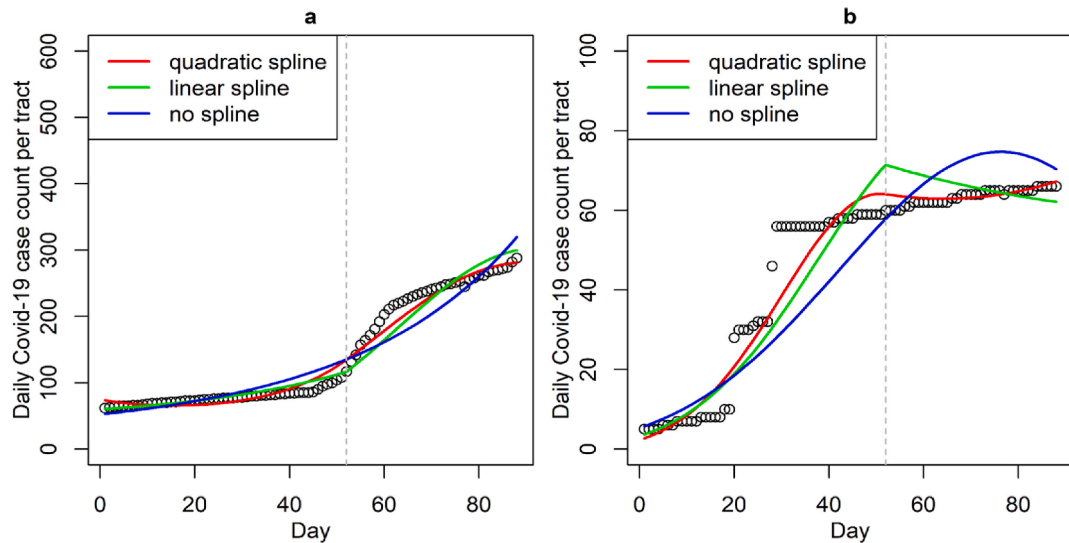


Fig. 3. Covid-19 count trajectories and model predictions of case counts of two census tracts in region 1 using various spline functions (a) a tract in Volusia County (b) a tract in Okaloosa County. Vertical dashed line corresponds to the known land fall day $\kappa = 52$.

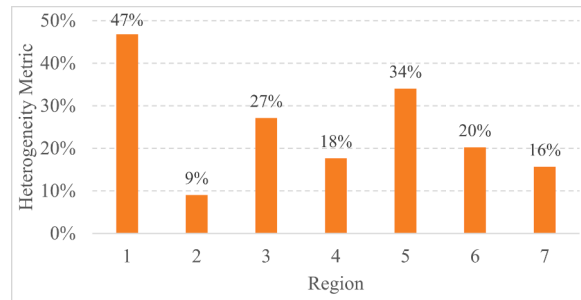


Fig. 4. Heterogeneity metric computed based on the mixed-effects models of the regions representing the effect of hurricane landfall on Covid-19 case rates.

tifies the impact of the hurricane landfall on the spatial heterogeneity of Covid-19 case rate accelerations within census tracts. It can be seen that, as expected, the hurricane has the greatest impact on the heterogeneity of increased Covid-19 case rates in region 1, with the closest proximity to the landfall location, among all regions. In addition, region 5, that has the second highest impact due to the hurricane, even though it is located farther away from the landfall, warrants further consideration, and will be evaluated in what follows. In addition (from [Appendix 1a](#)) the linear-spline model, model (2), provides the best fit compared to non-spline benchmark model (4) and the quadratic-spline model (3), in all but one of the regions, as indicated by the smaller AIC values (the smallest AIC value is shown in bold fonts). Therefore, for identifying the regions most impacted by the hurricane and selecting the covariates of the models, we consider a linear-spline random-effects structure in the second level analysis. [Supplementary Appendix 1b](#) shows the model AICs of the regions fitted with the raw data. It can be seen that similar conclusions are obtained with raw data to those from smoothed data (regions 1 and 5 are impacted greatest and the linear spline random effects model fits best in all but one of the regions). While the use of smoothed data is preferred in this analysis, this comparison is useful in highlighting the robustness of the conclusions obtained with the proposed method with both data types.

The second stage analysis will be applied to both regions 1 and 5 using the 7-day moving-averaged case count data. The observed Covid-19 case count trajectories and the predictions obtained by the fitted multi-level linear-spline mixed-effects model (2) are shown in [Figs. 5 and 6](#) for regions 1 and 5, respectively. The predicted daily infection counts for each census-tract are shown with a different color, with a solid line corresponding to the prediction and a dashed line corresponding to the observed data, and panels corresponding to different counties (The curves are grouped for counties in separate panels for ease of viewing). The predicted case counts follow the observed case counts closely, demonstrating that the model describes census-tract level trends well.

In the second stage of the analysis, we identify which of the demographic, socioeconomic and health covariates given in [Table 2](#) are most strongly associated with increased Covid-19 case rates in the highly impacted regions. As the number of potential covariates is relatively high, using all of them in the model before a dimension reduction can result in highly unstable models and degraded predictive performance if the covariates are highly multicollinear. In order to avoid multicollinearity between the predictors a heuristic approach [54] is used to remove a minimum number of covariates to ensure that all pairwise correlations are below a threshold of 0.6. [Supplementary Appendix 2a and 2b](#) show the correlation matrix of the covariates with data from regions 1 and 5, respectively. Each pairwise correlation is colored according to its magnitude and the covariates have been grouped using a clustering technique so

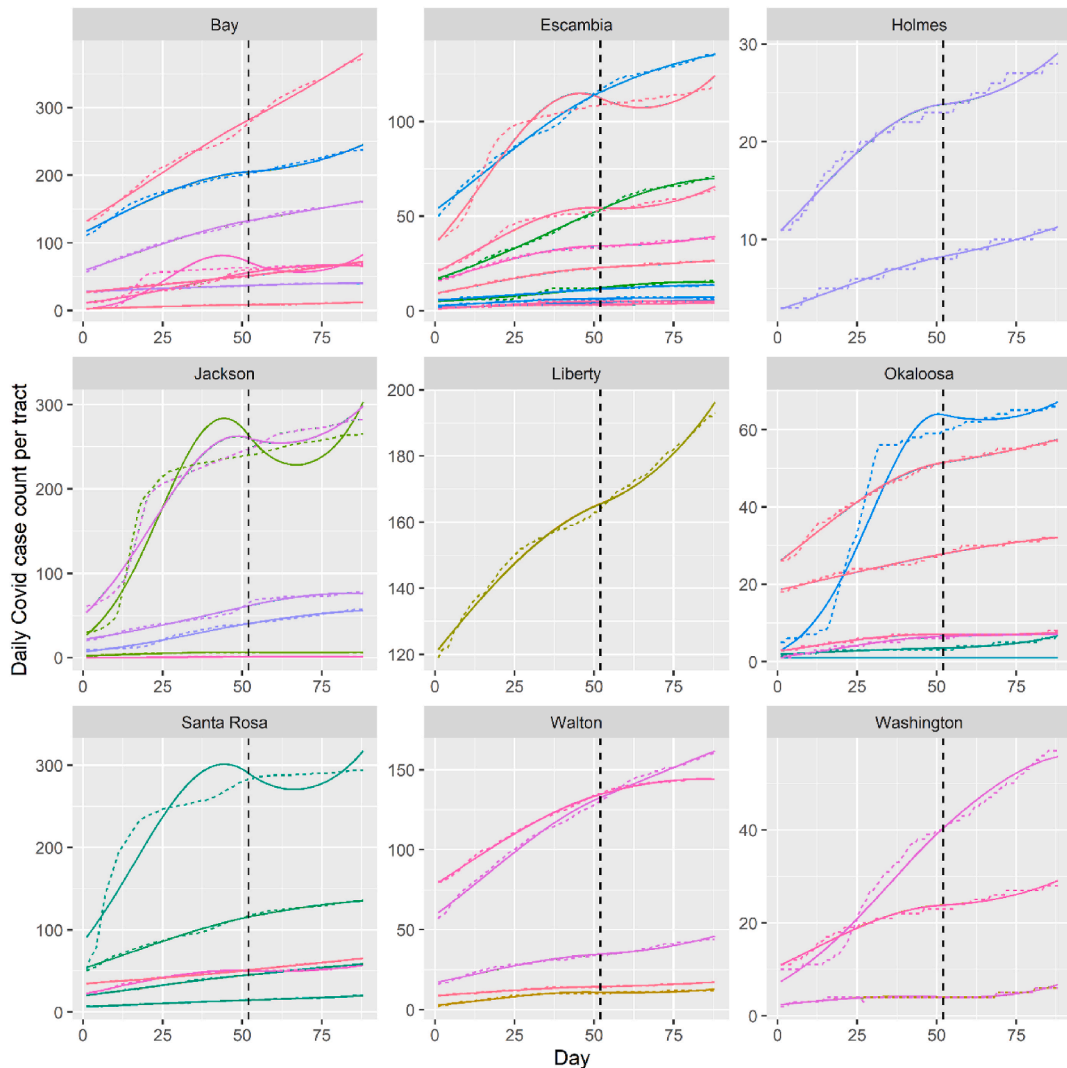


Fig. 5. Census-tract level daily infection count data and predictions for region 1. The time period covers from 7/22/2020 to 10/17/2020. Vertical line shows the land-fall date November 9, 2020 of hurricane Sally. Dashed lines: Confirmed daily counts; Solid lines: Mixed-effects model predictions. Each solid or dashed line pair with the same color corresponds to a different census tract within the county. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

that collinear groups of predictors are adjacent to one another [55]. For example, for region 1 (in [Appendix 2a](#)) it can be seen that, looking along the diagonal, there are 2 blocks of strong positive correlations that indicate clusters of collinearity. Based on the heuristic algorithm, EP_NOHSDP, EP_PCI, EP_POV and Obesity were dropped for the region 1 analysis and EP_MINRTY, EP_NOHSDP, EP_S-NGPNT, EP_POV, EP_AGE65 and EP_NOVEH were dropped from the region 5 analysis.

The proposed top-down variable selection strategy is then applied to the resulting list of reduced dimensional variables and choose which covariates should be added to determine the optimal mixed-effects model (6). The coefficients of the resulting optimal mixed-effect models for region 1 and region 5 are shown in [Table 3](#) and [Table 4](#), respectively. The estimated β slopes provide the fixed-effects of the covariates, while the random-effects part provides the estimated standard deviations and the corresponding 95 % confidence interval, for each random slope b_i . At 95 % confidence, a fixed effect is deemed significant (i.e., the region level effect is different than 0) if the corresponding p-value is less than 0.05, while a random effect is deemed significant (i.e., the census-tract level heterogeneity is large) if the 95 % confidence interval of the standard deviation does not encompass 0.

On the basis of the fixed-effects structure of the models following insights are gained. In region 1, percentage of persons in group quarters (EP_GROUPQ), per capita number of hospital beds (PctBeds), percentage of persons who speak English less than well (EP_LI-MENG), air pollution level (AirPollution) and multi-unit housing (EP_MUNIT), were predictive of increased Covid-19 daily case rates. A percentage increase in persons in group quarters (EP_GROUPQ) was associated with a 3.9 % increase in daily case rates ($e^\beta = 1.039$, $p = <0.0001$), a unit decrease in per capita number of hospital beds (PctBeds) is associated with a $1/0.7427 = 1.35$ -fold increase in daily case rates ($e^\beta = 0.7427$, $p = <0.0001$) and a percentage increase in multi-unit housing (EP_MUNIT) is associated

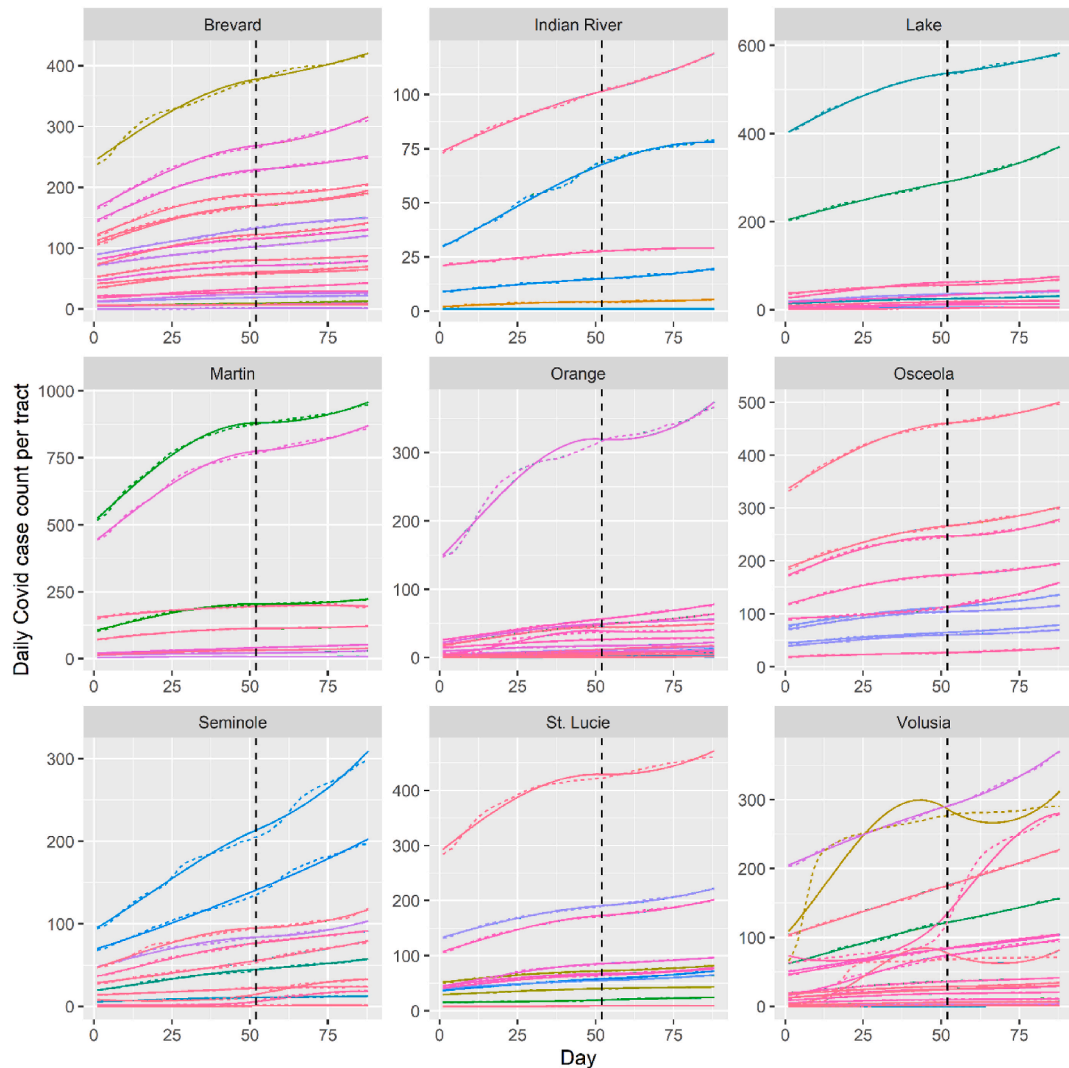


Fig. 6. Census-tract level daily infection count data and predictions for region 5. The time period covers from 7/22/2020 to 10/17/2020. Vertical line shows the land-fall date November 9, 2020 of hurricane Sally. Dashed lines: Confirmed daily counts; Solid lines: Mixed-effects model predictions. Each solid or dashed line pair with the same color corresponds to a different census tract within the county. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

with a 1.42 % increase in daily case rates ($e^{\beta} = 1.0142$, $p = 0.005$). The negative relationship between percentage of persons who speak English less well (EP_LIMENG) and the daily case counts and the negative relationship between AirPollution and the daily case counts were counter-intuitive. The counter-intuitive negative coefficient of a variable may be due to a common confounder that was not included in the model but has an unknown effect on both the variable and the infection counts. Confounding variables commonly arise in non-causal statistical models, like the ones we employ here, and their coefficients do not necessarily imply causal relationship ([56], pp. 77). For example, a third confounder that is jointly causing EP_LIMENG to be lower and the infection counts to be higher in these rural communities may result in such negative coefficients. In region 5, percentage of population with a disability (EP_DISABL), percentage of persons in group quarters (EP_GROUPQ), percentage of persons in multi-unit housing (EP_MUNIT), and percentage of persons who speak English less than well (EP_LIMENG) were predictive of increased Covid-19 daily case rates. A percentage increase in population with a disability (EP_DISABL) was associated with a 4.91 % increase in daily case rates ($e^{\beta} = 1.0491$, $p = 0.005$), a percentage increase in persons in group quarters (EP_GROUPQ) was associated with a 3.34 % increase in daily case rates ($e^{\beta} = 1.0334$, $p < 0.0001$), and a percentage increase in multi-unit housing (EP_MUNIT) was associated with a 1.79 % increase in daily rates ($e^{\beta} = 1.0179$, $p = 0.001$). Similar to region 1, the negative relationship between percentage of persons who speak English less well (EP_LIMENG) and the daily case counts is counter intuitive: a percentage increase in persons who speak English less well was associated with a 12 % decrease in daily case rates ($e^{\beta} = 0.8800$, $p = 0.017$). However, EP_LIMENG also appears significant as a random effect and this association is better described by local spatial variations between census tracts as modeled by the random-effect structure of the model.

Table 3
Mixed-effects model for region 1.

Fixed Effects				
Variable	β	t-value	e^β	p-value
Intercept	24.210	4.337	3.25E10	0
Linear trend, t_{ij}	0.0092	33.836	1.0092	0
Quadratic trend, t_{ij}^2	-0.0002	-69.127	0.9997	0
Quadratic spline, $Q(t_{ij}) = (t_{ij} - \kappa)_+^2$	0.0002	26.830	1.0002	0
EP_GROUPQ	0.0383	6.861	1.0390	0
PctBeds	-0.2970	-4.386	0.7427	0
EP_LIMENG	-0.0977	-1.950	0.9068	0.051
EP_MUNIT	0.0141	2.767	1.0142	0.005
AirPollution	-2.8220	-5.182	0.0594	0
Random Effects				
Variable	Std. Dev.	95 % LCB	95 % UCB	
Intercept	1.1351	1.023	1.284	
EP_GROUPQ	0.0380	0.024	0.055	
Linear trend, t_{ij}	0.0095	0.008	0.010	
Linear spline, $L(t_{ij}) = (t_{ij} - \kappa)_+$	0.0140	0.012	0.015	
Residual	0.0982	0.097	0.099	

Table 4
Mixed-effects model for region 5.

Fixed Effects				
Variable	β	t-value	e^β	p-value
Intercept	-5.906	-16.988	0.0027	0
Linear trend, t_{ij}	0.0092	33.858	1.0093	0
Quadratic trend, t_{ij}^2	-0.0002	-69.169	0.9997	0
Quadratic spline, $Q(t_{ij}) = (t_{ij} - \kappa)_+^2$	0.0002	26.827	1.0002	0
EP_DISABL	0.0479	2.794	1.0491	0.005
EP_MUNIT	0.0177	3.147	1.0179	0.001
EP_GROUPQ	0.0329	5.59	1.0334	0
EP_LIMENG	-0.1277	-1.928	0.8800	0.017
Random Effects				
Variable	Std. Dev.	95 % LCB	95 % UCB	
Intercept	1.2894	1.0552	1.519	
EP_LIMENG	0.2630	0.0785	0.5087	
EP_GROUPQ	0.0284	0.0183	0.0529	
Linear trend, t_{ij}	0.0096	0.0087	0.0107	
Linear spline, $L(t_{ij}) = (t_{ij} - \kappa)_+$	0.0140	0.0127	0.0156	
Residual	0.0981	0.0970	0.9900	

On the basis of the random-effects structures given in Tables 3 and 4 and it can be observed that the spatial variation of the relationship between the Covid-19 daily case rates and EP_GROUPQ (for region 1) and the relationship between the daily case rates and EP_LIMENG and EP_GROUPQ (for region 5) are very heterogeneous. For example, in region 5, the large heterogeneity measured by the standard deviation 0.263 of the effect of EP_LIMENG helps explain the counter-intuitive result of this coefficient in the fixed-effect structure. It implies that the fixed effect (found as $e^\beta = 0.8800$ or a 12 % decrease in the daily case rates), has a tract-to-tract variation represented by the 95 % confidence interval of $e^{\beta \pm 2\sigma} = e^{-0.1277 \pm 2(0.263)} = (0.52, 1.48)$. Therefore, depending on the census tract the effect of speaking English less well on the daily Covid-19 case rates after the hurricane landfall varies between a decrease by 48 % and an increase by 48 %. This result demonstrates a unique benefit of the mixed-effects models with spatially varying coefficients in quantifying large spatial variations of the effects of such demographic variables.

In addition, the large and positive random slope of the linear spline $L(t_{ij})$ (in both regions 1 and 5) tells us that the of daily case rate accelerations after the hurricane is highly heterogeneous for both regions. The mixed-effects model provides an estimated slope for every census tract for each random effect. The larger values of the random slope of $L(t_{ij})$ indicate the census-tracts with large increases in the acceleration of Covid-19 case rates after the hurricane landfall. The larger (positive) values of the random slope of the covariate EP_GROUPQ for region 1 and, the covariates EP_LIMENG and EP_GROUPQ for region 5, indicate the census-tracts with large increases in the Covid-19 case rates due to a percentage increase in the respective covariate. The joint effect of the pandemic and the hurricane, therefore, can be assessed by considering these two random coefficients simultaneously: the census tracts with the large (positive) random slope of $L(t_{ij})$ and the large (positive) slope of the covariate can be identified as the highly vulnerable areas because these areas experience the large increases in the case rates after the hurricane and have the strongest associations with the covariate.

The spatially varying coefficients of mixed-effects models can be visualized to aid further interpretation. For the model fitted for region 1 (Table 3), Fig. 7a shows a bivariate choropleth map of the census tracts colored according to both the random slope of the linear spline $L(t_{ij})$ (labelled “trendSally”) and the random slope of EP_GROUPQ. Similarly, for the model fitted for region 5 (Table 4), Fig. 8a shows bivariate choropleth map of region 5 according to the random slope of the spline basis and the random slope of EP_LIMENG (The tracts shaded white represent tracts with missing population data and are excluded from the analysis). The bivariate map for EP_GROUPQ for region 5 is omitted here to save space. Each bivariate map is constructed by dividing the joint distribution of the corresponding pair of slopes into 9 bins based on the 33-rd and the 66-th percentiles of each slope and classifying them into Low, Medium, and High categories. To isolate the impact of the hurricane landfall, Fig. 7b shows the census tracts of region 1 colored *only* according to the variable EP_GROUPQ and the centroids of census tracts for which *both* random slopes are in the High category from Fig. 7a. Similarly, Fig. 8b shows the census tracts of region 5 colored only according to the variable EP_LIMENG and the centroids of census tracts for which *both* slopes are in the High category from Fig. 8a. In these figures (Fig. 7b and 8b), the vulnerability of the census tracts shown with darker colors and the centroid markers are driven by both the social factors and the hurricane landfall. By contrast, the vulnerability in the tracts that are darker but do not have their centroid indicated are driven only by social factors. This shows the unique insights that can be gained by joint modeling of pandemic and hurricane effects using a mixed-effects model and the use of bivariate maps such as those in Fig. 7a and 8a are crucial in identifying the needs of populations during hurricane-pandemics.

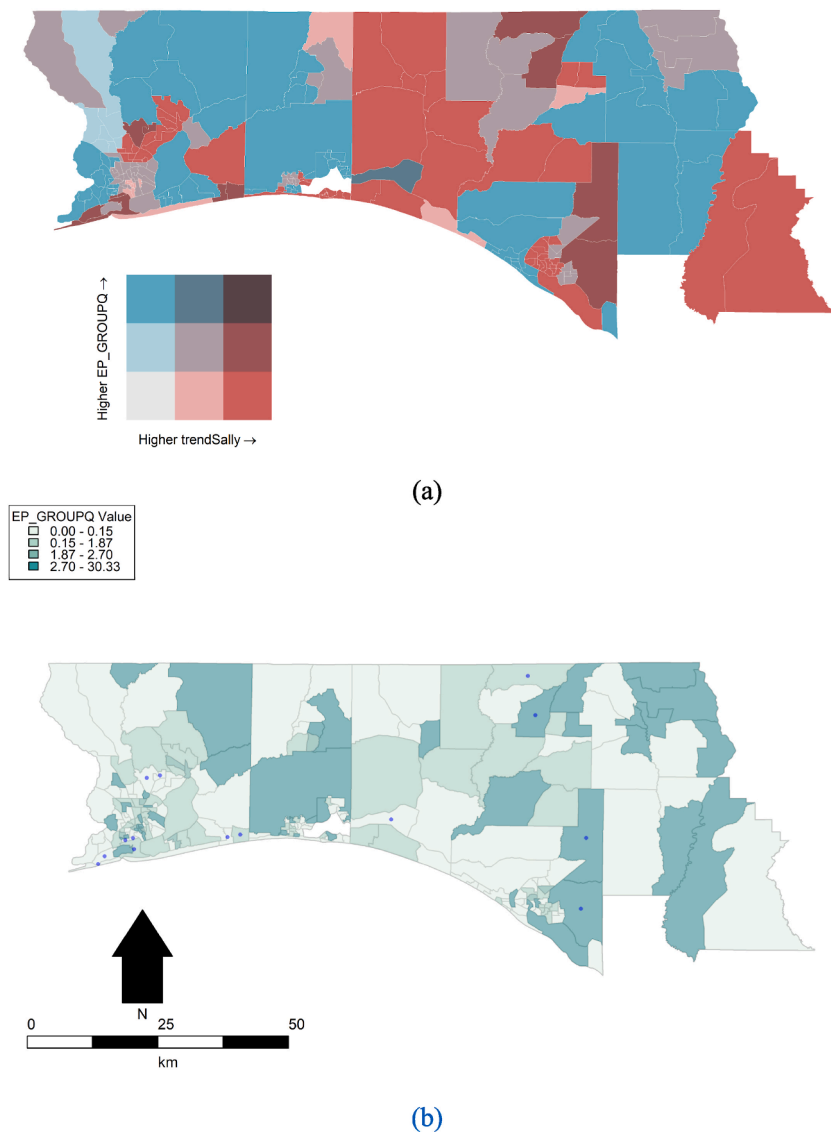


Fig. 7. (a) Bivariate map of region 1 census tracts colored according to the random slope of the linear spline (measuring acceleration/deceleration in Covid-19 case rates following hurricane Sally) and the random slope of EP_GROUP (measuring the effect of this factor on increased Covid-19 case rates) (b) Census tracts of region 1 colored according to the EP_GROUP variable and the centroids of the census tracts in the High-Medium and Medium-High categories of the bivariate map shown with markers.

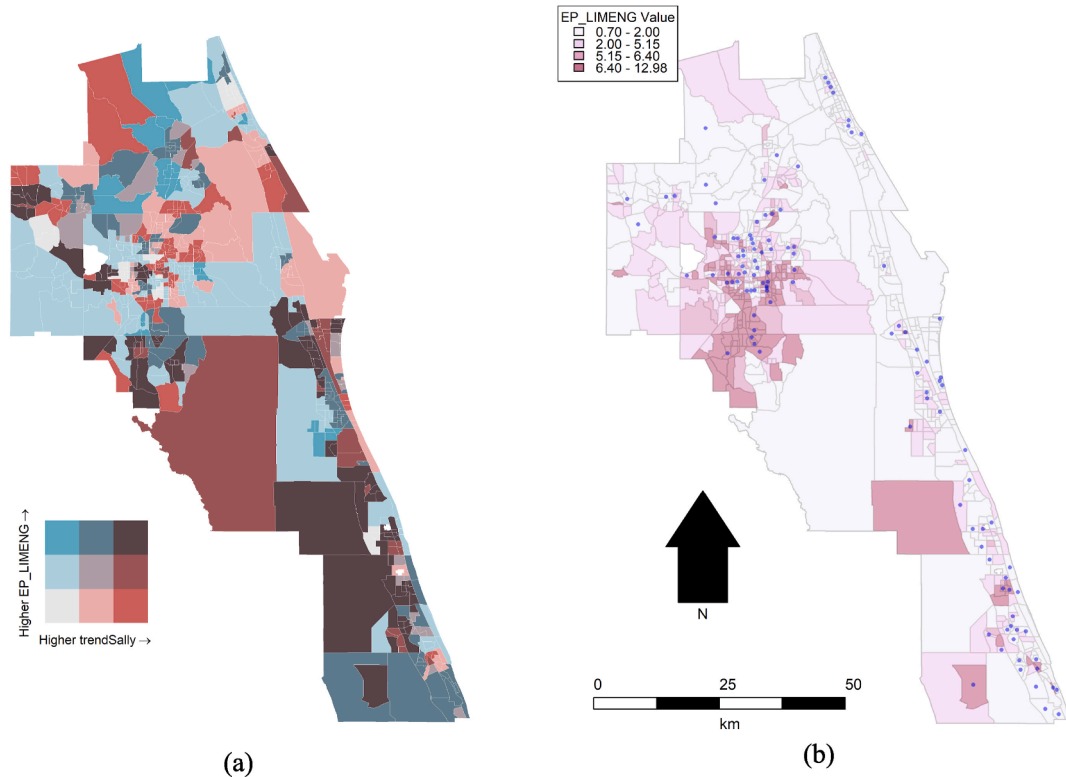


Fig. 8. (a) Bivariate map of region 5 census tracts colored according to the random slope of the linear spline (measuring acceleration/deceleration in Covid-19 case rates following hurricane Sally) and the random slope of EP_LIMENG (measuring the effect of this factor on increased Covid-19 case rates) (b) Census tracts of region 5 colored according to the EP_LIMENG variable and the centroids of the census tracts in the High-High category of the bivariate map shown with blue markers. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

A remark on the choice of multiple models for regions: For this same data set, a single mixed-effects model for the entire state could be estimated. However, in addition to the obvious computational benefits, analyzing the data for regions separately, as we did in this paper, has the added benefit of identifying patterns that are unique to the regions and allowing the regional emergency management officials and governments to take response and preparedness actions for allocating resources more effectively. It must be remembered, depending on the goals of the analysis, a multi-regional/multi-level approach may present limitations in adequately capturing the common characteristics or shared patterns across the entire study area. If the goal is to better understand the local patterns, then a multi-regional model is preferable over a single-region model. However, if the goal is to understand global patterns, then the latter approach is preferable. In our multiple region approach, Region 1 has EP_GROUPQ as the dominant factor impacting spatial heterogeneity, while Region 5 has EP_LIMENG as the dominant factor, a conclusion that may not be reached with a single model. In addition, the time trends of infection counts from different regions were distinct depending on the proximity to the hurricane landfall, as the AICs of our models of the regions have shown. Similar conclusions can be reached with a single model with random coefficients for the entire state. However, this would require an additional analysis step, such as a spatial partitioning and grouping of the tract-level model coefficients (e.g., Ref. [57]), to understand the patterns of variations of these coefficients under different regions.

6. Discussion: practical considerations and limitations

The proposed multi-level modeling of the compound impacts of geophysical disasters co-occurring with pandemics has practical implications for emergency management and public health policy for coordinating resource allocation efforts between state, region, and community (i.e., census tract) levels as shown in Fig. 2. By implementing the approach, the longitudinal infection data from more than 4000 census tracts of the state of Florida can be analyzed in a unified framework to efficiently to extract patterns and heterogeneities of vulnerability unique to the local communities.

By enabling practitioners to analyze the temporal disease spread dynamics and quantifying the significance of the spatial differences of the adaptive capacities of the communities, the method provides critical insights. For example, in the case study, the percentage of persons in group quarters (EP_GROUPQ), a random-effect, shown a significant variation depending on the location within the region with respect to increased infection rates, while fixed-effects, such as the effect of hospital beds, was constant throughout the region. To improve resiliency against compound hazards, the fixed effects can guide region-wide resource allocations, while the random effects provide guidance on how allocate resources at the community level.

Here, regions that are comprised of multiple contiguous counties, can be considered as emergency management districts. The grouping of counties to regions was needed to ensure the computational tractability of the linear mixed effects model by maintaining a reasonable number of parameters to be estimated from a large spatial data set and the entire state. While in practice, the emergency management plans are coordinated between county and state levels, to illustrate the application of the approach for the entire state, the hierarchical structure of state→region→census tract of Fig. 2 was adopted in this study. Returning to the policy front, the methodology is general, however, and it can be conducted for county level planning by following a region→county→census tract hierarchy. Although the outcomes should be treated with caution, since the selected levels may impact the results due to the possible spatial relations within the levels, this type of approach can help develop and implement more efficient and integrated pre- and post-disaster plans if incorporation between state and local agencies can be successfully achieved. Policymakers, thus, should focus on integrating these plans with considering disparities at multiple levels (i.e., county-level vs. region-level).

7. Conclusions

The paper presented a new approach to systematically study the effects of compound hazards and distinct patterns of infectious disease spread during hurricanes. By contrast to existing methods, the methodology provides unique contributions in modeling the (dynamic) associations between infection case rates and social factors that vary over time and also quantifying how these associations vary geographically between communities. An application with data from the first wave of the Covid-19 pandemic during a concurrent hurricane in northwest Florida is considered for illustration. It has been shown that the spatial heterogeneity assessed by the proposed modeling method would provide crucial information to local health officials and emergency managers to help them allocate resources and plan for community-level impacts of concurrently occurring hurricane pandemics.

The proposed approach can be extended to other locations in the U.S.; however, the efficiency of the model at a particular location for a given hurricane-pandemic situation can be qualitatively impacted by several critical factors needed to be studied in detail, including economic conditions, cultural traditions, local and political constraints, and spatial characteristics.

The multilevel models built for the regions of Florida helped highlight factors associated with increased Covid-19 infections both at the region level (fixed-effects) and those associated with increased heterogeneity in their association with increased infections at the census tract-level (random effects). The regression-spline based mixed-effects model identified one region near the hurricane landfall location, and another somewhat remote region, to have significantly increased Covid-19 case counts after the hurricane. Given the high uncertainty in the hurricane track forecast (Fig. 1) the increased case counts implied by the model for the remote region may be due to the evacuations of the communities between this region and the other regions or may be due to other reasons. As future work, GPS or cell-phone human movement data, if available, can be incorporated into the mixed-effects model to uncover the causes of the high infection counts in such regions using human mobility network approaches [58,59]. As another future work, a multivariate clustering of geographically distributed random effects coefficients, similar to Ref. [57], can be undertaken for the regionalization of similarly vulnerable areas for effective resource allocation purposes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by National Science Foundation (NSF) grant CMMI-2101091. The authors would like to acknowledge the funding support from NSF.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijdr.2023.104095>.

References

- [1] Our World in Data, United States: Coronavirus Pandemic Country Profile, 2022. <https://ourworldindata.org/coronavirus/country/united-states>.
- [2] J. Malilay, M. Heumann, D. Perrotta, A.F. Wolkin, A.H. Schnall, M.N. Podgornik, J.R. Greenspan, The role of applied epidemiology methods in the disaster management cycle, *Am. J. Publ. Health* 104 (11) (2014) 2092–2102.
- [3] T. Kawano, Y. Tsugawa, K. Nishiyama, H. Morita, O. Yamamura, K. Hasegawa, Shelter crowding and increased incidence of acute respiratory infection in evacuees following the Great Eastern Japan Earthquake and tsunami, *Epidemiol. Infect.* 144 (4) (2016) 787–795.
- [4] F.D. Orata, P.S. Keim, Y. Boucher, The 2010 cholera outbreak in Haiti: how science solved a controversy, *PLoS Pathog.* 10 (4) (2014) e1003967.
- [5] FEMA, Mass Care/Emergency Assistance Pandemic Planning Considerations, Federal Emergency Management Agency, 2020. https://www.fema.gov/media-library-data/1591805537743-dd381cfb5c45eea9999dac1637815716/MCEA_Pandemic_Planning_Considerations_Guide_508.pdf.
- [6] National Academies of Sciences, Engineering, and Medicine, Framework for Equitable Allocation of COVID-19 Vaccine, National Academies Press, 2020.
- [7] B. Wisner, P. Blaikie, T. Cannon, I. Davis, *At Risk: Natural Hazards, People's Vulnerability and Disasters*, Routledge, 2014.
- [8] S.L. Cutter, B.J. Boruff, W.L. Shirley, Social vulnerability to environmental hazards, *Soc. Sci. Q.* (2003) 242–261.
- [9] B.E. Flanagan, E.J. Hallisey, E. Adams, A. Lavery, Measuring community vulnerability to natural and anthropogenic hazards: the Centers for disease control and prevention's social vulnerability index, *J. Environ. Health* 80 (10) (2018) 34.

- [10] I.M. Karaye, J.A. Horney, The impact of social vulnerability on COVID-19 in the US: an analysis of spatially varying relationships, *Am. J. Prev. Med.* 59 (3) (2020) 317–325.
- [11] S.W. Marvel, J.S. House, M. Wheeler, K. Song, Y.H. Zhou, F.A. Wright, D.M. Reif, The COVID-19 Pandemic Vulnerability Index (PVI) Dashboard: monitoring county-level vulnerability using visualization, statistical modeling, and machine learning, *Environ. Health Perspect.* 129 (1) (2021) 017701.
- [12] M.S. Rahman, K.C. Paul, M.M. Rahman, J. Samuel, J.C. Thill, M.A. Hossain, G.M.N. Ali, Pandemic vulnerability index of US cities: a hybrid knowledge-based and data-driven approach, *Sustain. Cities Soc.* 95 (2023) 104570.
- [13] J. Lighter, M. Phillips, S. Hochman, S. Sterling, D. Johnson, F. Francois, A. Stachel, Obesity in patients younger than 60 years is a risk factor for Covid-19 hospital admission, *Clin. Infect. Dis.* 71 (15) (2020) 896–897.
- [14] L. Fang, G. Karakiulakis, M. Roth, Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir. Med.* 8 (4) (2020) e21.
- [15] L.M. Andersen, S.R. Harden, M.M. Sugg, J.D. Runkle, T.E. Lundquist, Analyzing the spatial determinants of local Covid-19 transmission in the United States, *Sci. Total Environ.* 754 (2021) 142396.
- [16] M. Nuutinen, I. Haavisto, A.J. Niemi, A. Rissanen, M. Ikkivuo, R.L. Leskelä, Statistical model for factors correlating with COVID-19 deaths, *Int. J. Disaster Risk Reduc.* 82 (2022) 103333.
- [17] X. Wu, R.C. Nethery, M.B. Sabbath, D. Braun, F. Dominici, Exposure to Air Pollution and COVID-19 Mortality in the United States: A Nationwide Cross-Sectional Study, *MedRxiv*, 2020.
- [18] A. Ehler, The socio-economic determinants of COVID-19: a spatial analysis of German county level data, *Soc. Econ. Plann. Sci.* 78 (2021) 101083.
- [19] A. Henning, C. McLaughlin, S. Armen, S. Allen, Socio-spatial influences on the prevalence of COVID-19 in central Pennsylvania, *Spatial and Spatio-temporal Epidemiol.* 37 (2021) 100411.
- [20] E. Vynnycky, R. White, *An Introduction to Infectious Disease Modelling*, OUP oxford, 2010.
- [21] F. Schlosser, B.F. Maier, O. Jack, D. Hinrichs, A. Zachariae, D. Brockmann, COVID-19 lockdown induces disease-mitigating structural changes in mobility networks, *Proc. Natl. Acad. Sci. USA* 117 (52) (2020) 32883–32890.
- [22] J. Pinheiro, D. Bates, *Mixed-effects Models in S and S-PLUS*, Springer, New York, 2006.
- [23] A.F. Zuur, E.M. Ieno, N.J. Walker, A.A. Saveliev, G.M. Smith, *Mixed Effects Models and Extensions in Ecology with R*, Springer, New York, 2009.
- [24] C. Fritz, G. Kauermann, On the interplay of regional mobility, social connectedness and the spread of COVID-19 in Germany, *J. Roy. Stat. Soc.* 185 (1) (2022) 400.
- [25] T. Boschi, J. Di Iorio, L. Testa, M.A. Cremona, F. Chiaromonte, Functional data analysis characterizes the shapes of the first COVID-19 epidemic wave in Italy, *Sci. Rep.* 11 (1) (2021) 1–15.
- [26] Paul M. McKeigue, Amanda Weir, Jen Bishop, J. Stuart, McGurnaghan, Sharon Kennedy, David McAllister, Chris Robertson, et al., Rapid epidemiological analysis of comorbidities and treatments as risk factors for COVID-19 in Scotland (REACT-SCOT): a population-based case-control study, *PLoS Med.* 17 (10) (2020) e1003374.
- [27] F. Jiang, Z. Zhao, X. Shao, Modelling the COVID-19 infection trajectory: a piecewise linear quantile trend model, *J. Roy. Stat. Soc. B* 84 (5) (2021) 1589–1607.
- [28] U. Basellini, C.G. Camarda, Explaining regional differences in mortality during the first wave of Covid-19 in Italy, *Popul. Stud.* 76 (1) (2022) 99–118.
- [29] B. Nikparvar, M.M. Rahman, F. Hatami, J.C. Thill, Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network, *Sci. Rep.* 11 (1) (2021) 21715.
- [30] F. Shahid, A. Zameer, M. Muneeb, Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM, *Chaos, Solitons & Fractals* 140 (2020) 110212.
- [31] J.P. LeSage, R.K. Pace, *Introduction to Spatial Econometrics*, Chapman & Hall/CRC, 2009.
- [32] A.S. Fotheringham, C. Brunsdon, M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons, 2003.
- [33] L.A. Waller, L. Zhu, C.A. Gotway, D.M. Gorman, P.J. Gruenewald, Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models, *Stoch. Environ. Res. Risk Assess.* 21 (5) (2007) 573–588.
- [34] L.A. Waller, C.A. Gotway, *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons, 2004.
- [35] J.P. Elhorst, Specification and estimation of spatial panel data models, *Int. Reg. Sci. Rev.* 26 (3) (2003) 244–268.
- [36] N.A.C. Cressie, *Statistics for Spatial Data*, Wiley, New York, 1993.
- [37] P.J. Diggle, K.-Y. Liang, S.L. Zeger, *Analysis of Longitudinal Data*, Oxford University Press, New York, 1994.
- [38] S. McLafferty, Placing pandemics: geographical dimensions of vulnerability and spread, *Eurasian Geogr. Econ.* 51 (2) (2010) 143–161.
- [39] B.F. Snyder, V. Parks, Spatial variation in socio-ecological vulnerability to Covid-19 in the contiguous United States, *Health Place* 66 (2020) 102471.
- [40] C. Song, H. Yin, X. Shi, M. Xie, S. Yang, J. Zhou, J. Pan, Spatiotemporal disparities in regional public risk perception of COVID-19 using Bayesian Spatiotemporally Varying Coefficients (STVC) series models across Chinese cities, *Int. J. Disaster Risk Reduc.* (2022) 103078.
- [41] WHO, WHO Director-General's Opening Remarks at the Media Briefing on COVID-19, 11 March 2020, World Health Organization, 2020.
- [42] NHC, SALLY Graphics Archive, 5-day Forecast Track and Watch/Warning Graphic, 2022. https://www.nhc.noaa.gov/archive/2020/SALLY_graphics.php?product=5day_cone_with_line Other packages.
- [43] Florida department of health, FDOT Healthcare Coalitions Regions, 2022. https://www.floridahealth.gov/programs-and-services/emergency-preparedness-and-response/community-preparedness/_documents/HCCMapwithEmails.pdf.
- [44] Florida COVID-19 Hub, <https://covid19-usflibrary.hub.arcgis.com/>, 2022.
- [45] U.S. Census Bureau, Zip Code Tabulation Area 5-Digit (ZCTA5) Relationship Files (2022). <https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html>. (Accessed 3 November 2022).
- [46] U.S. Census Bureau, DP1 Profile of General Population and Housing Characteristics, Prepared by the United States Census Bureau, 2020 Washington, D.C. [https://data.census.gov/table?q=census+tract+Florida+2020&g=040XX00US12\\$1400000&d=DEC+Demographic+Profile](https://data.census.gov/table?q=census+tract+Florida+2020&g=040XX00US12$1400000&d=DEC+Demographic+Profile).
- [47] CDC, CDC's Social Vulnerability Index, CDC SVI Documentation 2018 | Place and Health, ATSDR, 2018.
- [48] CDC, CDC's Social Vulnerability Index Data and Documentation, 2018. <https://svi.cdc.gov/SVIDataToolsDownload.html>.
- [49] COVID-19 PVI, COVID-19 Pandemic Vulnerability Index (PVI) Dashboard Public Repository, 2022. <https://github.com/COVID19PVI/data>. (Accessed 24 October 2022).
- [50] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Stat. Computing, Vienna, Austria, 2022. <http://www.R-project.org>.
- [51] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning*, Springer, New York, 2013.
- [52] J. Fox, *Applied Regression Analysis and Generalized Linear Models*, Sage Publications, 2016.
- [53] S. Unkel, C.P. Farrington, P.H. Garthwaite, C. Robertson, N. Andrews, Statistical methods for the prospective detection of infectious disease outbreaks: a review, *J. Roy. Stat. Soc. Stat. Soc.* 175 (1) (2012) 49–82.
- [54] M. Kuhn, K. Johnson, *Applied Predictive Modeling*, Springer, New York, 2013.
- [55] B. Everitt, S. Landau, M. Leese, D. Stahl, *Cluster Analysis*, Wiley, 2011.
- [56] J. Pearl, M. Glymour, N.P. Jewell, *Causal Inference in Statistics: A Primer*, John Wiley & Sons, 2016.
- [57] D. Nicholson, O.A. Vanli, S. Jung, E.E. Ozguven, A spatial regression and clustering method for developing place-specific social vulnerability indices using census and social media data, *Int. J. Disaster Risk Reduc.* 38 (2019) 101224.
- [58] D.T. Citron, C.A. Guerra, A.J. Dolgert, S.L. Wu, J.M. Henry, H.M. Sánchez, C.D.L. Smith, Comparing metapopulation dynamics of infectious diseases under different models of human movement, *Proc. Natl. Acad. Sci. USA* 118 (18) (2021) e2007488118.
- [59] M. Grímée, M.B.N. Dunbar, F. Hofmann, L. Held, Modelling the effect of a border closure between Switzerland and Italy on the spatiotemporal spread of COVID-19 in Switzerland, *Spatial statistics* 49 (2022) 100552.