Moran's Eigenvector Spatial Filtering Regression Modeling of Interacting Disasters

by

ALAYNA DAWS

(Under the Direction of Lynne Seymour)

Abstract

This study aims to quantify spatiotemporal relationships among hurricane impacts, social vulnerability, and community resilience with COVID-19 deaths per capita. Data used in this study included the Southeastern US with Hospital Service Areas (HSAs) and Hospital Referral Regions (HRRs) instead of county-level data. The Moran eigenvector spatial filtering (MESF) modeling method was employed to index and filter spatial autocorrelation (SA) from the residuals. SA was visualized to understand better the spatial structure meaningful covariates contain. The MESF model outperformed the linear model, not accounting for SA. Adjusted R^2 was 49.8% higher, and residual SE decreased 29.3% when comparing a 4-nearest-neighbor HRR to the linear model not considering SA. Analyzing cumulative COVID-19 deaths with this methodology is more effective at the HRR than at the HSA level. Additionally, vulnerability and resilience indicators were more substantial than hurricane impacts.

Index words: [Spatial Autocorrelation, Eigenvectors Spatial Filtering, Moran's I, Disasters Colliding]

Moran's Eigenvector Spatial Filtering Regression Modeling of Interacting Disasters

by

ALAYNA DAWS

B.S., University of Georgia, 2022

A Thesis Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Master of Science

Athens, Georgia

2023

Moran's Eigenvector Spatial Filtering Regression Modeling of Interacting Disasters

by

ALAYNA DAWS

Major Professor: Lynne Seymour

Committee: Jaxk Reeves
Paul Schliekelman

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
August 2023

Chapter 1

Introduction

1.1 Motivation

In 2022, an interdisciplinary team of academics from two Southeastern United States (US) universities joined forces to identify, quantify, and gain a better understanding of the relationship between compounded disturbances – extreme weather events, community vulnerability, and resilience – and infectious diseases.

The team consisted of researchers from the **University of Georgia (UGA)** Center for the Ecology of Infectious Diseases (CEID), UGA's: D.B. Warnell School of Forestry and Natural Resources, Department of Anthropology, College of Engineering Institute for Resilient Infrastructure Systems, Department of Geography and Atmospheric Sciences Program, Department of Statistics, and the **Mississippi State University**, College of Forest Resources.

It is hypothesized that the interplay between severe weather events and epidemics could be substantial and reliant on community resilience, the ability to prepare for anticipated hazards, adapt to changing conditions and withstand and recover rapidly from disruptions. The team focused on hurricane impacts, which limited the regional scope to the southeast, an area historically prone to frequent hurricanes and tornadic activity (Gensini & Brooks, 2018). Furthermore, the Southeastern US was heavily impacted by the COVID-19 pandemic. Motivated by these events, the team published a concept paper titled: **"Disasters collide at the intersection of extreme weather and infectious diseases,"** which highlighted the need for further research and analysis about the effects of compounding community disturbances and infectious diseases (Drake et al., 2023).

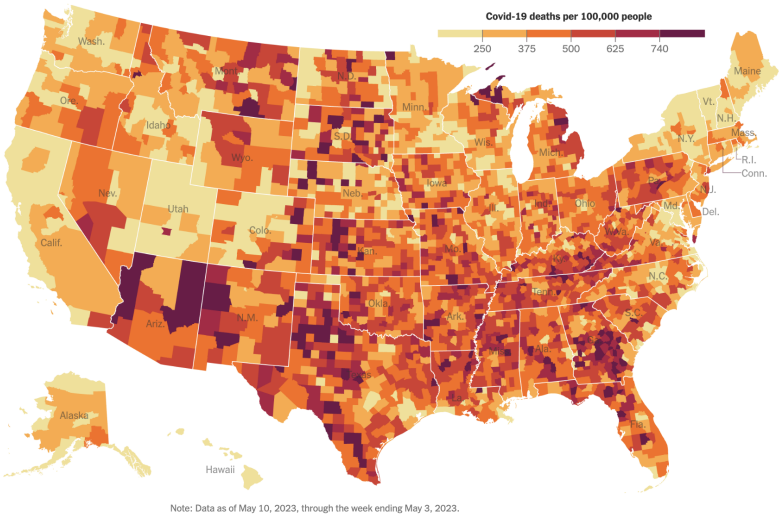

Figure 1.1: This map, which utilizes data from the Centers for Disease Control and Prevention (CDC), shows where people have died of COVID-19 at the highest rates (Lazaro et al., 2023). It illustrates that many Southeastern states have higher COVID-19 deaths per 100k people (dark purple).

Several spatial regression models were employed to quantify the spatiotemporal relationships among hurricane impacts, social vulnerability, and community resilience with COVID-19 deaths per capita. The models utilized data from the Southeastern US in HHS districts 4 and 6, excluding New Mexico, because these areas are subject to hurricanes.

1.2 Hurricane Impacts

Hurricane impacts are particularly interesting because they represent the intersection between severe weather events and increased human aggregation - posing a significant risk when infection rampantly spreads in communities. Hurricanes can damage health infrastructure, impeding the community's recovery and response time needed to curtail the further spread of disease. Hurricane intensity factors in this study are quantified as the maximum wind speed and maximum daily precipitation.

1.3 Community Vulnerability and Resilience

The University of South Carolina Hazards and Vulnerability Research Institute (HVRI) designed two analytical tools – the **Social Vulnerability Index (SoVI)®** and the **Baseline Resilience Indicators for Communities (BRIC)**, a baseline for monitoring existing natural hazards resilience – to help policymakers identify vulnerable communities and more effectively allocate resources to high-risk populations. Both variables are based on a community's ability to prepare, act, and recover from hazards.

1.3.1 Social Vulnerability Index (SoVI)

The SoVI score measures the vulnerability of all counties in the US to environmental hazards (Institute, 2015). SoVI is built from 29 socioeconomic Census variables. A lower SoVI score indicates less vulnerability, while a higher SoVI score indicates more vulnerability as defined by the socioeconomic components. Additional information about SoVI can be found in **2.2.1**.

1.3.2 Baseline Resilience Indicators for Communities (BRIC)

The Baseline Resilience Indicators for Communities (BRIC) was created to provide a baseline for measuring changes in community resiliency over time. BRIC comprises six broad capitals of community disaster resilience (Cutter et al., 2010). The capitals include Social, Economic, Community Capacity, Institutional, Infrastructural, and Environmental. A complete list of the 49 variables, split by capital, is found in **Appendix B**. A higher BRIC score is favorable and indicates higher community resilience. Additional information about BRIC can be found in **2.2.2**.

1.4 COVID-19

Cumulative COVID-19 deaths per 100,000 in 2020 and 2021 were analyzed as the response variables using Johns Hopkins University data (Hopkins, 2023). COVID-19 data have been widely reported at the county level where COVID-related deaths occurred. However, this data did not always align with the county the person resided and likely contracted COVID-19. As such, county-level analysis may misrepresent or obscure valuable information.

The following analysis takes a **Hospital Service Areas (HSAs)** and **Hospital Referral Regions (HRRs)** approach in place of county-level data. HSAs and HRRs represent healthcare utilization patterns of hospital and specialized care services. These defining regions have proven to be more reliable study regions than county-level data when seeking to understand where individuals seek and receive treatment (Skinner et al., 2022). Additional information about HRRs and HSAs can be found in **2.1.2** (HSAs) and **2.1.3** (HRRs).

Crosswalks were employed to aggregate COVID-19 death reports to HRR and HSA levels. Crosswalks are a method used to map equivalent, identical, or similar information across two or more distinct data sets (MB&C, 2022). The predictor variables were also aggregated to the HRR and HSA levels as defined by the (Sledge, 2001). Additional information about crosswalks can be found in **2.1.4**.

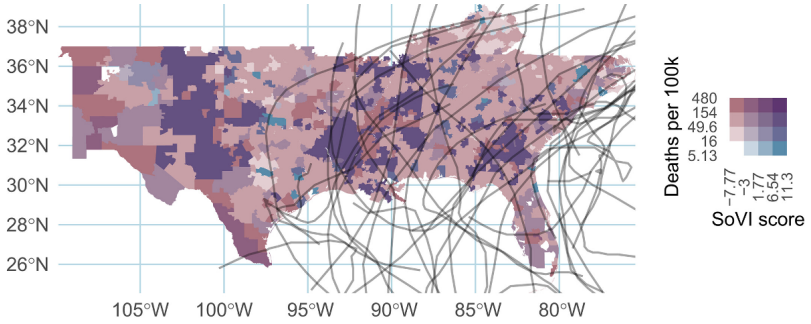

Figure 1.2: The Intersection of the Social Vulnerability Indicator (SoVI) (Institute, 2015), Hurricane Tracks, and COVID-19 Outcomes for Hospital Service Areas (HSAs) (Drake et al., 2023). The HSAs included are located in regions 4 and 6 as defined by the United States Department of Health and Human Services (HHS). The regions are colored based on SoVI score and cumulative COVID-19 deaths per 100,000. Higher SoVI scores (blue) represent greater community social vulnerability and are aggregated to HSA by county-level populations. Higher cumulative deaths per 100,000 from COVID-19 appear in mauve (top left corner of the legend color matrix). HSAs with higher COVID-19 cumulative deaths per 100,000 and higher SoVI scores appear in dark purple (top right corner of the legend color matrix). The hurricanes in $2015 - 2020$ are visualized as grey tracks (Landsea & Franklin, 2013).

1.5 Spatial Autocorrelation (SA)

Independent observations are an underlying assumption in many statistical modeling techniques. This assumption, however, is only upheld in some real-world phenomena. Spatial autocorrelation (SA) may be expected in geospatial data. The concept of SA can be related to Tobler's First Law of Geography, which states that "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). SA is a correlation occurring in geospatial data and can indicate how much a map pattern changes throughout the landscape.

1.5.1 Types of Spatial Autocorrelation (SA)

SA can be positive or negative when describing a geospatial trend. Positive spatial autocorrelation (PSA) occurs when geographically nearby values tend to be similar on a map. PSA is the most common type of SA observed. When negative spatial autocorrelation (NSA) is present, locations with large values have neighbors with small values, and when small values occur, the neighbors have large values. When SA is not observed, the map will not have a pattern, and response values will be mixed without an underlying pattern.

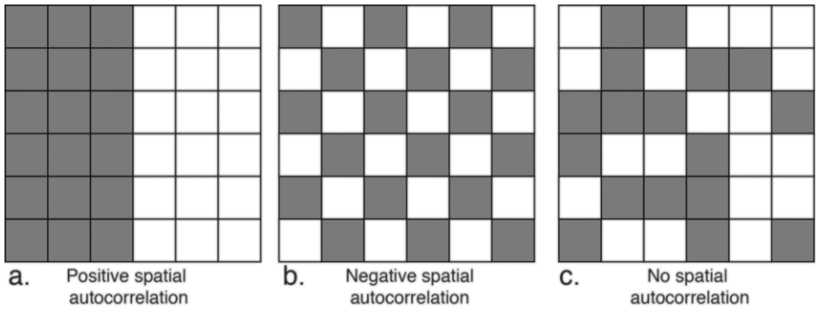

Figure 1.3: Spatial Autocorrelation Illustration (Boraks et al., 2021).

1.6 Goals

This study aims to quantify the spatiotemporal relationships among hurricane impacts, social vulnerability, and community resilience with COVID-19 deaths per capita. SA is indexed and filtered during the statistical modeling process to account for these spatiotemporal relationships more accurately.

Chapter 5

Results

5.1 Eigenvectors Extracted from Centered Spatial Weights Matrix (SWM)

5.1.1 Distance-Based $\mathbf{C}$

When a distance-based $\mathbf{C}$ is defined, the standard eigenvalue threshold is $\lambda_n > 0$, where λ_n is the nth eigenvalue. This threshold considers all elements that describe positive spatial dependence.

Hospital Referral Regions (HRRs): When the eigenvectors are extracted from the distance-based $\mathbf{C}$ at the HRR level, $16/103$ (15.5%) of the eigenvectors describe positive spatial dependence.
Hospital Service Areas (HSAs): When the eigenvectors are extracted from the distance-based $\mathbf{C}$ at the HSA level, $88/1094$ (8.04%) of the eigenvectors describe positive spatial dependence.

5.1.2 4-nearest-neighbor-based $\mathbf{C}$

In the 4-nearest-neighbor-based approach, the eigenvector extraction threshold is set to be $\lambda_n/\lambda_1 > 0.25$. The threshold $\lambda_n/\lambda_1 > 0.25$ is implemented because it attempts to capture approximately 5% of variance in the explanatory variables that is attributable to positive spatial dependence (D. Griffith & Chun, 2014).

Hospital Referral Regions (HRRs): When the eigenvectors are extracted from the 4-nearest-neighbor-based at the HRR level, $26/103$ (25.2%) of the eigenvectors exceed the threshold.
Hospital Service Areas (HSAs): When the eigenvectors are extracted from the 4-nearest-neighbor-based at the HSA level, $294/1094$ (26.9%) of the eigenvectors exceed the threshold.

5.2 Hospital Referral Regions (HRRs) Model Summary Tables

5.2.1 COVID-19 Cumulative Deaths Per 100K in 2020

Minimizing BIC

Linear ESF model at the HRR Level with the objective function: minimizing the BIC and maximizing the adjusted R^2 for the **COVID-19 Cumulative Deaths Per 100K in 2020** and **COVID-19 Cumulative Deaths Per 100K in 2021** as the responses for a linear model without higher order or interaction terms, distance-based $\mathbf{C}$, and 4NN-based $\mathbf{C}$. ThePredictors selected in all HRR modeling are bolded blue.

Minimizing BIC 2020 HRR Model Fit	Variables Selected	Adj. R^2	Residual SE	AIC	BIC	Moran's I and SE	Number of Eigenvectors Selected
Linear Model without Higher Order or Interaction Terms	**SoVI**, BRIC economic, **BRIC infrastructure**, **BRIC community**, BRIC institutional, BRIC environment	0.502	34.04	1027.75	1048.83	NA	NA
MESF Disance-Based C	**SoVI**, **BRIC infrastructure**, **BRIC community**, Wind 2019, Rain 2020	0.692	26.78	982.8	1017.05	0.452, SE: 27.56	6/16
MESF 4NN-Based C	**SoVI**, **BRIC infrastructure**, **BRIC community**, BRIC environment, Rain 2019, Rain 2020, Wind 2020	0.731	25.05	973.26	1020.68	0.754 SE: 30.02	9/26

Figure 5.1: HRRs Model Summary Table **COVID-19 Cumulative Deaths Per 100K in 2020** minimizing the BIC. The BIC is minimized when the MESF utilizing the distance-based $\mathbf{C}$ modeling technique is completed.

The BIC is 3% lower for the distance-based $\mathbf{C}$ compared to the BIC for the linear model without higher order or interaction terms. The MESF utilizing the 4NN-based $\mathbf{C}$ resulted in the lowest residual SE, lowest AIC, and greatest adjusted R^2. The Moran's I value for the distance-based $\mathbf{C}$ is in the weak range. However, the Moran's I value for the 4NN-based $\mathbf{C}$ is in the strong range. The distance-based $\mathbf{C}$ resulted in fewer eigenvectors being selected compared to the 4NN-based $\mathbf{C}$. The residual SE decreased 21.3% when comparing the distance-based $\mathbf{C}$ to the linear model without higher order or interaction terms. All of the three models fit with minimizing the BIC criterion, resulted in the selection of **SoVI**, **BRIC infrastructure**, and **BRIC community**.

The distance-based $\mathbf{C}$ method resulted in the selection of **SoVI**, **BRIC infrastructure**, **BRIC community**, **Wind 2019**, and **Rain 2020** as important predictors.

Maximizing Adjusted R^2

Maximizing R^2 2020 HRR Model Fit	Variables Selected	Adj. R^2	Residual SE	AIC	BIC	Moran's I and SE	Number of Eigenvectors Selected
Linear Model without Higher Order or Interaction Terms	SoVI, BRIC economic, BRIC infrastructure, BRIC community, BRIC institutional, BRIC environment	0.502	34.04	1027.75	1048.83	NA	NA
MESF Disance-Based C	SoVI, BRIC infrastructure, BRIC community, BRIC environment, Rain 2019, Rain 2020, Wind 2019	0.719	25.59	978.44	1028.5	0.450, SE: 30.28	10/16
MESF 4NN-Based C	SoVI, BRIC infrastructure, BRIC community, BRIC environment, Rain 2019, Rain 2020, Wind 2020	0.752	24.06	967.24	1022.57	0.735, SE: 30.91	12/26

Figure 5.2: HRRs Model Summary Table **COVID-19 Cumulative Deaths Per 100K in 2020** maximizing the adjusted R^2. The adjusted R^2 is maximized when the MESF utilizing the 4NN-based **C** modeling technique is completed.

The adjusted R^2 is 49.8% higher for the 4NN-based **C** compared to the adjusted R^2 for the linear model without higher order or interaction terms. The MESF utilizing the 4NN-based **C** also resulted in the lowest residual SE, AIC, and BIC. The Moran's I value for the distance-based **C** is in the weak range. However, the Moran's I value for the 4NN-based **C** is in the strong range. The distance-based **C** resulted in fewer eigenvectors being selected compared to the 4NN-based **C**. The residual SE decreased 29.3% when comparing the 4NN-based **C** to the linear model without higher order or interaction terms. All of the three models fit with maximizing the adjusted R^2 criterion, resulted in the selection of **SoVI**, **BRIC infrastructure**, **BRIC community**, and **BRIC environment**.

The 4NN-based **C** method resulted in the selection of **SoVI**, **BRIC infrastructure**, **BRIC community**, **BRIC environment**, **Rain 2019**, **Rain 2020**, and **Wind 2020** as important predictors.

5.3 Hospital Service Areas (HSAs) Model Summary Tables

5.3.1 COVID-19 Cumulative Deaths Per 100K in 2020

Minimizing BIC

Linear ESF model at the HSA Level with the objective function: minimizing the BIC and maximizing the adjusted R^2 for the **COVID-19 Cumulative Deaths Per 100K in 2020** as the response for a linear model without higher order or interaction terms, distance-based **C**, and 4NN-based **C**.

Minimizing BIC 2020 HSA Model Fit	Variables Selected	Adj. R^2	Residual SE	AIC	BIC	Moran's I and SE	Number of Eigenvectors Selected
Linear Model without Higher Order or Interaction Terms	SoVI, BRIC economic, BRIC infrastructure, BRIC community, BRIC environment, Rain 2019, Wind 2019, Wind 2020	0.29	57.34	11974.33	12024.31	NA	NA
MESF Disance-Based C	SoVI, BRIC economic, BRIC infrastructure, BRIC community, BRIC environment, Rain 2019, Wind 2019, Wind 2020	0.585	43.82	11413.72	11603.63	0.212, SE: 41.15	28/88
MESF 4NN-Based C	SoVI, BRIC infrastructure, BRIC community, BRIC environment, Rain 2019, Rain 2020, Wind 2019	0.594	43.35	11400.01	11639.9	0.879, SE: 40.29	39/294

Figure 5.3: HSAs Model Summary Table **COVID-19 Cumulative Deaths Per 100K in 2020** minimizing the BIC. The BIC is minimized when the MESF utilizing the distance-based **C** modeling technique is completed.

The BIC is 3.5% lower for the distance-based **C** compared to the BIC for the linear model without higher order or interaction terms. The MESF utilizing the 4NN-based **C** resulted in the lowest residual SE, lowest AIC, and greatest adjusted R^2. The Moran's I value for the distance-based **C** is below the weak range. However, the Moran's I value for the 4NN-based **C** is in the strong range. The distance-based **C** resulted in fewer eigenvectors being selected compared to the 4NN-based **C**. The residual SE decreased 23.6% when comparing the distance-based **C** to the linear model without higher order or interaction terms. All of the three models fit with minimizing the BIC criterion, resulted in the selection of **SoVI**, **BRIC infrastructure**, **BRIC community**, and **Wind 2019**.

The distance-based **C** method resulted in the selection of **SoVI**, **BRIC economic**, **BRIC infrastructure**, **BRIC community**, **BRIC environment**, **Rain 2019**, **Wind 2019**, and **Wind 2020** as important predictors.

Maximizing Adjusted R^2

Maximizing R^2 2020 HSA Model Fit	Variables Selected	Adj. R^2	Residual SE	AIC	BIC	Moran's I and SE	Number of Eigenvectors Selected
Linear Model without Higher Order or Interaction Terms	SoVI, BRIC economic, BRIC infrastructure, BRIC community, BRIC environment, Rain 2019, Wind 2019, Wind 2020	0.29	57.34	11974.33	12024.31	NA	NA
MESF Disance-Based C	SoVI, BRIC infrastructure, BRIC community, Rain 2019, Wind 2019	0.611	42.46	11368.43	11683.28	0.187, SE: 43.77	56/88
MESF 4NN-Based C	SoVI, BRIC infrastructure, BRIC community, Rain 2020, Wind 2019	0.683	38.32	11237.09	12076.68	0.830, SE: 50.05	161/294

Figure 5.4: HSAs Model Summary Table **COVID-19 Cumulative Deaths Per 100K in 2020** maximizing the adjusted R^2. The adjusted R^2 is maximized when the MESF utilizing the 4NN-based **C** modeling technique is completed.

The adjusted R^2 is 135.5% higher for the 4NN-based **C** compared to the adjusted R^2 for the linear model without higher order or interaction terms. The MESF utilizing the 4NN-based **C** also resulted in the lowest residual SE and AIC. The distance-based **C** resulted in the lowest BIC. The Moran's I value for the distance-based **C** is below the weak range. However, the Moran's I value for the 4NN-based **C** is in the strong range. The distance-based **C** resulted in fewer eigenvectors being selected compared to the 4NN-based **C**. The residual SE decreased 33.1% when comparing the 4NN-based **C** to the linear model without higher order or interaction terms. All of the three models fit with maximizing the adjusted R^2 criterion, resulted in the selection of **SoVI**, **BRIC infrastructure**, **BRIC community**, and **Wind 2019**.

The 4NN-based **C** method resulted in the selection of **SoVI**, **BRIC infrastructure**, **BRIC community**, **Rain 2020**, and **Wind 2019** as important predictors.

5.3.2 COVID-19 Cumulative Deaths Per 100K in 2021

Minimizing BIC

Linear ESF model at the HSA Level with the objective function: minimizing the BIC and maximizing the adjusted R^2 for the **COVID-19 Cumulative Deaths Per 100K in 2021** as the response for a linear model without higher order or interaction terms, distance-based **C**, and 4NN-based **C**.

Minimizing BIC 2021 HSA Model Fit	Variables Selected	Adj. R^2	Residual SE	AIC	BIC	Moran's I and SE	Number of Eigenvectors Selected
Linear Model without Higher Order or Interaction Terms	**SoVI**, BRIC economic, **BRIC infrastructure**, **BRIC community**, BRIC institutional, BRIC environment, Rain 2019, Rain 2020, **Wind 2019**, Wind 2020	0.414	68.33	12360.36	12420.33	NA	NA
MESF Disance-Based C	**SoVI**, BRIC economic, **BRIC infrastructure**, **BRIC community**, BRIC institutional, BRIC environment, Rain 2019, Rain 2020, **Wind 2019**	0.62	55.05	11910.32	12085.23	0.268, SE: 49.63	24/88
MESF 4NN-Based C	**SoVI**, BRIC economic, **BRIC infrastructure**, **BRIC community**, BRIC institutional, BRIC environment, Rain 2019, Rain 2020, **Wind 2019**, Wind 2020	0.65	52.83	11841.95	12131.82	0.859, SE: 47.36	46/294

Figure 5.5: HSAs Model Summary Table **COVID-19 Cumulative Deaths Per 100K in 2021** minimizing the BIC. The BIC is minimized when the MESF utilizing the distance-based **C** modeling technique is completed.

The BIC is 2.7% lower for the distance-based **C** compared to the BIC for the linear model without higher order or interaction terms. The MESF utilizing the 4NN-based **C** resulted in the lowest residual SE, lowest AIC, and greatest adjusted R^2. The Moran's I value for the distance-based **C** is in the weak range. However, the Moran's I value for the 4NN-based **C** is in the strong range. The distance-based **C** resulted in fewer eigenvectors being selected compared to the 4NN-based **C**. The residual SE decreased 19.4% when comparing the distance-based **C** to the linear model without higher order or interaction terms. All of the three models fit with minimizing the BIC criterion resulting in the selection of **SoVI**, **BRIC infrastructure**, **BRIC community**, and **Wind 2019**.

The distance-based **C** method resulted in the selection of **SoVI**, **BRIC economic**, **BRIC infrastructure**, **BRIC community**, **BRIC institutional**, **BRIC environment**, **Rain 2019**, **Rain 2020**, and **Wind 2019** as important predictors.

Maximizing Adjusted R^2

Maximizing R^2 2021 HSA Model Fit	Variables Selected	Adj. R^2	Residual SE	AIC	BIC	Moran's I and SE	Number of Eigenvectors Selected
Linear Model without Higher Order or Interaction Terms	**SoVI**, BRIC economic, **BRIC infrastructure**, **BRIC community**, BRIC institutional, BRIC environment, Rain 2019, Rain 2020, **Wind 2019**, Wind 2020	0.414	68.33	12360.36	12420.33	NA	NA
MESF Disance-Based C	**SoVI**, **BRIC infrastructure**, **BRIC community**, BRIC institutional, BRIC environment, Rain 2019, **Wind 2019**	0.649	52.93	11855.71	12195.54	0.276, SE: 58.95	59/88
MESF 4NN-Based C	**SoVI**, BRIC economic, **BRIC infrastructure**, **BRIC community**, BRIC environment, Rain 2019, **Wind 2019**	0.746	44.99	11615.59	12630.11	0.809, SE: 63.01	190/294

Figure 5.6: HSAs Model Summary Table **COVID-19 Cumulative Deaths Per 100K in 2021** maximizing the adjusted R^2. The adjusted R^2 is maximized when the MESF utilizing the 4NN-based **C** modeling technique is completed.

The adjusted R^2 is 80.2% higher for the 4NN-based **C** compared to the adjusted R^2 for the linear model without higher order or interaction terms. The MESF utilizing the 4NN-based **C** also resulted in the lowest residual SE and AIC. The distance-based **C** resulted in the lowest BIC. The Moran's I value for the distance-based **C** is in the weak range. However, the Moran's I value for the 4NN-based **C** is in the strong range. The distance-based **C** resulted in fewer eigenvectors being selected compared to the 4NN-based **C**. The residual SE decreased 34.2% when comparing the 4NN-based **C** to the linear model without higher order or interaction terms. All of the three models fit with maximizing the adjusted R^2 criterion, resulted in the selection of **SoVI**, **BRIC infrastructure**, **BRIC community**, and **Wind 2019**.

The 4NN-based **C** method resulted in the selection of **SoVI**, **BRIC infrastructure**, **BRIC community**, **BRIC environment**, **Rain 2019**, and **Wind 2019** as important predictors.

5.4 Model Fitting

5.4.1 HRRs COVID-19 Cumulative Deaths Per 100K in 2020

The linear model with **COVID-19 Cumulative Deaths Per 100K in 2020** as the response that does not account for SA is provided. The models under the distance-based C and 4NN-based C fit with both objective criteria are also provided. It should be noted that the eigenvectors shown for the models under the distance-based C and 4NN-based C all of the eigenvectors selected in the model.

Linear Model

Linear model for the HRR data with **COVID-19 Cumulative Deaths Per 100K in 2020** as the response:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 99.163 + 16.713 * SoVI - 11.392 * BRIC.Economic - 11.056 * BRIC.Infrastructure + 18.081 * BRIC.Community + 8.095 * BRIC.Institutional - 10.254 * BRIC.Environment + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	99.163	5.338	18.577	0.000
SoVI	16.713	2.179	7.671	0.000
BRIC Economic	−11.392	4.898	−2.326	0.022
BRIC Infrastructure	−11.056	5.596	−1.976	0.051
BRIC Community	18.081	5.058	3.575	0.001
BRIC Institutional	8.095	3.312	2.444	0.016
BRIC Environment	−10.254	3.898	−2.631	0.010

The 4NN-Based C - Maximizing Adjusted R^2

Linear ESF model at the HRR level with the objective function: maximizing the adjusted R^2 for ESF forward stepwise method and **COVID-19 Cumulative Deaths Per 100K in 2020** as the response for the 4NN-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 105.305 + 12.886 * SoVI - 8.852 * BRIC.Infrastructure + 8.366 * BRIC.Community - 6.551 * BRIC.Environment + 1.535 * Wind.2020 - 0.078 * Rain.2019 - 0.166 * Rain.2020 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	105.305	10.960	9.608	0.000
SoVI	12.886	1.848	6.973	0.000
BRIC Infrastructure	−8.852	4.050	−2.186	0.032
BRIC Community	8.366	4.263	1.963	0.053
BRIC Environment	−6.551	2.800	−2.339	0.022
Rain 2019	−0.078	0.056	−1.410	0.162
Wind 2020	1.535	0.357	4.304	0.000
Rain 2020	−0.166	0.070	−2.385	0.019

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
2	−146.006	27.971	−5.220	0.000
10	140.610	26.904	5.226	0.000
9	108.700	27.062	4.017	0.000
14	78.597	25.354	3.100	0.003
24	74.214	25.695	2.888	0.005
3	−100.848	30.595	−3.296	0.001
19	−53.638	24.656	−2.175	0.032
1	−72.878	30.205	−2.413	0.018
21	66.430	25.684	2.586	0.011
8	−66.457	31.400	−2.116	0.037
15	52.158	25.370	2.056	0.043
22	−54.557	30.492	−1.789	0.077

The Distance-Based C - Maximizing Adjusted R^2

Linear ESF model at the HRR level with the objective function: maximizing the adjusted R^2 for ESF forward stepwise method and **COVID-19 Cumulative Deaths Per 100K in 2020** as the response for the distance-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 109.392 + 13.294 * SoVI - 5.857 * BRIC.Infrastructure + 13.159 * BRIC.Community - 4.753 * BRIC.Environment + 2.090 * Wind.2020 - 0.091 * Rain.2019 - 0.149 * Rain.2020 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	109.392	12.420	8.808	0.000
SoVI	13.294	1.953	6.807	0.000
BRIC Infrastructure	−5.857	4.436	−1.320	0.190
BRIC Community	13.159	3.951	3.331	0.001
BRIC Environment	−4.753	3.652	−1.302	0.197
Wind 2019	2.090	0.661	3.161	0.002
Rain 2019	−0.091	0.049	−1.879	0.064
Rain 2020	−0.149	0.067	−2.221	0.029

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	t- value	p-value
1	−175.813	30.753	−5.717	0.000
9	−170.394	29.139	−5.848	0.000
3	107.322	30.236	3.549	0.001
5	82.466	26.639	3.096	0.003
11	63.397	26.749	2.370	0.020
7	52.995	28.054	1.889	0.062
12	−53.569	28.084	−1.907	0.060
15	−47.901	26.793	−1.788	0.077
6	−45.896	28.642	−1.602	0.113
16	−34.455	27.602	−1.248	0.215

The 4NN-Based C - Minimizing the BIC

Linear ESF model at the HRR level with the objective function: minimizing the BIC for ESF forward stepwise method and **COVID-19 Cumulative Deaths Per 100K in 2020** as the response for the 4NN-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 119.033 + 11.899 * SoVI - 9.206 * BRIC.Infrastructure + 11.095 * BRIC.Community - 4.516 * BRIC.Environment + 1.401 * Wind.2020 - 0.154 * Rain.2019 - 0.179 * Rain.2020 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	t- value	p-value
(Intercept)	119.033	10.044	11.851	0.000
SoVI	11.899	1.843	6.456	0.000
BRIC Infrastructure	−9.206	4.104	−2.243	0.027
BRIC Community	11.095	4.248	2.612	0.011
BRIC Environment	−4.516	2.832	−1.595	0.114
Rain 2019	−0.154	0.049	−3.172	0.002
Wind 2020	1.401	0.339	4.137	0.000
Rain 2020	−0.179	0.072	−2.498	0.014

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
2	−160.872	28.583	−5.628	0.000
10	149.876	27.736	5.404	0.000
9	102.782	27.760	3.703	0.000
14	81.943	26.077	3.142	0.002
24	79.103	26.603	2.973	0.004
3	−92.364	30.750	−3.004	0.003
19	−58.339	25.618	−2.277	0.025
1	−70.699	31.442	−2.249	0.027
21	55.506	26.416	2.101	0.039

The Distance-Based C - Minimizing the BIC

Linear ESF model at the HRR level with the objective function: minimizing the BIC for ESF forward stepwise method and **COVID-19 Cumulative Deaths Per 100K in 2020** as the response for the distance-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 110.113 + 13.678 * SoVI - 7.377 * BRIC.Infrastructure + 17.106 * BRIC.Community + 0.967 * Wind.2019 - 0.118 * Rain.2020 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	110.113	11.256	9.783	0.000
SoVI	13.678	1.736	7.879	0.000
BRIC Infrastructure	−7.377	4.355	−1.694	0.094
BRIC Community	17.106	3.729	4.588	0.000
Wind 2019	0.967	0.488	1.980	0.051
Rain 2020	−0.118	0.067	−1.771	0.080

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
9	−158.942	29.411	−5.404	0.000
1	−155.378	29.090	−5.341	0.000
3	107.402	30.526	3.518	0.001
5	85.937	27.722	3.100	0.003
11	70.621	27.583	2.560	0.012
7	64.643	28.234	2.290	0.024

5.4.2 HSAs COVID-19 Cumulative Deaths Per 100K in 2020

The linear model with **COVID-19 Cumulative Deaths Per 100K in 2020** as the response that does not account for SA is provided. The models under the distance-based C and 4NN-based C fit with both objective criteria are also provided. At the HSA level, the number of eigenvectors selected varies from 28 to 161. As a result, only the first 10 selected eigenvectors, ranked by the absolute value of their slope coefficient, are shown for each model.

Linear Model

Linear model for the HSA data with **COVID-19 Cumulative Deaths Per 100K in 2020** as the response:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 86.122 + 10.873 * SoVI - 9.525 * BRIC.Economic - 6.301 * BRIC.Infrastructure + 12.146 * BRIC.Community - 8.410 * BRIC.Environment + 3.339 * Wind.2019 - 0.302 * Wind.2020 - 0.121 * Rain.2019 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	86.122	5.033	17.111	0.000
SoVI	10.873	0.849	12.806	0.000
BRIC Economic	−9.525	2.098	−4.539	0.000
BRIC Infrastructure	−6.301	2.119	−2.973	0.003
BRIC Community	12.146	2.214	5.487	0.000
BRIC Environment	−8.410	2.360	−3.563	0.000
Wind 2019	3.339	0.396	8.436	0.000
Wind 2020	−0.302	0.224	−1.350	0.177
Rain 2019	−0.121	0.036	−3.392	0.001

The 4NN-Based C - Maximizing Adjusted R^2

Linear ESF model at the HSA level with the objective function: maximizing the adjusted R^2 for ESF forward stepwise method and **COVID-19 Cumulative Deaths Per 100K in 2020** as the response for the 4NN-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 111.84 + 9.88 * SoVI - 9.41 * BRIC.Infrastructure + 11.21 * BRIC.Community - 0.82 * Wind.2019 + 0.10 * Rain.2020 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	111.841	7.029	15.912	0.000
SoVI	9.879	0.685	14.413	0.000
BRIC Infrastructure	−9.407	1.763	−5.337	0.000
BRIC Community	11.209	1.819	6.162	0.000
Wind 2019	−0.820	0.419	−1.958	0.050
Rain 2020	0.104	0.045	2.318	0.021

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
11	555.603	40.296	13.788	0.000
8	−452.103	43.300	−10.441	0.000
21	389.416	40.170	9.694	0.000
9	316.566	41.197	7.684	0.000
27	265.203	41.983	6.317	0.000
3	−274.494	40.702	−6.744	0.000
71	−221.473	39.152	−5.657	0.000
57	−255.333	39.746	−6.424	0.000
12	−229.436	39.267	−5.843	0.000
58	217.610	38.492	5.653	0.000

The Distance-Based C Maximizing Adjusted R^2

Linear ESF model at the HSA level with the objective function: maximizing the adjusted R^2 for ESF forward stepwise method and **COVID-19 Cumulative Deaths Per 100K in 2020** as the response for the distance-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 90.02 + 8.47 * SoVI - 9.37 * BRIC.Infrastructure + 9.47 * BRIC.Community + 1.04 * Wind.2019 + 0.12 * Rain.2019 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	90.018	4.706	19.127	0.000
SoVI	8.468	0.747	11.342	0.000
BRIC Infrastructure	−9.374	1.754	−5.345	0.000
BRIC Community	9.466	2.233	4.239	0.000
Wind 2019	1.044	0.377	2.768	0.006
Rain 2019	0.125	0.040	3.138	0.002

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
9	724.490	45.399	15.958	0.000
5	369.141	49.021	7.530	0.000
2	−356.856	50.241	−7.103	0.000
13	392.431	43.843	8.951	0.000
25	−309.593	43.137	−7.177	0.000
3	−417.178	65.252	−6.393	0.000
24	−239.292	43.870	−5.455	0.000
1	−127.104	48.060	−2.645	0.008
65	207.880	42.708	4.867	0.000
15	218.447	44.532	4.905	0.000

The 4NN-Based C - Minimizing the BIC

Linear ESF model at the HSA level with the objective function: minimizing the BIC for ESF forward stepwise method and **COVID-19 Cumulative Deaths Per 100K in 2020** as the response for the 4NN-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 79.98 + 10.06 * SoVI - 10.02 * BRIC.Infrastructure + 13.39 * BRIC.Community - 3.51 * BRIC.Environment + 1.63 * Wind.2019 - 0.08 * Rain.2019 + 0.16 * Rain.2020 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	79.981	5.290	15.120	0.000
SoVI	10.065	0.663	15.186	0.000
BRIC Infrastructure	−10.023	1.691	−5.928	0.000
BRIC Community	13.387	1.744	7.675	0.000
BRIC Environment	−3.505	1.861	−1.884	0.060
Wind 2019	1.626	0.301	5.407	0.000
Rain 2019	−0.077	0.028	−2.788	0.005
Rain 2020	0.162	0.038	4.238	0.000

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
11	491.798	44.501	11.051	0.000
8	−343.682	46.329	−7.418	0.000
21	359.490	44.721	8.039	0.000
9	267.665	44.891	5.963	0.000
27	275.105	46.366	5.933	0.000
3	−205.174	44.785	−4.581	0.000
57	−274.432	44.256	−6.201	0.000
71	−255.909	44.002	−5.816	0.000
12	−226.934	44.110	−5.145	0.000
58	216.315	43.716	4.948	0.000

The Distance-Based C - Minimizing the BIC

Linear ESF model at the HSA level with the objective function: minimizing the BIC for ESF forward stepwise method and **COVID-19 Cumulative Deaths Per 100K in 2020** as the response for the distance-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 95.55 + 8.60 * SoVI + 2.96 * BRIC.Economic - 8.51 * BRIC.Infrastructure + 15.89 * BRIC.Community - 3.99 * BRIC.Environment + 2.67 * Wind.2019 - 1.01 * Wind.2020 + 0.05 * Rain.2019 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	95.550	4.534	21.072	0.000
SoVI	8.601	0.719	11.961	0.000
BRIC Economic	2.964	1.835	1.616	0.106
BRIC Infrastructure	−8.514	1.769	−4.814	0.000
BRIC Community	15.887	2.189	7.258	0.000
BRIC Environment	−3.992	2.221	−1.798	0.072
Wind 2019	2.665	0.369	7.227	0.000
Rain 2019	0.052	0.037	1.412	0.158
Wind 2020	−1.010	0.213	−4.749	0.000

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
9	681.113	48.287	14.106	0.000
3	−511.982	68.177	−7.510	0.000
13	419.943	45.517	9.226	0.000
25	−324.040	44.474	−7.286	0.000
5	405.676	54.351	7.464	0.000
24	−277.178	45.255	−6.125	0.000
2	−265.035	53.526	−4.952	0.000
65	234.926	44.113	5.326	0.000
1	−166.057	50.559	−3.284	0.001
12	130.394	49.949	2.611	0.009

5.4.3 HSAs COVID-19 Cumulative Deaths Per 100K in 2021

The linear model with **COVID-19 Cumulative Deaths Per 100K in 2020** as the response that does not account for SA is provided. The models under the distance-based C and 4NN-based C fit with both objective criteria are also provided. At the HSA level, the number of eigenvectors selected varies from 24 to 190. As a result, only the first 10 selected eigenvectors, ranked by the absolute value of their slope coefficient, are shown for each model.

Linear Model

Linear model for the HSA data with **COVID-19 Cumulative Deaths Per 100K in 2021** as the response:

Model COVID-19 Cumulative Deaths Per 100K in 2021:
$\widehat{COVID-19} = 242.057 + 7.598 * SoVI - 8.492 * BRIC.Economic - 21.149 * BRIC.Infrastructure + 42.049 * BRIC.Community - 5.782 * BRIC.Instiutional - 14.485 * BRIC.Environment - 2.009 * Wind.2019 - 0.948 * Wind.2020 + 0.1301 * Rain.2019 - 0.114 * Rain.2020 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	242.057	7.450	32.498	0.000
SoVI	7.598	1.019	7.455	0.000
BRIC Economic	−8.492	2.503	−3.393	0.001
BRIC Infrastructure	−21.149	2.597	−8.142	0.000
BRIC Community	42.049	2.705	15.545	0.000
BRIC Instiutional	−5.782	2.575	−2.246	0.025
BRIC Environment	−14.485	2.840	−5.100	0.000
Wind 2019	−2.009	0.477	−4.209	0.000
Wind 2020	−0.948	0.295	−3.210	0.001
Rain 2019	0.1301	0.0430	3.040	0.002
Rain 2020	−0.114	0.053	−2.162	0.031

The 4NN-Based C - Maximizing Adjusted R^2

Linear ESF model at the HSA level with the objective function: maximizing the adjusted R^2 for ESF forward stepwise method and **COVID-19 Cumulative Deaths per 100K in 2021** as the response for the 4NN-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2021:
$\widehat{COVID-19} = 215.26 + 12.32 * SoVI - 7.16 * BRIC.Economic - 11.67 * BRIC.Infrastructure + 19.08 * BRIC.Community - 9.65 * BRIC.Environment - 2.74 * Wind.2019 + 0.24 * Rain.2019 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	215.264	6.702	32.118	0.000
SoVI	12.318	0.970	12.695	0.000
BRIC Economic	−7.159	2.438	−2.937	0.003
BRIC Infrastructure	−11.667	2.190	−5.328	0.000
BRIC Community	19.083	2.900	6.581	0.000
BRIC Environment	−9.653	2.947	−3.275	0.001
Wind 2019	−2.740	0.501	−5.468	0.000
Rain 2019	0.241	0.049	4.903	0.000

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
21	713.284	47.614	14.981	0.000
24	321.552	50.681	6.345	0.000
1	−537.202	48.473	−11.083	0.000
50	402.121	46.247	8.695	0.000
38	264.346	46.711	5.659	0.000
23	−239.105	55.803	−4.285	0.000
58	298.510	45.996	6.490	0.000
71	278.539	46.478	5.993	0.000
57	−254.662	45.802	−5.560	0.000
18	370.954	47.546	7.802	0.000

The Distance-Based C - Maximizing Adjusted R^2

Linear ESF model at the HSA level with the objective function: maximizing the adjusted R^2 for ESF forward stepwise method and **COVID-19 Cumulative Deaths per 100K in 2021** as the response for the distance-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2021:
$\widehat{COVID-19} = 208.17 + 14.10 * SoVI - 18.29 * BRIC.Infrastructure + 21.92 * BRIC.Community - 7.48 * BRIC.Institutional + 5.64 * BRIC.Environment - 2.20 * Wind.2019 + 0.19 * Rain.2019 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	208.165	8.920	23.336	0.000
SoVI	14.095	0.898	15.697	0.000
BRIC Infrastructure	−18.292	2.267	−8.069	0.000
BRIC Community	21.925	2.959	7.410	0.000
BRIC Institutional	−7.485	2.719	−2.753	0.006
BRIC Environment	5.639	3.189	1.769	0.077
Wind 2019	−2.200	0.627	−3.506	0.000
Rain 2019	0.189	0.043	4.403	0.000

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
5	858.365	65.124	13.181	0.000
3	949.244	103.360	9.184	0.000
4	−316.566	110.990	−2.852	0.004
28	−336.452	54.541	−6.169	0.000
15	359.815	56.607	6.356	0.000
55	320.795	54.302	5.908	0.000
10	411.155	59.075	6.960	0.000
11	−420.412	80.110	−5.248	0.000
2	−518.083	75.520	−6.860	0.000
18	297.056	53.578	5.544	0.000

The 4NN-Based C - Minimizing the BIC

Linear ESF model at the HSA level with the objective function: minimizing the BIC for ESF forward stepwise method and **COVID-19 Cumulative Deaths per 100K in 2021** as the response for the 4NN-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2021:
$\widehat{COVID-19} = 237.67 + 9.72 * SoVI - 9.39 * BRIC.Economic - 15.35 * BRIC.Infrastructure + 35.70 * BRIC.Community - 4.24 * BRIC.Institutional - 13.38 * BRIC.Environment - 3.36 * Wind.2019 - 0.53 * Wind.2020 + 0.36 * Rain.2019 - 0.14 * Rain.2020 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	237.671	6.850	34.698	0.000
SoVI	9.715	0.853	11.395	0.000
BRIC Economic	−9.388	2.278	−4.121	0.000
BRIC Infrastructure	−15.354	2.184	−7.029	0.000
BRIC Community	35.702	2.362	15.117	0.000
BRIC Institutional	−4.240	2.233	−1.898	0.058
BRIC Environment	−13.384	2.405	−5.564	0.000
Wind 2019	−3.364	0.443	−7.587	0.000
Rain 2019	0.360	0.046	7.850	0.000
Wind 2020	−0.531	0.271	−1.964	0.050
Rain 2020	−0.139	0.048	−2.908	0.004

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
21	589.040	55.128	10.685	0.000
1	−400.521	54.938	−7.290	0.000
24	321.524	60.169	5.344	0.000
50	349.811	53.865	6.494	0.000
23	−317.496	58.753	−5.404	0.000
71	306.004	53.951	5.672	0.000
58	276.764	53.442	5.179	0.000
38	260.377	54.730	4.758	0.000
74	284.503	54.345	5.235	0.000
57	−236.317	54.310	−4.351	0.000

The Distance-Based C - Minimizing the BIC

Linear ESF model at the HSA level with the objective function: minimizing the BIC for ESF forward stepwise method and **COVID-19 Cumulative Deaths per 100K in 2021** as the response for the distance-based **C**:

Model COVID-19 Cumulative Deaths Per 100K in 2021:
$\widehat{COVID-19} = 220.19 + 13.70 * SoVI - 3.94 * BRIC.Economic - 20.45 * BRIC.Infrastructure + 25.93 * BRIC.Community - 8.08 * BRIC.Institutional - 3.81 * BRIC.Environment - 1.76 * Wind.2019 + 0.12 * Rain.2019 - 0.11 * Rain.2020 + E_k\hat{\gamma}_k$

Estimated Coefficients on X:

Predictor	β Estimate	SE	$t-value$	$p-value$
(Intercept)	220.192	9.226	23.868	0.000
SoVI	13.702	0.901	15.214	0.000
BRIC Economic	−3.942	2.145	−1.838	0.066
BRIC Infrastructure	−20.446	2.183	−9.367	0.000
BRIC Community	25.928	2.735	9.480	0.000
BRIC Institutional	−8.077	2.241	−3.603	0.000
BRIC Environment	−3.811	2.750	−1.386	0.166
Wind 2019	−1.759	0.560	−3.142	0.002
Rain 2019	0.116	0.036	3.205	0.001
Rain 2020	−0.107	0.046	−2.329	0.020

Estimated Coefficients on Moran's Eigenvectors:

EV # from **C**	γ Estimate	SE	$t-value$	$p-value$
5	769.950	65.426	11.768	0.000
3	735.018	95.424	7.703	0.000
4	−288.918	95.220	−3.034	0.002
28	−367.430	56.256	−6.531	0.000
6	229.863	61.550	3.735	0.000
15	376.717	58.670	6.421	0.000
55	296.641	56.265	5.272	0.000
29	252.381	56.990	4.428	0.000
10	364.055	59.465	6.122	0.000
2	−409.864	71.218	−5.755	0.000

5.5 Spatially Dependent Component Maps - Distance-Based C

The following 6 figures provide a visualization of the residual spatial process. For each of the figures, the number of dots are the number of either hospital referral regions (HRRs) or hospital service areas (HSAs). The values mapped are the estimated spatially dependent components in the residuals, $\boldsymbol{E\gamma}$, at each respective HRR or HSA. $\boldsymbol{E\gamma}$ is made up of the eigenvectors selected in each respective forward stepwise modeling process. All of the plots are from the distance-based **C** ESF modeling and utilize the `plot_s` function in the `spmoran` R package. Additional information about the spatially dependent component maps can be found in the **Methods**.

- **Positive $\boldsymbol{E\gamma}$** values indicate that the HRR/HSA has greater **COVID-19, Cumulative Deaths Per 100K in 2020/2021** than would have been estimated by a typical linear model that did not account for SA.
- Values around 0 indicate that, after accounting for SA occurring in the residuals, the response is not impacted heavily by the spatially dependent component in the residuals.
- **Negative $\boldsymbol{E\gamma}$** values indicate that the HRR/HSA has fewer **COVID-19 Cumulative Deaths Per 100K in 2020/2021** than would have been estimated by a typical linear model that did not account for SA.

5.5.1 Hospital Referral Regions (HRRs)

COVID-19 Cumulative Deaths Per 100K in 2020 - Minimizing the BIC

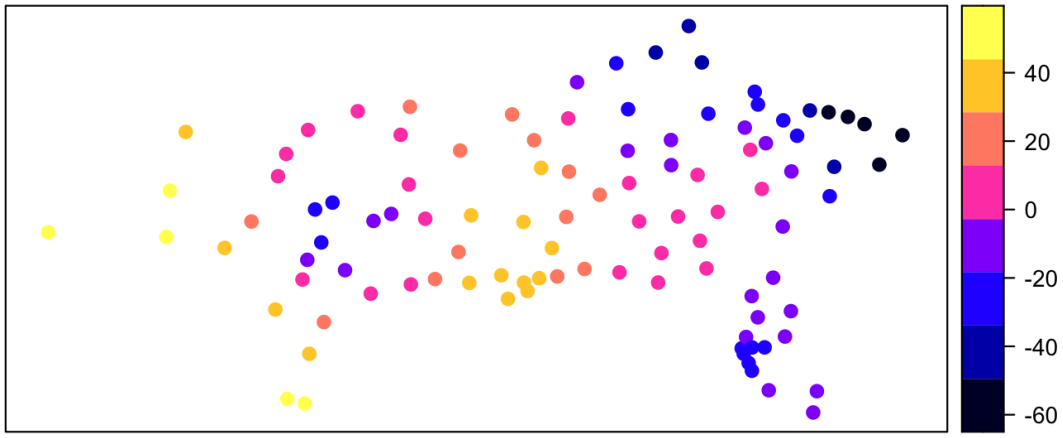

Figure 5.7: Linear ESF Model Spatially Dependent Component Under a Distance-Based $\mathbf{C}$ at the HRR Level with the Objective Function: Minimizing the BIC. The estimated spatially dependent components in the residuals tend to be lowest in North Carolina and Kentucky. The estimated spatially dependent components in the residuals are highest in South and West Texas and Louisiana.

COVID-19 Cumulative Deaths Per 100K in 2020 - Maximizing Adjusted $\mathbf{R^2}$

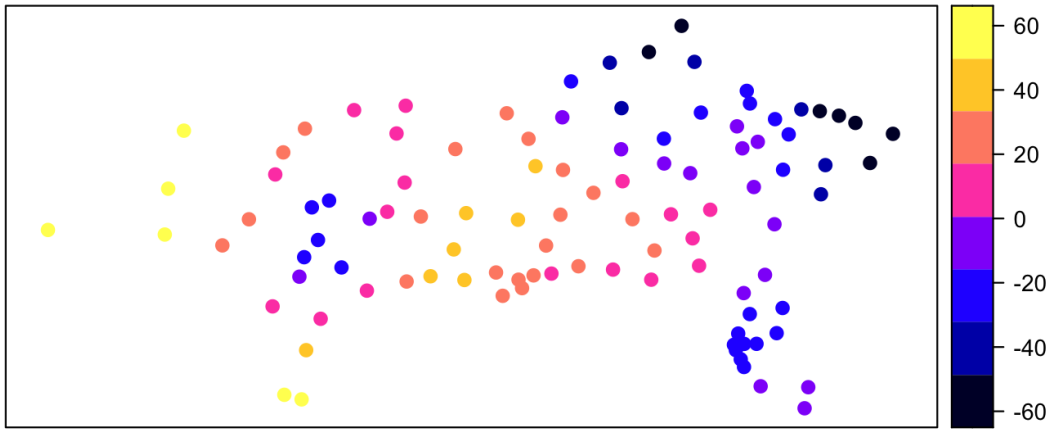

Figure 5.8: Linear ESF Model Spatially Dependent Component Under a Distance-Based $\mathbf{C}$ at the HRR Level with the Objective Function: Maximizing the adjusted R^2. The estimated spatially dependent components in the residuals tend to be lowest in North Carolina and Kentucky. The estimated spatially dependent components in the residuals are highest in South and West Texas and Louisiana.

5.5.2 Hospital Service Areas (HSAs)

COVID-19 Cumulative Deaths Per 100K in 2020 - Minimizing the BIC

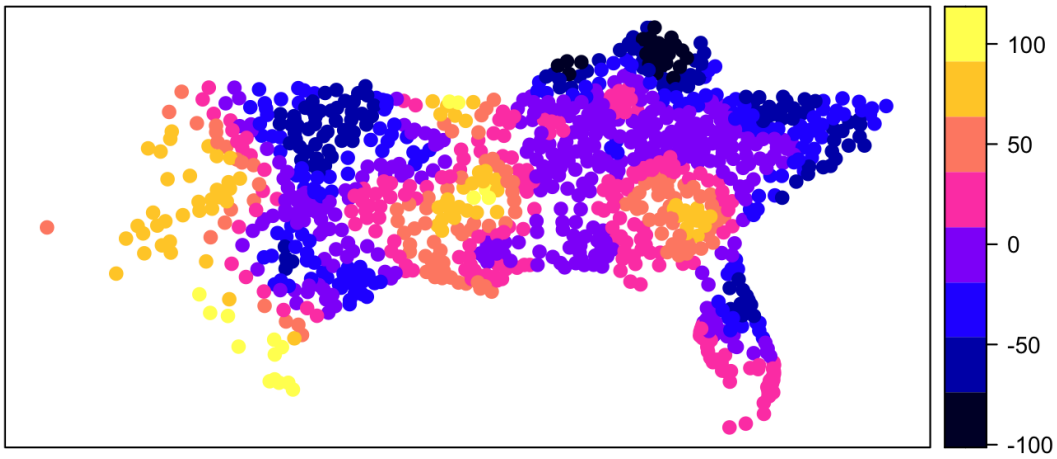

Figure 5.9: Linear ESF Model Spatially Dependent Component Under a Distance-Based $\mathbf{C}$ at the HSA Level with the Objective Function: Minimizing the BIC. The estimated spatially dependent components in the residuals tend to be lowest in Kentucky. The estimated spatially dependent components in the residuals are highest in South Texas and Mississippi.

COVID-19 Cumulative Deaths Per 100K in 2020 - Maximizing Adjusted $\mathbf{R^2}$

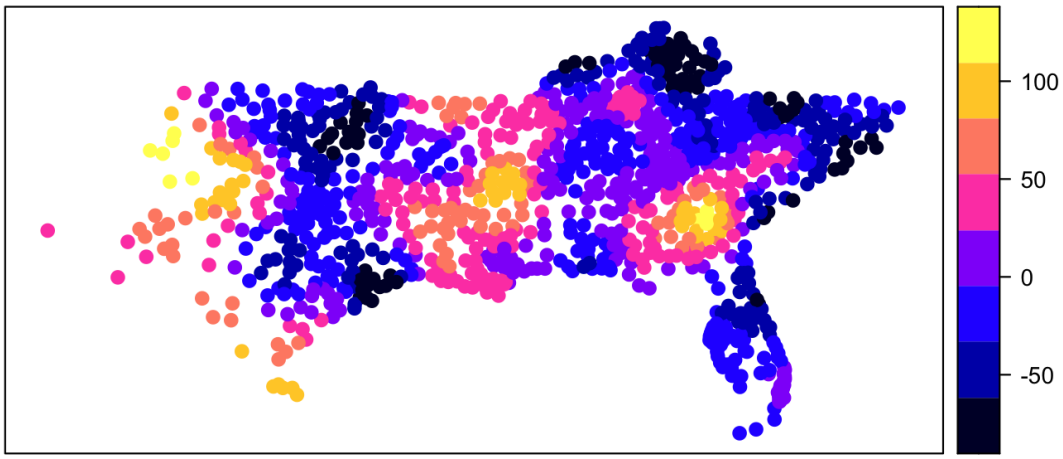

Figure 5.10: Linear ESF Model Spatially Dependent Component Under a Distance-Based $\mathbf{C}$ at the HSA Level with the Objective Function: Maximizing the adjusted R^2. The estimated spatially dependent components in the residuals tend to be lowest in Southeast Texas, Kentucky, and North Carolina. The estimated spatially dependent components in the residuals are highest in West Texas and South Georgia.

COVID-19 Cumulative Deaths Per 100K in 2021 - Minimizing the BIC

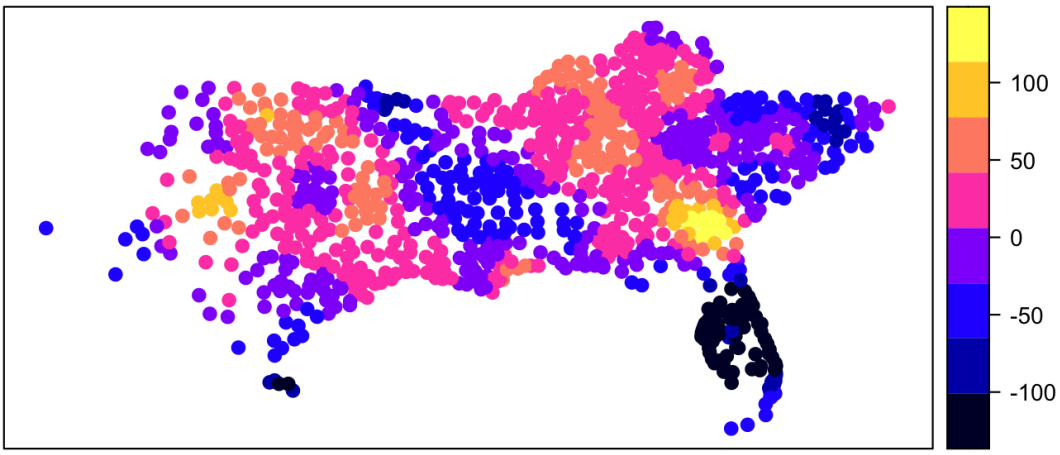

Figure 5.11: Linear ESF Model Spatially Dependent Component Under a Distance-Based $\mathbf{C}$ at the HSA Level with the Objective Function: Minimizing the BIC. The estimated spatially dependent components in the residuals tend to be lowest in Southeast Texas, Kentucky, and North Carolina. The estimated spatially dependent components in the residuals are highest in West Texas and South Georgia.

COVID-19 Cumulative Deaths Per 100K in 2021 - Maximizing Adjusted $\mathbf{R^2}$

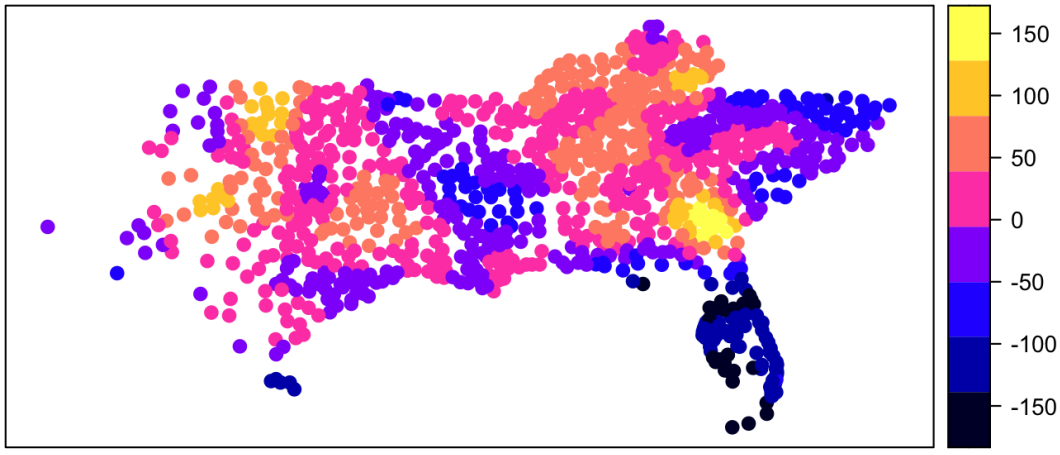

Figure 5.12: Linear ESF Model Spatially Dependent Component Under a Distance-Based $\mathbf{C}$ at the HSA Level with the Objective Function: Maximizing the adjusted R^2. The estimated spatially dependent components in the residuals tend to be lowest in Florida. The estimated spatially dependent components in the residuals are highest in South Georgia and Oklahoma.

5.6 Visualization of Important Eigenvectors for 4NN-based C

The eigenvectors visualized in the **Chapter 5 Results** are selected for being the greatest contributors to maximizing the adjusted R^2/minimzing the BIC under the model selected utilizing the 4NN-based **C** at the HRR and HSA level. Two eigenvectors are visualized for each 4NN-based **C** model. The selected eigenvectors are ranked by the absolute value of their slope coefficient. As a result, the eigenvectors are ranked by the strength of their association with the response variable. As a result, the eigenvectors shown have stronger association with cumulative COVID-19 deaths per capita.

Recall that the set of real numbers that has the largest MC from the geographic arrangement defined in **C** is captured by the first eigenvector, E_1. The second eigenvector, E_2, is the set of real numbers that has the largest achievable MC in which the set is orthogonal and uncorrelated with E_1. This intuition follows so that E_3 through E_n make up the set of real numbers that have the largest negative MC from any set of real numbers that is orthogonal and uncorrelated with the preceding $(n-1)$ eigenvectors. The result is a set of n eigenvectors that is said to provide the full range of SA (D. Griffith et al., 2019). The number the eigenvector is in **C** is in the header of each plot.

Some of the eigenvectors are selected as being the most influential for multiple models. It is important to note that since eigenvectors are not unique in sign, eigenvectors' elements' distributions do not show if the true clusters comprise H or L values. Thus, the red-blue gradient is reversible due to the non-uniqueness in sign for eigenvectors. The eigenvector maps are made up of quantiles with 7 classes.

The **Getis-Ord** $\mathbf{G_i^*}$ values are mapped for each of the eigenvectors selected. The shade itself is based on significance levels of .01, .5, and .1, respectively, based on z values (-2.58, -1.96, -1.645, 1.645, 1.96, and 2.58). The statistical significance is based on deviations from the Moran eigenvector map 0 mean. The color gradient for the H-H and L-L clusters are relative due to arbitrary eigenvector signs - the hot spots and cold spots can be interchanged for the other. The size and distribution of G_i^* clusters are considered the most essential aspect of a Moran eigenvector map. Clusters indicate that the SA map components are expressed well.

The G_i^* statistic allows for one to determine if the polygon map pattern is different from the SA over the entire landscape. Stronger PSA can be associated with bigger clusters across the overall region. Moderate and smaller clusters can be associated with moderate and weak PSA respectively. The G_i^* statistic utilizes three cluster classifications: global, regional, and local to better understand the level of PSA.

Hospital Referral Regions (HRRs)

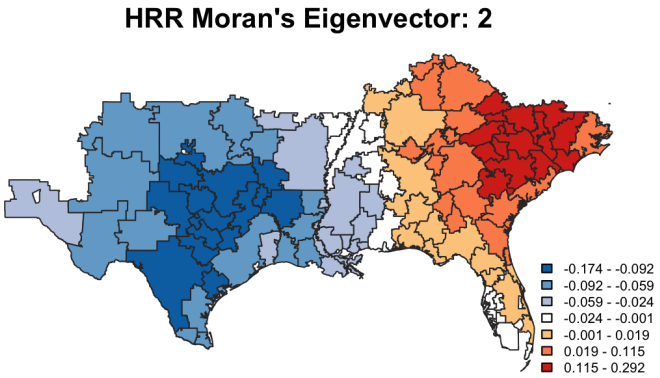

Figure 5.13: The first eigenvector selected at the HRR Level with the Objective Function: Maximizing the Adj. R^2 and the Objective Function: Minimizing the BIC for the **COVID-19 Cumulative Deaths Per 100K in 2020** as the response. The HRR neighbors in Central Texas have eigenvector's elements with similar lower values and higher values are seen in North/South Carolina.

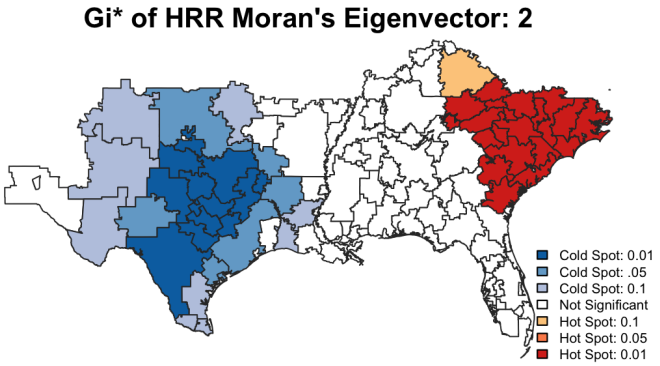

Figure 5.14: The G_i^* shows one large L-L cluster and one large H-H cluster. This portrays global polygon patterns as it has two sizeable clusters of the geographic landscape-wide polygon patterns. The global pattern is associated with stronger PSA. HRRs in Georgia, North Carolina, and South Carolina have similar high G_i^* values, while HRRs in Texas have similar low G_i^* values. The central areas of the study region do not differ significantly from 0.

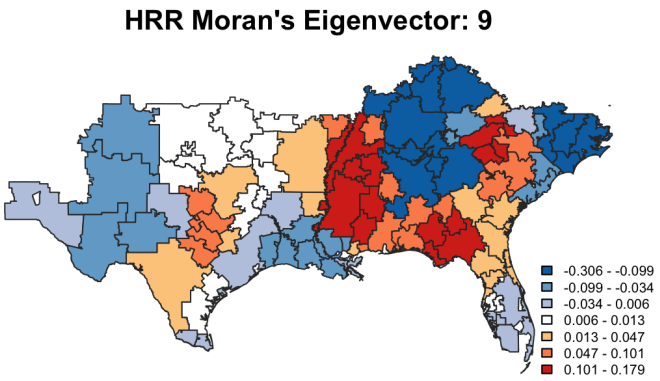

Figure 5.15: The third eigenvector selected at the HRR Level with the Objective Function: Maximizing the Adjusted R^2 and the Objective Function: Minimizing the BIC for the **COVID-19 Cumulative Deaths Per 100K in 2020** as the response. The HRR neighbors in Alabama, Tennessee, and Kentucky have eigenvector's elements with similar lower values and HRR neighbors with similar higher values are in Mississippi.

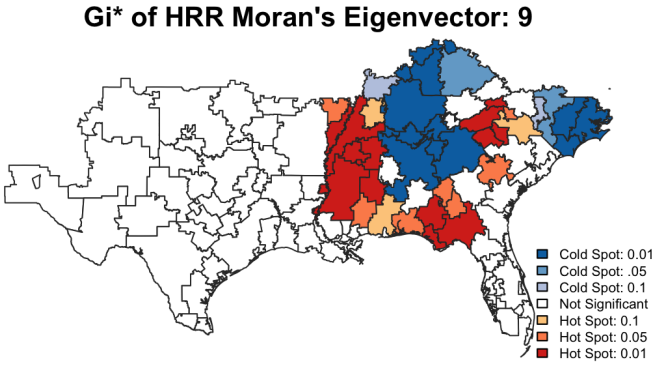

Figure 5.16: The G_i^* shows a few smaller L-L and H-H clusters. This portrays regional polygon patterns. HRRs in Mississippi have similar high G_i^* values, while HRRs in Alabama, Tennessee, Kentucky have similar low G_i^* values.

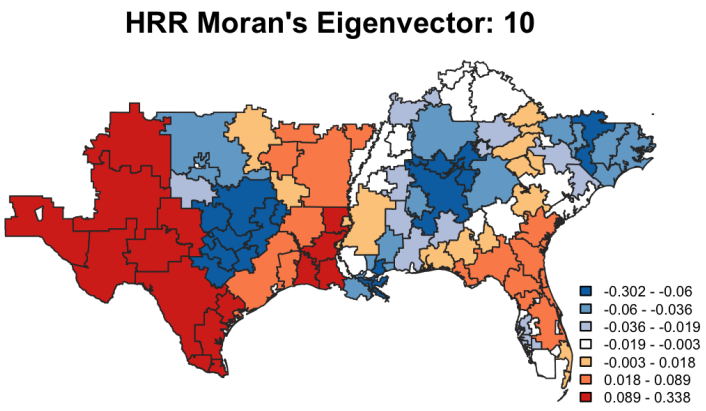

Figure 5.17: The second eigenvector selected at the HRR Level with the Objective Function: Maximizing the Adjusted R^2 and the Objective Function: Minimizing the BIC for the **COVID-19 Cumulative Deaths Per 100K in 2020** as the response. The HRR neighbors in Alabama and Central Texas have eigenvector's elements with similar lower values and HRR neighbors with similar higher values are in West Texas.

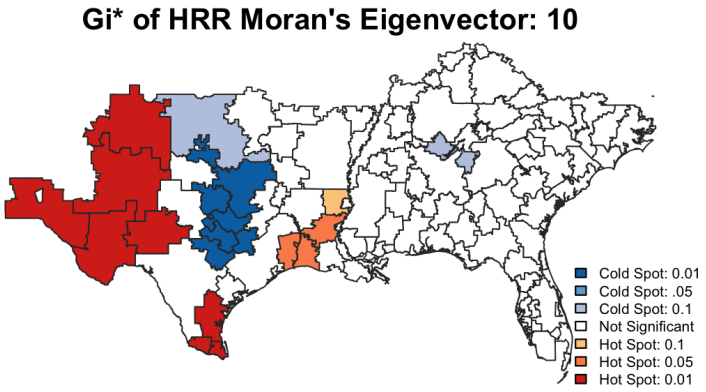

Figure 5.18: The G_i^* shows a few smaller L-L and H-H clusters. This portrays local polygon patterns. HRRs in West Texas have similar high G_i^* values, while HRRs in Central Texas have similar low G_i^* values.

Hospital Service Areas (HSAs)

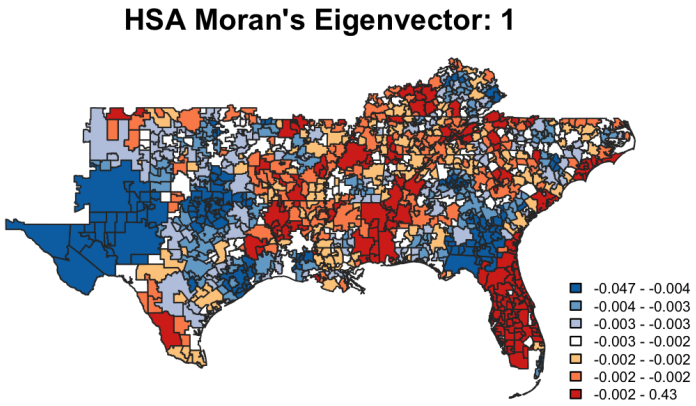

Figure 5.19: The second eigenvector selected at the HSA Level with the Objective Function: Minimizing the BIC for the **COVID-19 Cumulative Deaths Per 100K in 2021** as the response. The eigenvector's elements have similar lower values with HSA neighbors in West Texas and similar higher values in Florida.

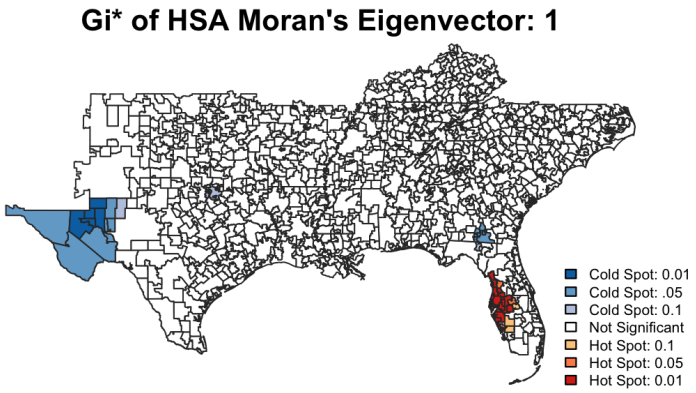

Figure 5.20: The G_i^* shows one small L-L cluster and one small H-H cluster representing a global polygon pattern. HSAs in West Florida have similar high G_i^* values, while HSAs in West Texas have similar low G_i^* values.

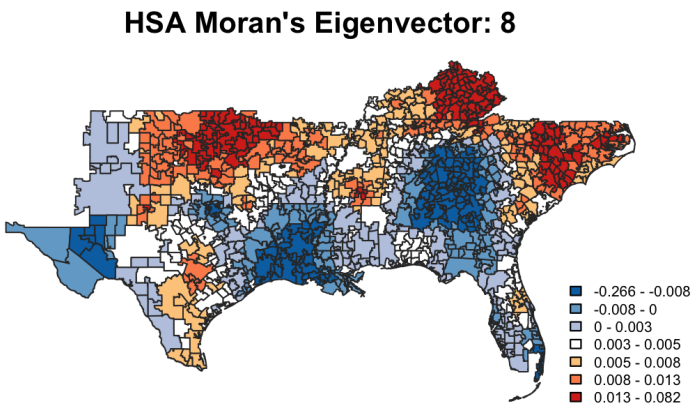

Figure 5.21: The second eigenvector selected at the HSA Level with the Objective Function: Minimizing the BIC and Maximizing the adjusted R^2 for the **COVID-19 Cumulative Deaths Per 100K in 2020** as the response. The eigenvector's elements have similar lower values with HSA neighbors in Georgia and Louisiana and similar higher values in Oklahoma, Kentucky, and South Carolina.

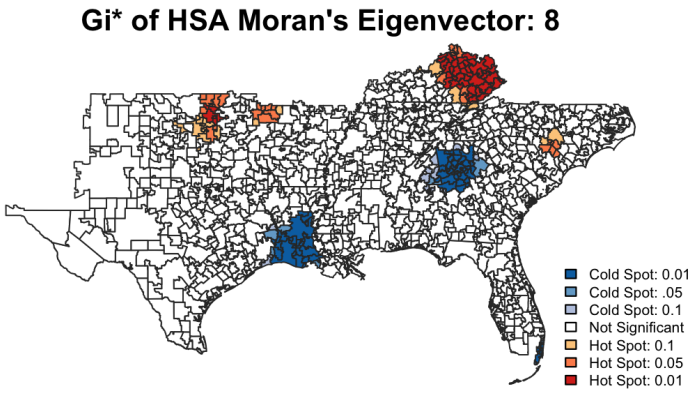

Figure 5.22: The G_i^* shows two small L-L clusters and four small H-H clusters appearing to represent local polygon patterns. HSAs in East Kentucky have similar high G_i^* values, while HSAs in North/Central Georgia and Louisiana have similar low G_i^* values.

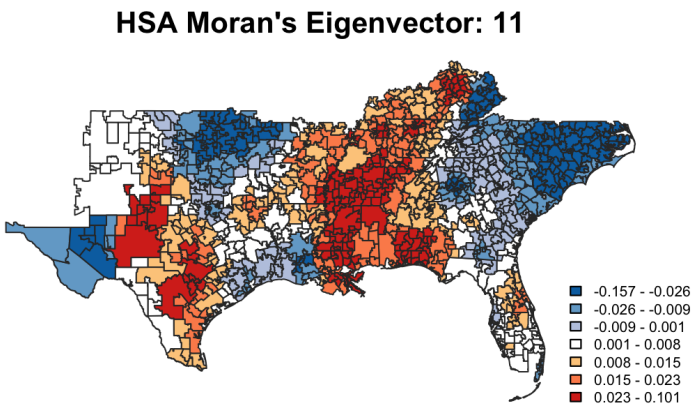

Figure 5.23: The first eigenvector selected at the HSA Level with the Objective Function: Minimizing the BIC and Maximizing the adjusted R^2 for the **COVID-19 Cumulative Deaths Per 100K in 2020** as the response. The eigenvector's elements have similar lower values with HSA neighbors in Oklahoma and North/South Carolina and similar higher values in Alabama and Mississippi.

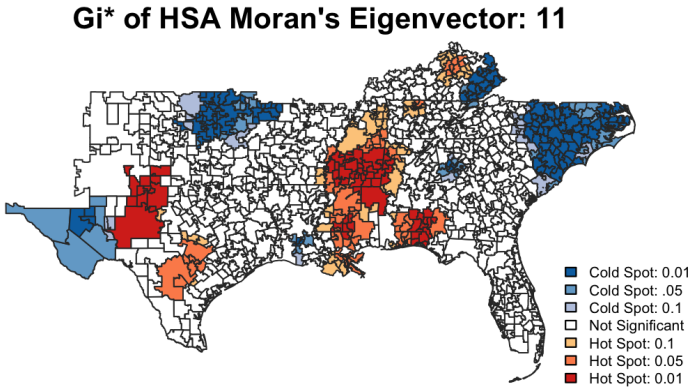

Figure 5.24: The G_i^* shows multiple L-L clusters and H-H clusters, portraying local patterns. HSAs in North/South Carolina and Oklahoma have similar low G_i^* values, while HSAs in Mississippi have similar high G_i^* values.

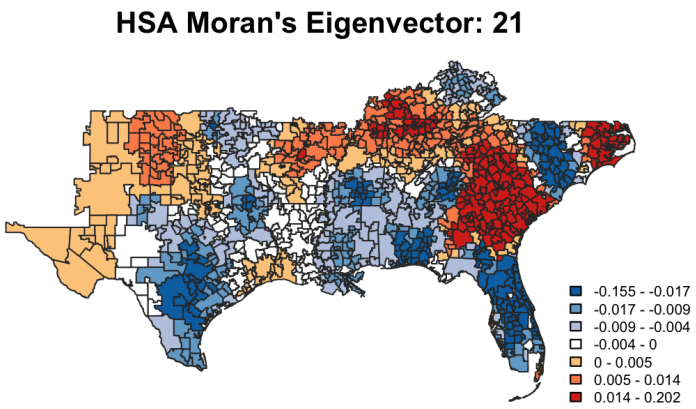

Figure 5.25: The first eigenvector selected at the HSA Level with the Objective Function: Minimizing the BIC and Maximizing the adjusted R^2 for the **COVID-19 Cumulative Deaths Per 100K in 2021** as the response. The eigenvector's elements have similar lower values with HSA neighbors in Florida, South Texas, and South Carolina and similar higher values in Georgia and North Carolina.

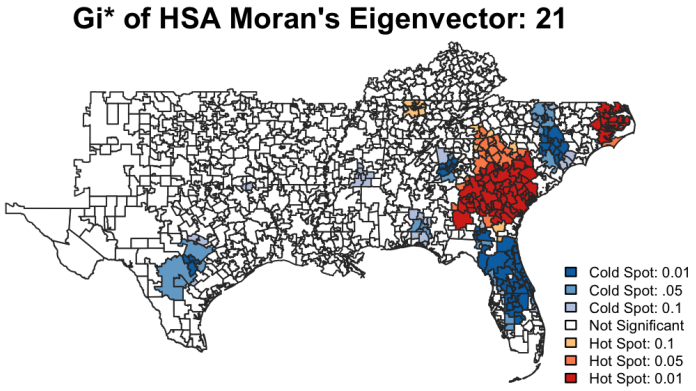

Figure 5.26: The G_i^* shows multiple L-L clusters and H-H clusters, portraying local patterns. HSAs in Florida have similar low G_i^* values, while HSAs in Georgia have similar high G_i^* values.

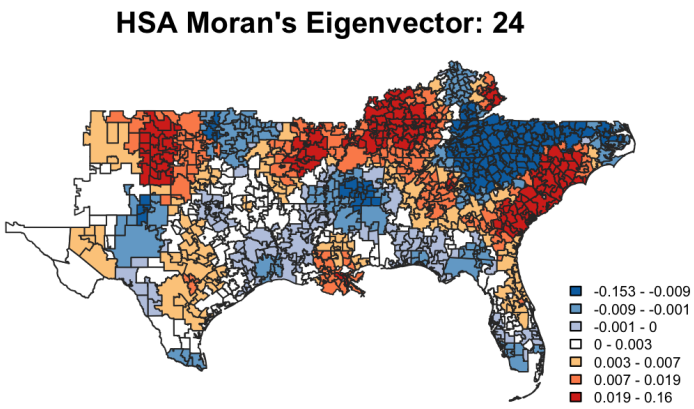

Figure 5.27: The second eigenvector selected at the HSA Level with the Objective Function: Maximizing the adjusted R^2 for the **COVID-19 Cumulative Deaths Per 100K in 2021** as the response. The eigenvector's elements have similar lower values in North Georgia and North Carolina with HSA neighbors and similar higher values on the Georgia and the South Carolina Coast.

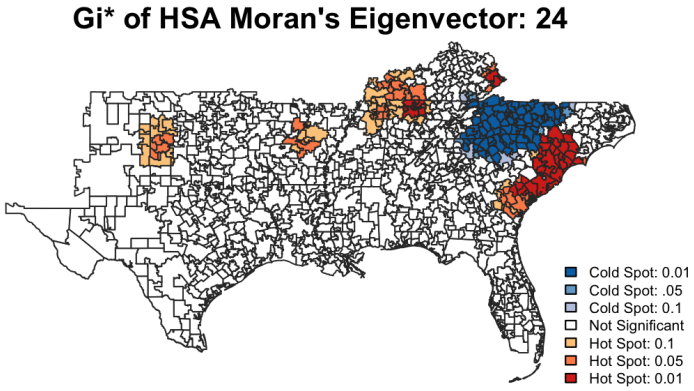

Figure 5.28: The G_i^* shows multiple L-L clusters and H-H clusters, portraying local patterns. HSAs in North/South Carolina have similar low G_i^* values, while HSAs on the Georgia and South Carolina coast have similar high G_i^* values.

5.7 Residual Analysis

5.7.1 Residual vs. Fitted Plots

Hospital Referral Regions (HRRs)

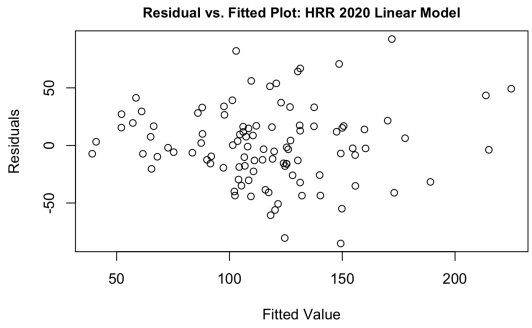

Figure 5.29: Residuals vs. Fitted Plots of the HRR Linear Model. The assumption of linearity holds moderately well. It should be noted that moving right across the x-axis shows some heteroskedasticity as the spread of the residuals seems to be increasing.

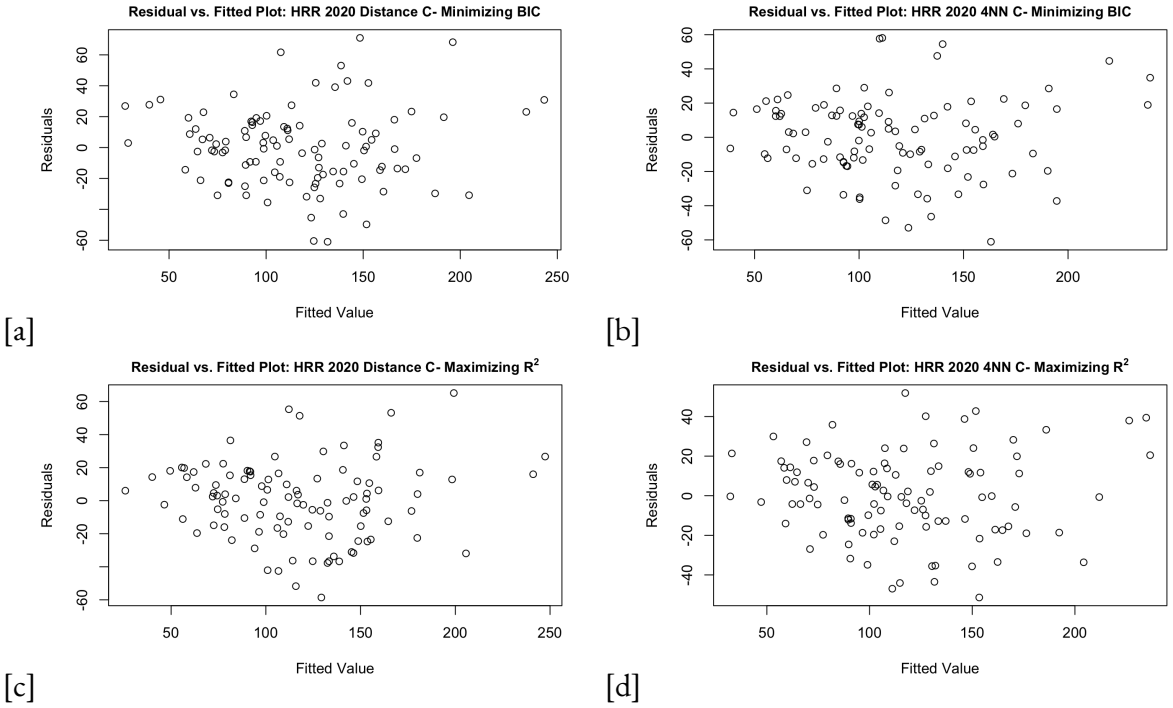

Figure 5.30: Residuals vs. Fitted Plots of the HRR ESF Models. The assumption of linearity holds moderately well for (a) and (b). The assumption of linearity holds well for (d). Heteroskedasticity is noted in (c).

Hospital Service Areas (HSAs)

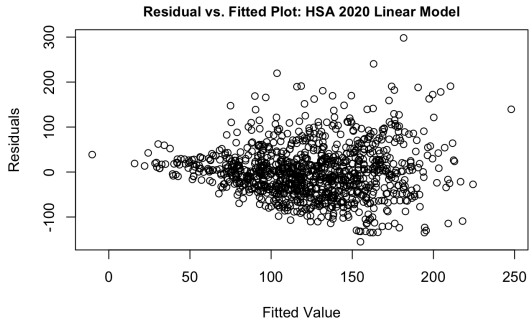

Figure 5.31: Residuals vs. Fitted Plots of the HSA Linear Model for **COVID-19 Cumulative Deaths Per 100K in 2020**. The assumption of linearity does not hold well because moving right across the x-axis shows some heteroskedasticity as the spread of the residuals show an increasing fanning pattern.

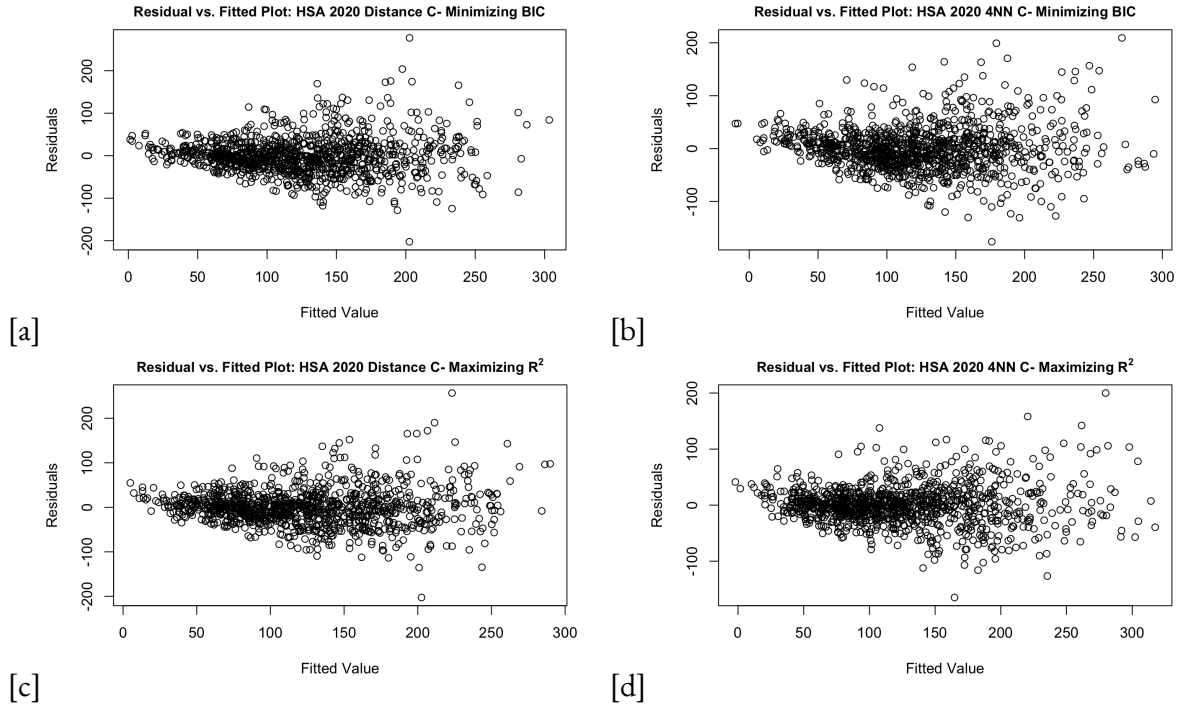

Figure 5.32: Residuals vs. Fitted Plots of the HSA ESF Models for **COVID-19 Cumulative Deaths Per 100K in 2020**. The assumption of linearity holds moderately well for (b), (c), and (d). Heteroskedasticity is more apparent in (a).

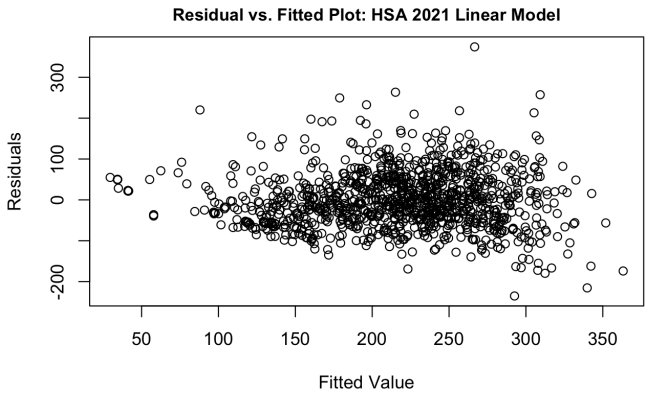

Figure 5.33: Residuals vs. Fitted Plots of the HSA Linear Model for **COVID-19 Cumulative Deaths Per 100K in 2021**. The assumption of linearity holds up moderately well.

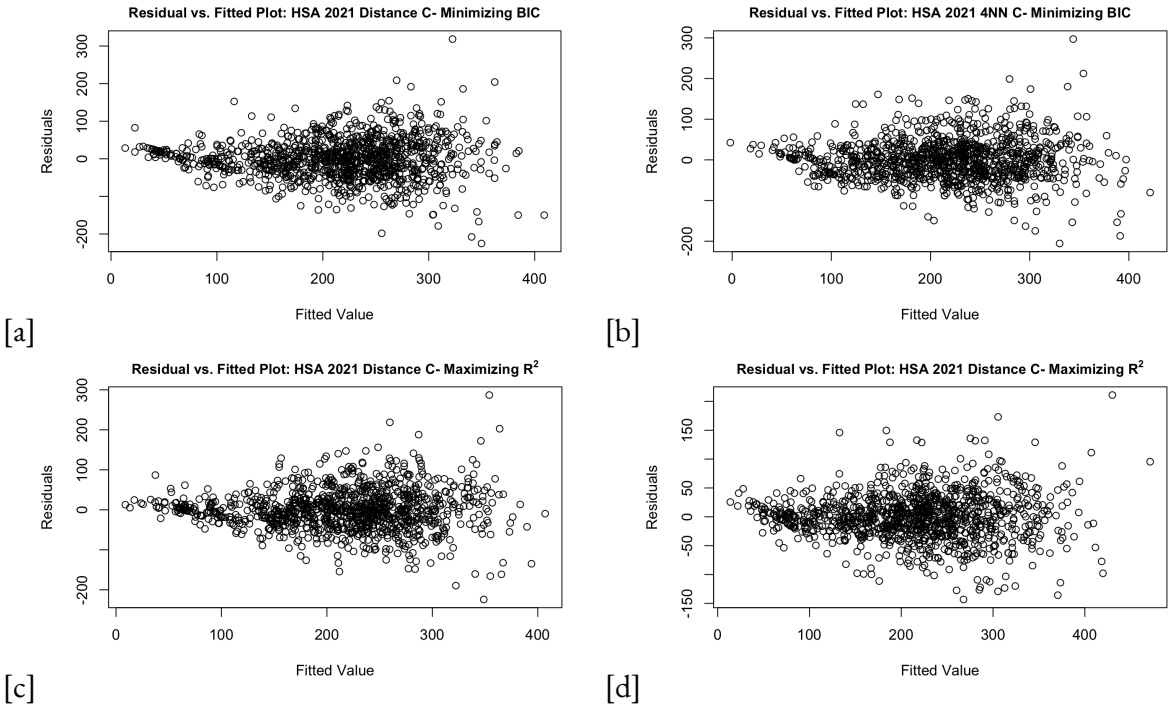

Figure 5.34: Residuals vs. Fitted Plots of the HSA ESF Models for **COVID-19 Cumulative Deaths Per 100K in 2021**. The assumption of linearity does not hold well for (a), (b), and (c) because moving right across the x-axis shows some heteroskedasticity as the spread of the residuals show an increasing fanning pattern. The assumption of linearity holds up moderately well for (d).

5.7.2 Moran Test for Residual Spatial Autocorrelation (SA)

To verify if the SA was filtered from the residuals and the residuals were considered spatially random, the residuals' global Moran's I was calculated for the 4NN-based $\mathbf{C}$. These values were compared to the global Moran's I of the linear models residuals. The global Moran's I was not calculated for the distance-based $\mathbf{C}$ as those models had Moran's I in the weak range. The Moran's I residual calculations are based on the procedure proposed by Cliff and Ord (1981). The null hypothesis: there is no spatial autocorrelation. The alternative hypothesis tested in the following tables: there is spatial autocorrelation greater than 0.

The following table headers interpretation:

- I: observed value of the Moran coefficient
- EI: expected value of Moran's I
- $VarI$: variance of Moran's I
- zI: standardized Moran coefficient
- pI: $p - value$ of the test statistic

Hospital Referral Regions (HRRs)

Table 5.1: The Moran's I of Model Residuals for the Hospital Referral Regions (HRRs) **COVID-19 Cumulative Deaths Per 100K in 2020 Models**

Model	I	EI	$VarI$	zI	pI	
2020 Linear Model	0.223	−0.034	0.004	4.101	0.000	***
2020 Minimizing BIC	−0.100	−0.040	2.080	−0.042	0.517	
2020 Maximizing Adj. R^2	−0.193	-0.040	2.080	−0.106	0.542	

The linear model has significant SA within its residuals as the $p - value$ of the test statistic is ≈ 0. The null hypothesis is rejected for the linear model. However, the MESF models successfully filtered the SA from the residuals, confirmed by the insignificant Moran's I.

Hospital Service Areas (HSAs)

Table 5.2: The Moran's I of Model Residuals for the Hospital Service Areas (HSAs) for **COVID-19 Cumulative Deaths Per 100K in 2020/2021** Models.

Model	I	EI	$VarI$	zI	pI	
2020 Linear Model	0.449	-0.006	0.001	19.271	0.000	***
2021 Linear Model	0.447	-0.007	0.002	9.551	0.000	***
2020 Minimizing BIC	0.068	-0.005	0.002	1.760	0.039	*
2020 Maximizing Adj. R^2	-0.196	-0.004	0.002	-4.732	1.000	
2021 Minimizing BIC	0.118	-0.007	0.002	2.641	0.004	**
2021 Maximizing Adj. R^2	-0.156	-0.005	0.001	-6.357	1.000	

Both linear models and both ESF models with the objective criterion: minimizing the BIC have significant SA within its residuals. The linear model has significant SA within its residuals as the $p-value$ of the test statistic is ≈ 0. The null hypothesis is rejected for the linear model and minimizing the BIC model at $\alpha = .05$. However, the ESF models with the objective criterion: maximizing the adjusted R^2 successfully filtered the SA from the residuals, confirmed by the insignificant Moran's I.

Chapter 6

Conclusion

6.1 Moran's Eigenvector Spatial Filtering (MESF) Modeling

6.1.1 Defining $\mathbf{C}$

The 4-Nearest Neighbor (4NN)-based $\mathbf{C}$ leads to a higher number of eigenvectors extracted from the centered spatial weights matrix compared to the distance-based $\mathbf{C}$ for the Hospital Referral Regions (HRRs) and the Hospital Service Areas (HSAs).

6.1.2 Objective Criteria

The two objective criteria used during the Moran's eigenvector spatial filtering (MESF): minimizing the BIC and maximizing the adjusted R^2 highlighted differences between the distance and 4NN-based $\mathbf{C}$.

The objective criteria of minimizing the BIC led to the selection of the distance-based $\mathbf{C}$ model at the HRR and HSA levels. While the distance-based $\mathbf{C}$ model leads to the greatest reduction in BIC, the value did not change drastically in comparison to the BIC of the linear model. For example, the greatest reduction observed in BIC is a decrease of 3.5% occurring at the HSA level for the **COVID-19 Cumulative Deaths Per 100K in 2020**.

The objective criteria of maximizing the adjusted R^2 led to the selection of the 4NN-based $\mathbf{C}$ model at the HRR and HSA levels. The greatest maximization in adjusted R^2 is an increase by 135.5% occurring at the HSA level for the **COVID-19 Cumulative Deaths Per 100K in 2020**. The increase in the adjusted R^2 after completing MESF is much more noticeable when comparing the adjusted R^2, rather than the minimizing BIC criteria, of the linear model not accounting for spatial autocorrelation (SA).

6.1.3 Model Predictors

Important Vulnerability and Resilience Predictors

SoVI, **BRIC infrastructure**, and **BRIC community** are found to be significant predictors of **COVID-19 Cumulative Deaths Per 100K** in all of the models fit. In addition, SoVI, BRIC infrastructure, and BRIC community typically had the greatest magnitude of the β estimate compared to other predictors.

A lower **SoVI** score is indicative of less vulnerability while a higher SoVI score is indicative of more vulnerability. SoVI had positive β estimates in all models, so higher SoVI scores, in general, lead to increases in the number of **COVID-19 Cumulative Deaths Per 100K**.

BRIC infrastructure is one of the six BRIC categories; a higher BRIC score indicates higher community resiliency. BRIC infrastructure had negative β estimates in all models, so higher BRIC infrastructure scores, in general, correspond to reductions in **COVID-19 Cumulative Deaths Per 100K**. The BRIC infrastructure predictor included temporary housing availability, temporary shelter availability, and evacuation routes.

BRIC community is another one of the six BRIC categories. BRIC Community had positive β estimates in all models, so a higher BRIC Community score, in general, correlates to increases in **COVID-19 Cumulative Deaths Per 100K**. However, it would be expected that an indicator of higher community capacity resilience would correlate with fewer **COVID-19 Cumulative Deaths Per 100K**. The BRIC community predictor included religious affiliation, attachment to place, and political engagement.

BRIC environment was selected in $1/2$ of the HRR and $6/8$ of the HSA models considering SA. BRIC environment had a negative coefficient in all but one of the models it was selected in. So, higher BRIC environment scores, in general, correspond to reductions in **COVID-19 Cumulative Deaths Per 100K**. The BRIC environment predictor included flood buffers, energy use, and water stress.

BRIC economic was selected in $0/2$ of the HRR and $4/8$ of the HSA models considering SA. BRIC economic had a negative coefficient in all but one of the models it was selected in, so higher BRIC economic scores correspond to reductions in **COVID-19 Cumulative Deaths Per 100K**. The BRIC economic predictor included homeownership, employment rate, and income inequality.

BRIC institutional was selected in $0/2$ of the HRR and $3/8$ of the HSA models considering SA. BRIC institutional had a negative coefficient in all of the models it was selected in, so higher BRIC institutional scores correspond to reductions in **COVID-19 Cumulative Deaths Per 100K**. The BRIC institutional predictor included population stability, nuclear accident planning, and crop insurance coverage.

Hurricane Impacts Predictors

Wind 2019 was selected in all HSA models. In general, the hurricane impacts: **Wind 2019**, **Wind 2020**, **Rain 2019**, and **Rain 2020**, if selected, had low magnitudes of positive and negative β coefficients. Throughout the modeling of **COVID-19 Cumulative Deaths Per 100K**, the impact of hurricanes are not found to be as influential when compared to measures of vulnerability and resilience.

6.1.4 Model Performance

In comparison to the linear models not accounting for SA, models under the distance and 4NN-based $\mathbf{C}$ always led to a reduction of residual standard error (SE), AIC, and BIC. In addition, models under the distance and 4NN-based $\mathbf{C}$ resulted in an increase in the adjusted R^2. When comparing model performance based on the objective function selected, models that maximized adjusted R^2 typically performed better than models that minimized the BIC - based on residual SE and AIC measures.

Another indicator of an effective model are collections of very small $p - values$. In general, eigenvectors selected under the 4NN-based $\mathbf{C}$ had a larger collection of smaller $p - values$ than under the distance-based $\mathbf{C}$. Eigenvectors shown in the **Chapter 5 Results** have smaller $p - values$ for the HSA rather than the HRR, but this is likely because only the most important 10 eigenvectors are shown at the HSA level, while all of the eigenvectors selected for the HRR are shown. The eigenvectors selected have low $p - values$ which supports their importance in accounting for SA when modeling cumulative COVID-19 deaths.

6.1.5 Residuals

Most residual plots exhibited heteroskedasticity, with variability increasing as the fitted values increase. To confirm if the SA was filtered from the residuals and the residuals were considered spatially random, the residuals' global Moran's I was calculated for the 4NN-based $\mathbf{C}$. The Moran's I values under the distance-based $\mathbf{C}$ models were always in the weak range: 0.25-0.50. However, Moran's I values were always in the strong range: 0.70-0.90 for all 4NN-based $\mathbf{C}$ models.

The alternative hypothesis tested: there is SA greater than 0. At the HRR level, the linear model had significant SA within its residuals as the $p-value$ of the test statistic was ≈ 0. However, the MESF models at the HRR level successfully filtered the SA from the residuals. At the HSA levels, both linear models and both ESF models with the objective criterion: minimizing the BIC had significant SA within its residuals. However, the ESF models with the objective criterion: maximizing the adjusted R^2 successfully filtered the SA from the residuals for both 2020 and 2021 data.

6.2 Selected Model

The linear ESF model at the HRR level modeling **COVID-19 Cumulative Deaths Per 100K in 2020** as the response under the 4NN-based **C** with the objective function: maximizing the adjusted R^2 had the greatest adjusted R^2 value and lowest residual SE of all models.

Model COVID-19 Cumulative Deaths Per 100K in 2020:
$\widehat{COVID-19} = 105.305 + 12.886 * SoVI - 8.852 * BRIC.Infrastructure + 8.366 * BRIC.Community - 6.551 * BRIC.Environment - 3.36 * Wind.2019 + 1.535 * Wind.2020 - 0.078 * Rain.2019 - 0.166 * Rain.2020 + E_k\hat{\gamma}_k$

At the HRR level, the adjusted R^2 is 49.8% higher for the 4NN-based **C** in comparison to the adjusted R^2 for the linear model not accounting for SA. In addition, the residual SE observed a decrease by 29.3% when comparing the 4NN-based **C** to the linear model. This model also successfully filtered the SA from the residuals.

6.3 Discussion

6.3.1 Hospital Referral Regions (HRRs) and Hospital Service Areas (HSAs)

Based on the fitted models residual SE, AIC, BIC, and adjusted R^2 values, aggregation at the HRR level is found to be more effective than the HSA level. Recall that HRRs contain groups of zip codes determined by the referral patterns of tertiary medical care and HSAs represent patterns of local healthcare utilization of hospital services where the greatest proportion of their Medicare residents were hospitalized (Sledge, 2001). Tertiary medical care is categorized as being highly specialized over an extended period of time (Flegel, 2015). It is theorized that deaths attributed to COVID-19 are more likely to occur in specialized medical facilities that treat patients over a longer period of time. The HSA defining regions are likely too small as most of the HSAs contain only one hospital. Additionally, the MESF models at the HRR level successfully filtered the SA from the residuals for both objective functions, but this did not occur for all 4NN-based $\mathbf{C}$ at the HSA level. As a result, analyzing cumulative COVID-19 deaths with this methodology is more effective at the HRR level than the HSA level.

6.3.2 Defining $\mathbf{C}$

In this analysis, two different definitions of $\mathbf{C}$ are considered: distance-based and 4-nearest-neighbor-based. Both definitions were selected to better understand the influence of the spatial weights matrix (SWM) and its impact on the overall model fit. The 4NN-based $\mathbf{C}$ known as topoligy-based ESF and is commonly used in regional science, while MESF under a distance-based $\mathbf{C}$ framework is standard in ecology (Murakami, 2023). In this analysis, the 4NN-based $\mathbf{C}$ and distance-based $\mathbf{C}$ had similarities in their predictor selection. However, the 4NN-based $\mathbf{C}$ consistently led to better model performance measures. The regional science approach with the the 4NN-based $\mathbf{C}$ was most appropriate for understanding HRR and HSA regional relationships with cumulative COVID-19 deaths.

6.3.3 Model Predictors

Throughout the modeling process at the HRR and HSA level, **SoVI**, **BRIC infrastructure**, and **BRIC community** were selected as important predictors for modeling cumulative COVID-19 deaths. Increased vulnerability according to SoVI and increased infrastructure vulnerability led to an increase in cumulative COVID-19 deaths. Lower community capital vulnerability was found to be an important predictor for increases in cumulative COVID-19 deaths. Community capital resilience is made up of place attachment-not recent immigrants and native born residents, political engagement, social capital-religious organizations, social capital-disaster volunteerism, and citizen disaster preparedness and response skills. Greater community capital resilience may suggest more human interaction and aggregation which could be a factor in spreading infection and not obeying pandemic protocols.

BRIC environment and **BRIC institutional** were selected in 3 and 4 out of 10 models fit utilizing ESF, respectively. As a result, home ownership, employment rate, income inequality, population stability,

nuclear accident planning, and crop insurance coverage are not believed to have a significant impact on cumulative COVID-19 death count. .

Throughout the modeling of cumulative COVID-19 deaths, the impact of hurricanes were not found to be as influential when compared to measures of vulnerability and resilience. It is still speculated that hurricane impacts lead to greater infection deaths due to an increase in aggregation and damage to health infrastructure. The maximum hurricane-associated sustained wind velocity in 2019 was selected as a significant predictor in all HSA models, but the coefficient was positive in some models and negative in others with a low magnitude. The maximum hurricane-associated sustained wind velocity and maximum daily precipitation likely do not fully capture the impact of hurricanes.

6.3.4 ESF Visualizations

Distance-Based C

From the spatially dependent component visualizations of the distance-based $\mathbf{C}$, analysis at the HRR levels observe similar (map) patters for both objective functions. HRRs in South and West Texas and Louisiana have greater cumulative COVID-19 deaths than would have been estimated by a typical linear model that did not account for SA. However, HRRs in North Carolina and Kentucky have fewer cumulative COVID-19 deaths than would have been estimated by a typical linear model that did not account for SA. At the HSA level, the patterns are different based on the year of cumulative COVID-19 deaths being analyzed and the objective function applied. In general, HSAs in South and West Texas and South Georgia have greater cumulative COVID-19 deaths than would have been estimated by a typical linear model that did not account for SA. In general, HSAs in Kentucky, North Carolina, and Florida have fewer cumulative COVID-19 deaths than would have been estimated by a typical linear model that did not account for SA.

At the HRR and HSA level, North Carolina and Kentucky have fewer cumulative COVID-19 deaths and South and West Texas have greater cumulative COVID-19 deaths than would have been estimated by a typical linear model that did not account for SA.

4NN-Based C

The Getis-Ord $\mathbf{G_i^*}$ statistic visualizations are important as they allow for better interpretation of the Moran's eigenvector maps. The statistical significance is based on deviations from the Moran eigenvector map 0 mean. The plots include the eigenvectors with the strongest association with the cumulative COVID-19 deaths per capita models.

At the HRR level, the eigenvectors selected exhibit global, regional, and local patterns. The HRR eigenvector with the strongest association with cumulative COVID-19 deaths showed that the central study region did not differ significantly from 0. The eastern and western most areas of the region of interest significantly deviated from the 0 mean and from one another.

At the HSA level, the eigenvectors selected generally exhibited local polygon patterns which is indicative of weak PSA. The shaded areas in the $\mathbf{G_i^*}$ plots are significantly different from 0 and should be further investigated to determine if they are more or less susceptible to COVID-19 deaths because the red-blue

gradient is reversible due to the non-uniqueness in sign for eigenvectors. The eigenvectors selected as being the most influential in the HSA models had weak PSA. This also aligns with the claim that HSA defining regions are likely too small as most of the HSAs contain only one hospital.

6.3.5 Future Work

Future work in this area would include considerations of flooding events to capture how hurricanes might affect the spread of infection - another primary interest of the interdisciplinary group. It is also of interest to analyze **COVID-19 Cumulative Deaths Per 100K in 2021** at the HRR level. This was not available during the duration of this analysis. Utilizing cross-validation to further confirm results and the impact of the MESF models would be useful. It would also be worth exploring the impact of interactions and higher-order terms in the linear model construction. Heterosckedasticity was apparent in residual plots. It may be of future interest to consider a response transformation.

During the analysis, newer BRIC data was provided by the Hazards and Vulnerability Research Institute (HVRI) at the University of South Carolina. It would be more beneficial to utilize newer data to better create predictive models for COVID-19 cumulative deaths.

Following the data analysis and thesis completion, the author was made aware by the outside source data supplier that data provided for the thesis had been mistakenly, incorrectly aggregated. It is of interest to re-analyze all data using the methods detailed in this thesis to draw proper conclusions about relationships among hurricane impacts, social vulnerability, and community resilience with COVID-19 deaths per capita.

Appendix A

A.1 Moran Coefficient

The MC is defined as:

$$\frac{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(y_i-\bar{y})}{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}}}{\sum_{i=1}^{n}(y_i-\bar{y})^2/n} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}} \frac{(y_i-\bar{y})\left[\sum_{j=1}^{n} c_{ij}(y_j-\bar{y})\right]}{(n-1)s^2}$$

$$= \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}} \frac{\sum_{i=1}^{n} z_i(\sum_{j=1}^{n} c_{ij}z_j)}{(n-1)}$$

where c_{ij} is the ith and jth entry of the SWM $\mathbf{C}$, $\bar{y}$ is the arithmetic mean, s^2 is the sample variance of Y, and z_i is the z-score for y_i. The RHS of the equation above shows the similarity between the MC and bivariate regression.

Using these equations: $Det(\mathbf{C}-\lambda\mathbf{I})=0$ and $(\mathbf{C}-\lambda\mathbf{I})E_i=0, i=1,2,...,n$, we can receive the following set of MC values:

$$MC_j = \frac{n}{\mathbf{1}^T\mathbf{C}\mathbf{1}} \frac{\mathbf{E}_j^T(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)\mathbf{E}_j}{\mathbf{E}_j^T(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)\mathbf{E}_j} \quad = \frac{n}{\mathbf{1}^T\mathbf{C}\mathbf{1}} = \frac{\mathbf{E}_j^T\mathbf{C}\mathbf{E}_j}{\mathbf{E}_j^T\mathbf{E}_j} \quad = \lambda_j \frac{n}{\mathbf{1}^T\mathbf{C}\mathbf{1}}$$

The MC in matrix form:

$$\frac{n}{\mathbf{1}^T\mathbf{C}\mathbf{1}} \frac{\mathbf{Y}^T(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)\mathbf{Y}}{\mathbf{Y}^T(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)\mathbf{Y}}$$

$(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)$ is the part of the matrix relating to the SA.

The eigenfunctions are calculated by the following:

$$Det\left[(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)-\lambda\mathbf{I}\right]=0.$$
$$\left[(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}^T/n)-\lambda_i\mathbf{I}\right]E_i=0, i=1,2,...,n$$

Det is the determinant and λ is the eigenvalue. Solving the polynomial of nth order results in the set of n eigenvalues denoted by $\lambda_i, i=1,2,...,n$. When a matrix is real and symmetric the eigenvectors are mutually orthogonal. With each eigenvalue, an $n\times 1$ vector occurs by solving the following equation: $(\mathbf{C}-\lambda\mathbf{I})E_i=0, i=1,2,...,n$. E_i is the ith eigenvector where $E_i\neq 0$. The eigenvectors E_i also meet

the following criteria: $E_i^T E_i = 1$ and $E_i^T E_j = 0, i \neq j$. The matrix $(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)$ is square and symmetric. As a result, the matrix can be decomposed into matrices that are rank 1 with the products of the eigenvalues and eigenvectors such that:$(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^\mathbf{T}$, $\mathbf{\Lambda}$ is the diagonal matrix that holds the n eigenvalues λ_j. λ_1 results in the maximum value of MC because it optimizes its Rayleigh quotient, while λ_n results in the minimum value of MC.

A.2 Getis-Ord $\mathbf{G_i^*}$ Statistic

The Getis-Ord G_i^* statistic is defined as:

$$\mathbf{G_i^*} = \frac{\sum_{j=1}^{n} c_{ij} y_j - \bar{y} \sum_{j=1}^{n} c_{ij}}{\sqrt{\frac{\sum_{j=1}^{n} y_j^2}{n} - (\bar{y})^2} \sqrt{\frac{n \sum_{j=1}^{n} c_{ij}^2 - (\sum_{j=1}^{n} c_{ij})^2}{n-1}}}$$

where c_{ij} is the ith and jth entry of the SWM $\mathbf{C}$, $\bar{y}$ is the arithmetic mean.

Appendix B

Table B.1: **BRIC Community**

Variable	Description
Place attachment-not recent immigrants	% Population not foreign-born persons who came to US within previous 5 years
Place attachment-native born residents	% Population born in state of current residence
Political engagement	% Voting age population participating in recent election
Social capital-religious organizations	# Affiliated with a religious organization per 10,000 persons
Social capital-disaster volunteerism	# Red Cross volunteers per 10,000 persons
Citizen disaster preparedness and response skills	# Red Cross training workshop participants per 10,000 persons

Table B.2: **BRIC Economic**

Variable	Description
Home ownership	% Owner-occupied housing units
Employment rate	% Labor force employed
Race/ethnicity income equality	Gini coefficient (Inverted; lower coefficient is more resilient)
Non-dependence on primary/tourism sectors	% Employees not in farming, fishing, forestry, extractive industry, or tourism
Gender income equality	Absolute difference of male and female median income (Inverted; less difference means more equality, resilience)
Business size	Ratio of large to small businesses
Large retail-regional/national geographic distribution	Large retail stores per 10,000 persons
Federal employment	% Labor force employed by federal government

Table B.3: **BRIC Environment**

Variable	Description
Local food suppliers	Farms marketing products through Community Supported Agriculture per 10,000 persons
Natural flood buffers	% Land in wetlands
Efficient energy use	Megawatt hours per energy consumer (Inverted; less consumption is more efficient and resilient)
Pervious surfaces	Average percent perviousness
Efficient water use	Water Supply Stress Index (Inverted; less stress is more efficient and resilient)

Table B.4: **BRIC Institutional**

Variable	Description
Mitigation spending	Ten year average per capita spending for mitigation projects
Flood insurance coverage	% Housing units covered by National Flood Insurance Program
Performance regimes-state capital	Distance from county seat to state capital (Inverted; closer is more resilient)
Performance regimes-nearest metro area	Distance from county seat to nearest county seat within a Metropolitan Statistical Area (Inverted; closer is more resilient)
Political and jurisdictional fragmentation	# Governments and special districts per 10,000 persons (Inverted; fewer districts, less fragmented is more resilient)
Disaster aid experience	# Presidential Disaster Declarations divided by # of loss-causing hazard events for 10-year period
Local disaster training	% Population in communities covered by Citizen Corps programs
Population stability	Population change over previous 5-year period (Inverted; less change is more resilient)
Nuclear plant accident planning	% Population within 10 miles of nuclear power plant
Crop insurance coverage	# Crop insurance policies per square mile

Table B.5: **BRIC Infrastructural**

Variable	Description
Sturdier housing types	% Housing units not mobile homes
Temporary housing availability	# Vacant rental units per 10,000 persons
Medical care capacity	# Hospital beds per 10,000 persons
Evacuation routes	Major road egress points per 10,000 persons
Housing stock construction quality	% Housing units built prior to 1970 or after 2000
Temporary shelter availability	# Hotels/motels per 10,000 persons
School restoration potential	# Public schools per 10,000 persons
Industrial re-supply potential	Rail miles per square mile
High speed internet infrastructure	% Population with access to broadband internet service

Table B.6: **BRIC Social**

Variable	**Description**
Educational attainment equality	Absolute difference between % population over 25 with college education and % population over 25 with less than high school education
Pre-retirement age	% Population below 65 years of age
Transportation access	% Households with at least one vehicle
Communication capacity	% Households with telephone service available
English language competency	% Population proficient English speakers
Non-special needs	% Population without sensory, physical, or mental disability
Health insurance	% Population under age 65 with health insurance
Mental health support	Psychosocial support facilities per 10,000 persons
Food provisioning capacity	Food insecurity rate (Inverted; lower insecurity is more resilient)
Physician access	Physicians per 10,000 persons

Bibliography

Affairs, A. S. P. (2015). Hhs regional offices.

Anderson, B., Yan, M., Ferreri, J., Crosson, W., Al-Hamdan, M., Schumacher, A., & Eddelbuettel, D. (2020). Hurricaneexposure: Explore and map county-level hurricane exposure in the united states. *R Packag. version 0.1*, *1*.

Anderson, G. B., Schumacher, A., Guikema, S., Ferreri, J. M., & Quiring, S. (2018). Details of stormwindmodel package.

Bi, Y., Xie, J., Sha, Z., Wang, M., Fu, Y., & Chen, W. (2018). Modeling spatiotemporal heterogeneity of customer preferences in engineering design. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, *51753*, V02AT03A050.

Boraks, A., Plunkett, G., Alo, F., Sam, C., Tuiwawa, M., Ticktin, T., & Amend, A. (2021). Scale-dependent influences of distance and vegetation on the composition of aboveground and belowground tropical fungal communities. *Microbial Ecology*, *81*, 1–10. https://doi.org/10.1007/s00248-020-01608-4

Bronner, K. K., & Goodman, D. C. (2018a). Dartmouth atlas project faq.

Bronner, K. K., & Goodman, D. C. (2018b). Dartmouth atlas project research methods.

College, P. E. (2018). Detecting multicollinearity using variance inflation factors.

Cutter, S. L., Burton, C. G., & Emrich, C. T. (2010). Disaster resilience indicators for benchmarking baseline conditions. *Journal of homeland security and emergency management*, *7*(1).

Drake, J. M., Marty, É., Gandhi, K. J., Welch-Devine, M., Bledsoe, B., Shepherd, M., Seymour, L., Fortuin, C. C., & Montes, C. (2023). Disasters collide at the intersection of extreme weather and infectious diseases. *Ecology letters*, *26*(4), 485–489.

Flegel, K. (2015). Tertiary hospitals must provide general care.

Gensini, V. A., & Brooks, H. E. (2018). Spatial trends in united states tornado frequency. *NPJ climate and atmospheric science*, *1*(1), 38.

Getis, A., & Griffith, D. A. (2002). Comparative spatial filtering in regression analysis. *Geographical analysis*, *34*(2), 130–140.

Griffith, D., & Chun, Y. (2014). Spatial autocorrelation and spatial filtering. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of regional science* (pp. 1478–1506). Springer Berlin Heidelberg.

Griffith, D., Chun, Y., & Li, B. (2019). *Spatial regression analysis using eigenvector spatial filtering*. Academic Press.

Griffith, D. A. (2021). Interpreting moran eigenvector maps with the getis-ord gi* statistic. *The Professional Geographer*, *73*(3), 447–463.

Griffith, D. A., & Griffith, D. A. (2003). *Spatial filtering*. Springer.

Hopkins, J. (2023). Covid-19 united states cases by county.

Institute, H. V. R. (2015). Social vulnerability index for the united states.

Landsea, C. W., & Franklin, J. L. (2013). Atlantic hurricane database uncertainty and presentation of a new database format. *Monthly Weather Review*, *141*(10), 3576–3592.

Lazaro, G., Lutz, E., & Sun, A. (2023). As emergency ends, a look at covid's u.s. death toll.

MB&C, M. B. C. (2022). Crosswalking in billing and coding.

Murakami, D. (2017). Spmoran: An r package for moran's eigenvector-based spatial regression analysis (version 2017/09).

Murakami, D. (2023). Spmoran: Fast spatial regression using moran eigenvectors. *0.2.2.9*.

Service, N. W. (2020). New mexico weather hazards.

Skinner, J. S., Gottlieb, D. J., & Carmichael, D. (2022). A new series of medicare expenditure measures by hospital referral region: 2003-2008: A report of the dartmouth atlas project.

Sledge, C. B. (2001). The dartmouth atlas of musculoskeletal health care.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic geography*, *46*(sup1), 234–240.

Zhang, J., Li, B., Chen, Y., Chen, M., Fang, T., & Liu, Y. (2018). Eigenvector spatial filtering regression modeling of ground pm2. 5 concentrations using remotely sensed data. *International Journal of Environmental Research and Public Health*, *15*(6), 1228.