ORIGINAL PAPER

Geographical and temporal weighted regression: examining spatial variations of COVID-19 mortality pattern using mobility and multi-source data

Nanzhou Hu[1], Ziyi Zhang[2], Nicholas Duffield[2], Xiao Li[3], Bahar Dadashova[4], Dayong Wu[5], Siyu Yu[6], Xinyue Ye[6], Daikwon Han[7] and Zhe Zhang[1,2]*

Abstract

The COVID-19 pandemic has had profound adverse effects on public health and society, with increased mobility contributing to the spread of the virus and vulnerable populations, such as those with pre-existing health conditions, at a higher risk of COVID-19 mortality. However, the specific spatial and temporal impacts of health conditions and mobility on COVID-19 mortality have yet to be fully understood. In this study, we utilized the Geographical and Temporal Weighted Regression (GTWR) model to assess the influence of mobility and health-related factors on COVID-19 mortality in the United States. The model examined several significant factors, including demographic and health-related factors, and was compared with the Multiscale Geographically Weighted Regression (MGWR) model to evaluate its performance. Our findings from the GTWR model reveal that human mobility and health conditions have a significant spatial impact on COVID-19 mortality. Additionally, our study identified different patterns in the association between COVID-19 and the explanatory variables, providing insights to policymakers for effective decision-making.

Keywords COVID-19, Mobility, Multi-scale geographically weighted regression, Geographical and temporal weighted regression

*Correspondence:
Zhe Zhang
zhezhang@tamu.edu
[1] Department of Geography, Texas A&M University, College Station, TX, USA
[2] Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA
[3] Transport Studies Unit, Oxford University, Oxford, UK
[4] Traffic Operations and Roadway Safety Division, Texas A&M Transportation Institute, College Station, TX, USA
[5] Research & Implementation, Texas A&M Transportation Institute, Dallas, TX, USA
[6] Department of Landscape Architecture & Urban Planning, Texas A&M University, College Station, TX, USA
[7] School of Public Health, Texas A&M University, College Station, TX, USA

1 Introduction

The ongoing COVID-19 disease has caused tremendous negative impacts on human health and society. Until February 2023, the COVID-19 pandemic has resulted in 1,140,000 deaths and over 100 million cases (JHU CSSE 2023). As a highly contagious virus, its transmission routes are divided into direct, aerosol, and contact transmission. The respiratory droplets the infected person produces when coughing, sneezing, or talking may spread COVID-19 (CDC Spread 2021). Social distancing has been approved as an effective way to prevent the pandemic, a pivotal countermeasure to control the spread of COVID-19 (Hu et al. 2022). Also, the emergence of the COVID-19 vaccine has significantly reduced the mortality rate (Steiger et al. 2021). The effectiveness of the

Hu *et al. Computational Urban Science*   (2024) 4:6

Page 2 of 13

COVID-19 vaccines in preventing severe illness, hospitalizations, and deaths has been a monumental achievement in public health. However, the pandemic continues to pose challenges due to the emergence of new COVID-19 variants. These variants have occasionally led to increases in cases and have tested the resilience of public health measures.

Geographic information system (GIS) represents a technical system that collects, stores, manages, calculates, analyzes, and displays spatial data with the help of computer hardware and software systems to support spatial decision-making (Rogers 1999; Zhang et al. 2021; Zhang et al. 2014; Zhang et al. 2020). Thus, GIS has played a vital role in examining the spatial pattern of infectious diseases (Mollalo et al. 2018; Li et al. 2020; Huang et al. 2022). For example, the emergence of Web-GIS provides various GIS dashboards that can help monitor the live pandemic distribution worldwide (JHU CSSE) (Dong et al. 2020; Berry et al. 2020). Also, GIS modeling can spatially examine the spread patterns of COVID-19 and help scholars to understand the associations between the pandemic and explanatory socioeconomic factors. Fortaleza et al. (2020) applied Cox regression to reinforce the spatial and hierarchical spread, which indicates the importance of travel distance (Fortaleza et al. 2020). Also, socioeconomic factors such as median household income and poverty have been identified as crucial factors affecting social distancing practices and causing the spread of the disease (Ehlert 2021).

Nevertheless, while GIS has provided valuable insights into the spatial aspects of the pandemic, there remains a gap in understanding the influence of individual health factors. Studies have shown that older people with health concerns, such as heart disease and diabetes, are more vulnerable to severe symptoms and mortality due to COVID-19 (Iyanda et al. 2020, CDC COVID 2021). For instance, (Abdi et al. 2020) highlighted diabetes as a significant contributor to COVID-19 severity and mortality, and (Gu et al. 2020) found that the mortality risk for patients with heart disease was three times higher than for those without. These findings underscore the need for a more comprehensive analysis integrating health and mobility factors with spatial and temporal data to fully understand COVID-19's impacts.

Geographically weighted regression (GWR) has been one of the most effective methodologies to examine the non-stationary spatial relationships between the dependent and independent variables (Fotheringham et al. 2002). GWR explores spatial object changes and related driving factors locally by establishing the local regression equation at each point in the spatial range and can predict future results (Fotheringham et al. 2002). Many studies can be found to apply the GWR model to COVID-19

studies (Liu et al. 2020; Jiao et al. 2021; Wu and Zhang 2021; Cui et al. 2022), which already proven the spatial heterogeneity issue in COVID-19 related studies. However, while GWR effectively captures spatial heterogeneity, it does not inherently account for temporal dynamics, which are crucial in understanding the evolution and progression of phenomena like the COVID-19 pandemic. The temporal dimension is particularly vital in infectious disease modeling, where transmission, recovery, and mortality rates can change significantly over time due to factors such as policy interventions, behavioral changes, and natural disease progression. The absence of this temporal component in traditional GWR models can limit their effectiveness in providing a comprehensive understanding of the pandemic's dynamics. As an extension of the GWR model, the Geographically and Temporally Weighted Regression (GTWR) model was introduced to model spatial and temporal variations simultaneously. It can deal with non-stationary spatial data and consider temporal effects (Fu and Li 2020). Additionally, to address the spatial heterogeneity evident in different scales in COVID-19 mortality data, we also employ the Multiscale Geographically Weighted Regression (MGWR) model. MGWR can analyze data across multiple scales and helps identify the differential impacts of various factors in distinct regions, further enriching our understanding of the spread and impact of COVID-19.

This study integrates various socioeconomic factors, mobility variables, and health conditions to understand their impact on COVID-19 mortality. This holistic approach allows for a more detailed and nuanced analysis than previous studies, which often considered these factors in isolation (Iyanda et al. 2020; Liu et al. 2020; Jiao et al. 2021; Wu and Zhang 2021; Cui et al. 2022). It aimed to quantitatively examine the non-stationary spatial and temporal associations between mortality and several socioeconomic factors such as mobility, health conditions, poverty level, and insurance coverage to assist policymakers in making better risk-informed decisions for public health management.

2 Background and theory

Geographic Weighted Regression (GWR) is a statistical tool used to explore the spatial variation in relationships between variables (Cui et al., 2022). They are widely used in various research fields, including geography, environmental sciences, and urban planning. This part aims to overview these three models and their applications in different research fields.

GWR is a spatial regression method that has been widely used to examine the spatial variation of relationships between variables. The method was first proposed by

Brunsdon et al. (1998) and has been extended by various researchers since then:

$$y_i = \alpha_0(i) + \sum_{k=1}^{p} \alpha_k(i)X_{ik} + \varepsilon_i, i = 1, \ldots, n$$

The parameters in the model will vary by location i, $\alpha'(i)$ allows for local variation in the relationship between the variables to be captured. The relationship between dependent and independent variables varies over space in GWR. The basic idea of GWR is to estimate a separate regression model for each location in the study area, using data from neighboring locations to inform the estimation of the model (Brunsdon et al. 1998). The distance between locations is usually measured using a spatial weight matrix, which can be defined based on geographic distance or some other spatial relationship (Deller and Lledo 2007; Leung et al. 2000; Hu et al. 2022; McMillen 1996).

MGWR is a significant extension of the GWR model. While GWR allows for the estimation of spatially varying coefficients at the local level, MGWR allows for estimating these coefficients at multiple scales, capturing the complexity of spatial relationships across different levels of analysis (Fotheringham et al. 2017):

$$y_i = \sum_{j=0}^{m} \beta_{bwj}(u_i, v_i)X_{ij} + \varepsilon_i, i = 1, \ldots, n$$

where y_i is the dependent variable for the ith observation and bwj in β_{bwj} indicates the bandwidth used for calibration of the *jth* conditional relationship. The concept of scale dependency is widely recognized in the spatial analysis literature, referring to the observation that spatial relationships between variables can vary at different scales. It means that the spatial processes and patterns observed at one scale may not hold at another scale. Therefore, analyzing spatial relationships at a single scale may not capture the full complexity of the relationship and may result in biased or incomplete results. To overcome this challenge, MGWR estimates coefficients for multiple spatial scales simultaneously, providing a more comprehensive and nuanced analysis of spatial relationships. The critical difference between GWR and MGWR is that MGWR models relationships between variables at multiple spatial scales, allowing for examining spatial non-stationarity across multiple scales (Fotheringham et al. 2017; Oshan et al. 2020).

GTWR model is another extension of GWR. It models the relationship between a dependent variable and one or more independent variables. GTWR extends the traditional regression models by incorporating spatial and temporal variations into the modeling process (Fotheringham et al. 2015):

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i, t_i)X_{ik} + \varepsilon_i, i = 1, \ldots, n$$

Compared with MGWR, a temporal element added to the formula of GTWR. GTWR allows for considering temporal effects and non-stationary spatial data (Fu and Li 2020). This method accounts for the heterogeneity of the relationships between the variables, which can vary based on geographic or temporal location. GTWR is particularly useful when there is evidence of non-stationarity in the data, which means that the relationships between variables change over space and time. While it may have limitations in terms of computational intensity and the selection of optimal bandwidths, GTWR remains a valuable tool for researchers looking to model their data's spatial or temporal variations.

3 Methods
This study investigated county-level variations of COVID-19 mortality across the United States by incorporating mobility, social, and other health-related data during July, August, and September 2021. The mobility data (home-dwell time) was retrieved from Safegraph (Safegraph 2021). Health-related data on heart disease and diabetes were collected from the Centers for Disease Control and Prevention Atlas (CDC, Atlas 2018). We collected other county-level demographic data from the National Historical Geographic Information System-NHGIS) (NHGIS (2021). The COVID-19 data was downloaded from the New York Times daily report on GitHub (Covid GitHub, 2020). All data were normalized during the model processing. Higher levels of mobility may exacerbate the impact of pre-existing health conditions on COVID-19 mortality. Increased mobility could result in greater exposure to the virus and a higher risk of infection, as suggested by previous studies (Fortaleza et al. 2020)). Additionally, populations with pre-existing health conditions, such as heart disease and diabetes, are known to be more vulnerable to COVID-19-related mortality. We hypothesized that the prevalence of these health conditions might play a critical role in the spatial distribution of COVID-19 mortality, with areas having higher rates of heart disease and diabetes experiencing a more significant number of deaths (Iyanda et al. 2020; CDC COVID 2021; Gu et al. 2020). Table 1 provides descriptions of the datasets and variables.

3.1 Ordinary least squares (OLS) model
In this study, we first applied OLS models to the variables to observe the spatial associations between mortality and the other six explanatory factors (health, socioeconomic, and demographic factors). The OLS regression model assumes the correlation between the dependent

Table 1 Description of the datasets

	Variables	Descriptions	Source	Scale
Independent Variable	Mobility	Mean home-dwell time	Safegraph (Safegraph 2021)	Census level in 2021, daily
	Income	The median household income	NHGIS (NHGIS 2021)	County level in 2021
	Non-Insurance	Rate of people who don't have insurance coverage	NHGIS (NHGIS 2021)	County level in 2021
	Old	Rate of people who are older than 60		County level in 2021
	Diabetes	Rate of diabetes cases in a county. (Age-adjusted percentage diagnosed diabetes)	Division of Diabetes Translation, CDC (CDC, Atlas 2018)	County level in 2021
	Heart disease	Rate of heart disease cases in a county	Division of Diabetes Translation, CDC (CDC, Atlas 2018)	County level in 2021
Dependent Variable	COVID-19 mortality	COVID-19 mortality rate	New York Times daily report (Covid GitHub 2020)	County level in 2021, daily

Table 2 Ordinary least squares (OLS) model results in July 2021

Variable	Coefficient	St. Errors	T-Statistic	*P*-value	VIF
Intercept	0.001	0.009	0.000	0.972	-
Diabetes	0.011	0.011	3.065	0.017	1.832
Heart disease	0.032	0.014	5.368	0.008	1.765
Old	0.978	0.017	8.596	0.001	1.239
Household income	-0.015	0.014	3.368	0.013	1.224
Noinsurance	-0.281	0.027	-1.235	0.035	1.372
Mobility	0.078	0.017	4.086	0.011	1.132
AICc	4114.392				
Adjusted R^2	0.471				
R^2	0.485				
RMSE	4325.51				

RMSE Root Mean Square Error

variable and explanatory factors is stationary and constant (Mahanty et al. 2021). It calculates a global model for variables, generating one equation for the entire dataset (Bacha 2003; Batisani and Yarnal 2009; Geri et al. 2010; Ali et al. 2007). Here, the dependent variable is the

COVID-19 mortality rate, and the independent variables refer to Mobility, Income, Non-Insurance, Old, Diabetes, and Heart disease, described in Table 1. The results of the OLS model are shown in Table 2. Table 2 shows that the OLS has a low-performance R^2 value, which indicates that the model only explained 47% of the independent variable. Also, it can reveal that the OLS model may not be able to capture all changes in the data, especially in the presence of spatial heterogeneity.

3.2 Multiscale geographically weighted regression (MGWR) model

According to Table 2, OLS has produced a low R^2 value (0.471), which indicates poor model fit outcomes. Therefore, a local regression model such as MGWR would be a better choice to improve the model accuracy and observe the spatial variations at the local level. The MGWR study uses a rigorous and comprehensive approach to model selection and parameter estimation to produce accurate and robust results. The statistical analysis results of the MGWR model for each month (July, August, and September 2020) are summarized in Tables 3, 4 and 5. Figure 1 illustrates the spatial distribution of local R^2 in July.

Table 3 Multiscale geographically weighted regression model for July 2021

	Variable		Diabetes	Heart Disease	Income level	Mobility	Old	No insurance
COVID-19 Mortality	Coefficients	Mean	0.007	0.023	-0.081	-0.013	0.45	0.167
		Median	0.059	0.231	-0.015	-0.008	-0.206	0.028
		STD	0.003	0.001	0.028	-0.001	0.224	0.019
	Bandwidth		128	79	173	43	66	87
	Adjusted R^2		0.861					
	AIC$_c$		3526.748					
	R^2		0.883					
	RMSE		2102.99					

Note: STD Standard deviation, *RMSE* Root Mean Square Error

Table 4 Multiscale geographically weighted regression model for August 2021

	Variable		Diabetes	Heart Disease	Income level	Mobility	Old	No insurance
COVID-19 Mortality	Coefficients	Mean	0.003	0.007	-0.081	-0.035	0.276	0.167
		Median	0.003	0.003	-0.015	-0.011	0.255	0.067
		STD	0.001	0.000	0.028	0.012	0.189	0.013
	Bandwidth		149	86	274	113	235	182
	Adjusted R^2		0.864					
	AIC$_c$		2926.894					
	R^2		0.879					
	RMSE		2263.28					

Note: STD Standard deviation, *RMSE* Root Mean Square Error

Table 5 Multiscale geographically weighted regression model for September 2021

	Variable		Diabetes	Heart Disease	Income level	Mobility	Old	No insurance
COVID-19 Mortality	Coefficients	Mean	0.008	0.014	-0.041	-0.112	0.081	0.040
		Median	0.014	0.014	-0.030	-0.113	0.084	0.045
		STD	0.032	0.001	0.028	0.003	0.049	0.023
	Bandwidth		131	171	354	66	125	245
	Adjusted R^2		0.882					
	AIC$_c$		2175.519					
	R^2		0.895					
	RMSE		2204.54					

Note: STD Standard deviation, *RMSE* Root Mean Square Error

Figure 1 represents the missing/null value data in white. The missing values appeared here due to the need for some data. For example, some counties lacked mobility data. Results showed that most counties received a high local R^2 value (over 0.75) except for some North Dakota and Montana counties. The results of the statistical analysis are illustrated in Table 3. According to Table 3, the MGWR model has a much higher adjusted R^2 (over 0.8 for July, August, and September) value compared to the OLS model (0.471) and a lower AICc value (2876.387) compared to the OLS model (4114.932). The MGWR model performs better than the OLS model in predicting COVID-19 mortality, mainly due to its ability to better handle spatial heterogeneity in data and provide more accurate local predictions. This performance is reflected in comparing the RMSE values of the two models. In addition, the bandwidth values of different parameters in the MGWR model vary in different months (as shown in Tables 3, 4 and 5), proving that the model can capture spatial heterogeneity at different time points. The dynamic changes in bandwidth reveal that the model can flexibly adapt to the spatial distribution characteristics of different periods, thereby providing more accurate analysis and prediction at the local level. Therefore,

the selection of the MGWR model is not only due to its overall statistical performance superior to the OLS model but also its unique advantages in spatial heterogeneity analysis.

3.3 Geographically temporal weighted regression (GTWR) model

Even though the MGWR has produced reliable results for the model fit (R^2) and AIC$_c$ value, the spatiotemporal kernel function, which consists of mixed spatial and time bandwidths, does not always seem reasonable (Fotheringham et al. 2015; Fotheringham et al. 1998). The MGWR function must first find and fix an optimized spatial bandwidth and the optimized temporal bandwidth to calculate the spatiotemporal weight, indicating that MGWR could not simultaneously optimize both temporal and spatial bandwidths. Therefore, we applied GTWR to estimate the temporal and spatial weights and to analyze the spatiotemporal associations between the variables. The statistical analysis results in Table 6 illustrate that the GTWR model has a higher adjusted R^2 value, lower AIC$_c$ value and RMSE value than the OLS model. GTWR produced a slightly higher R^2 than MGWR models, while it owned a lower RMSE value and AIC$_c$ value.
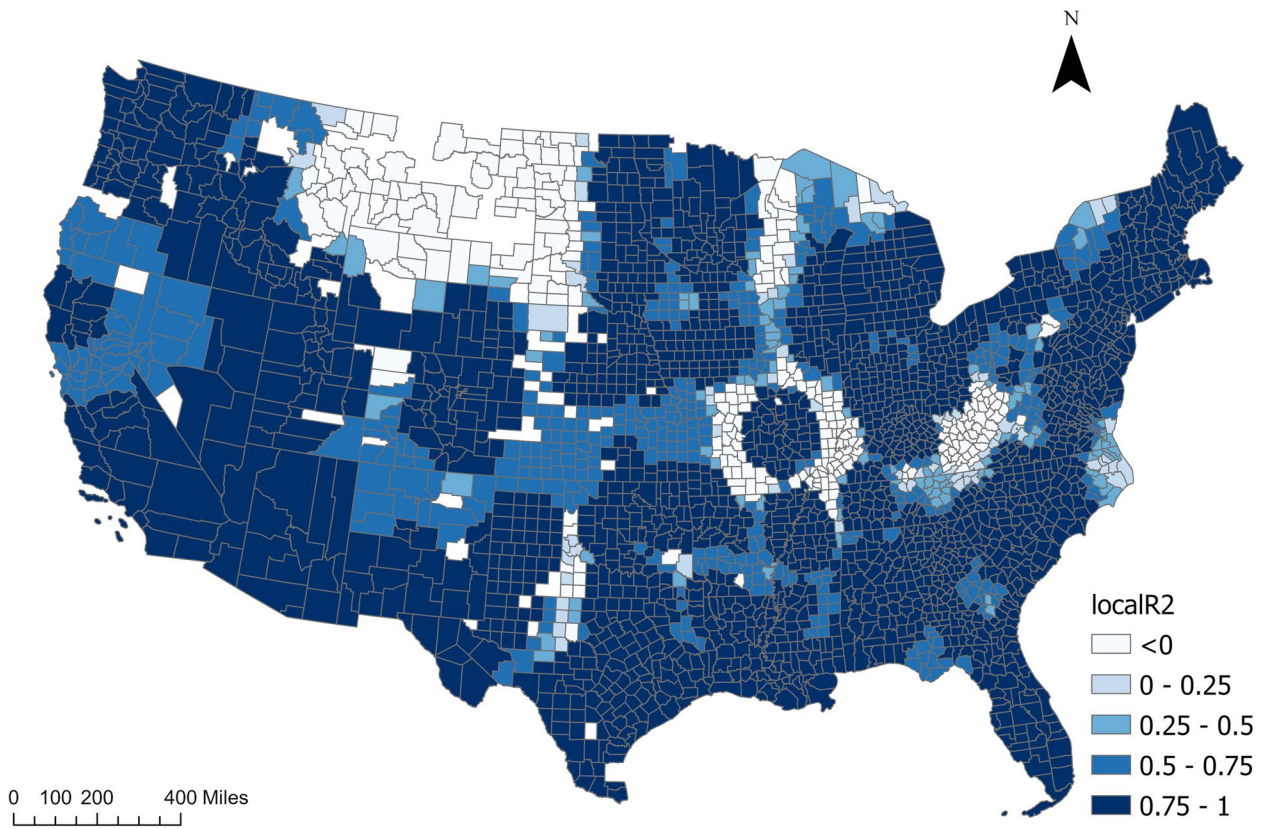
Hu *et al. Computational Urban Science*   (2024) 4:6

Page 6 of 13

Fig. 1 Local R square map in July 2021

Table 6 Geographically temporal weighted regression model results

Variable	Min	LQ	Mean	UQ	Max
Intercept	-4.652	-2.916	-1.167	2.691	5.224
Diabetes	1.126	2.335	4.012	5.173	12.860
Heart disease	-1.102	-0.245	4.896	6.287	7.214
Old	0.531	1.821	6.764	8.292	11.239
Household income	-4.681	-3.889	-2.957	-2.055	0.290
Noinsurance	-1.294	-0.198	3.893	4.321	5.501
Mobility	-6.117	-5.028	-4.662	-1.698	1.582
AICc	2656.38				
AdjustR^2	0.91				
R^2	0.93				
RMSE	1446.37				

Note: LQ Lower Quartile, *UQ* Upper Quartile, *RMSE* Root Mean Square Error

4 Results

This section introduces the key findings of GTWR results. In this project, we used six independent variables (number of diabetes and heart disease cases, income level, mobility, number of older people and non-insurance population) to test the relationship with the COVID-19 mortality rate. According to Table 2, the OLS model produced a low R^2 value, which indicates poor model performance (Fotheringham et al. 2017; Oshan et al. 2020). Both MGWR and GTWR have produced good R^2 and AICc values, indicating the models' robust goodness of fit. Figures 2, 3, 4, 5, 6 and 7

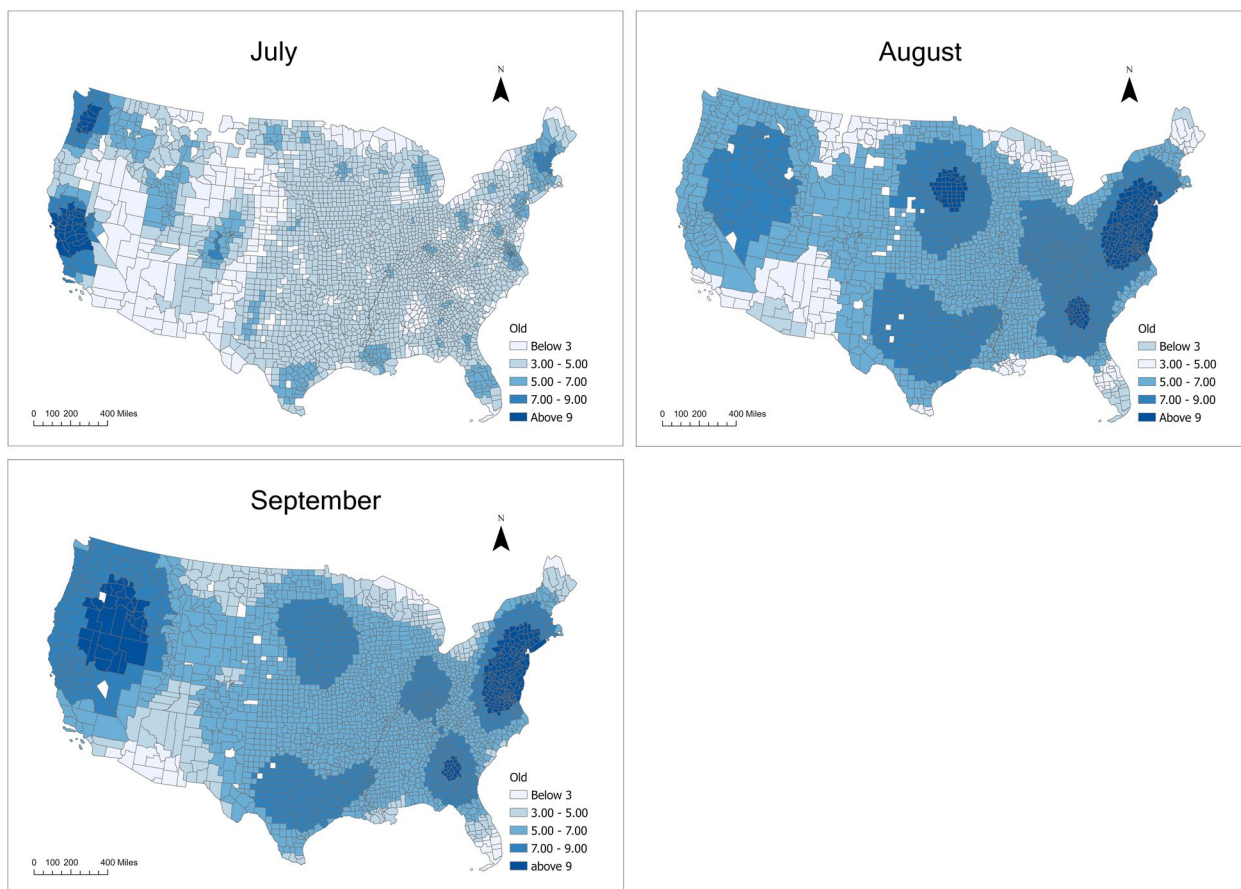Hu *et al. Computational Urban Science*   (2024) 4:6

Page 7 of 13

Fig. 2 Local effects of old factor

illustrate the local effects (coefficient maps) of independent variables using GTWR.

The degree of correlation between a county and its COVID-19 mortality rate can be represented on a map using a color gradient, with darker shades of color indicating higher correlation and lighter shades of color indicating lower correlation. According to Fig. 2, the importance of the old factor in COVID-19 mortality has increased over the study period. Although there was still a positive correlation in July, the correlation was relatively low in most regions. However, as time progressed, this factor became increasingly important. The Arizona cluster in the southwest, Maine in the northeast, and Florida in the southeast, appear to be relatively insensitive to time, as their COVID-19 mortality rates remained consistently high throughout the study period. The number of older adults in a region positively correlates with COVID-19 mortality rates. This relationship is statistically significant across most areas, except for a few small regions where the correlation is weak or nonexistent. These findings highlight the importance of protecting vulnerable elderly populations in the fight against the COVID-19 pandemic, particularly in areas where older people comprise a significant proportion of the population.

Figure 3 shows distinct spatial clusters in areas such as Arizona, Nevada, and California. Interestingly, these clusters exhibit a relative insensitivity to temporal changes compared to other regions, suggesting unique regional dynamics at play. Notably, the impact of income on COVID-19 mortality is intensified over the study period, with a notable new cluster emerging around the middle of September, particularly in New Mexico and Arizona. In contrast, some regions show a positive correlation with the pandemic, indicating a diminishing significance of the household income factor in these areas. Figure 4 presents a clustering pattern that mirrors the observations made in Fig. 3, with both figures highlighting clusters in the central and southeastern areas of the study region. However, it is crucial to interpret these findings cautiously, as they represent correlations rather than direct causations.

In the southern U.S., particularly in Texas, a cluster negatively correlated with mortality is observed,
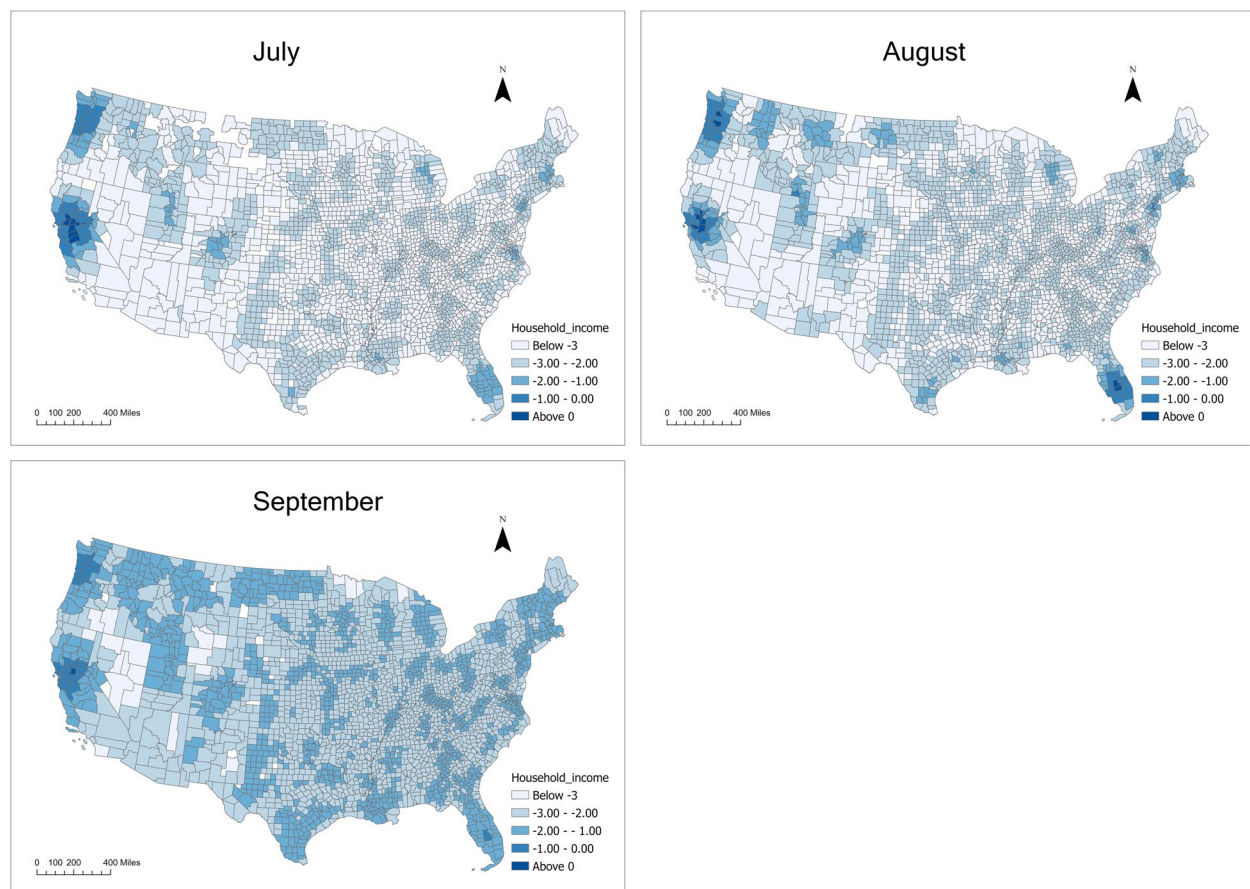
Hu *et al. Computational Urban Science*   (2024) 4:6

Page 8 of 13

Fig. 3 Local effects of median household income factor

suggesting a lower COVID-19 risk among uninsured populations. This counterintuitive finding coincides with when Texas reported a decrease in daily case and fatality numbers. Such observations underscore the complex interplay of socioeconomic factors in the pandemic's trajectory and highlight the importance of regional analyses in understanding COVID-19's impact.

Figure 5 shows a positive correlation between the heart disease factor and COVID-19 mortality across all regions, indicating that patients in areas with a higher prevalence of heart disease were more likely to experience higher levels of mortality, with a few exceptions. This pattern could be attributed to the already low mortality rates from coronary heart disease in areas like Kansas and Nebraska, potentially due to effective healthcare strategies and public health interventions targeting heart disease in these regions. Such interventions could have inadvertently strengthened the population's resilience against COVID-19. Additionally, these findings reflect broader socio-economic and

health system factors that contribute to these states' overall lower mortality rates.

Similar to the findings for heart disease, most regions showed a positive correlation between diabetes prevalence and COVID-19 mortality rates, suggesting that areas with a higher prevalence of diabetes patients were more likely to experience higher COVID-19 mortality rates in Fig. 6. Intriguingly, a shift is observed in August, with the central region, particularly a small cluster in the heart of Texas, demonstrating a sudden negative correlation with mortality rates. The persistence of a cluster in central Texas across various local effect maps suggests a regional anomaly. This unusual pattern could be linked to Texas experiencing a significant reduction in daily reported COVID-19 cases and fatalities during that time, potentially influencing the overall mortality correlation in the state. This reduction may reflect the impact of localized public health initiatives, changes in community behavior, or other region-specific factors, such as the timing of virus waves or the introduction of health policies and interventions. The emergence of this
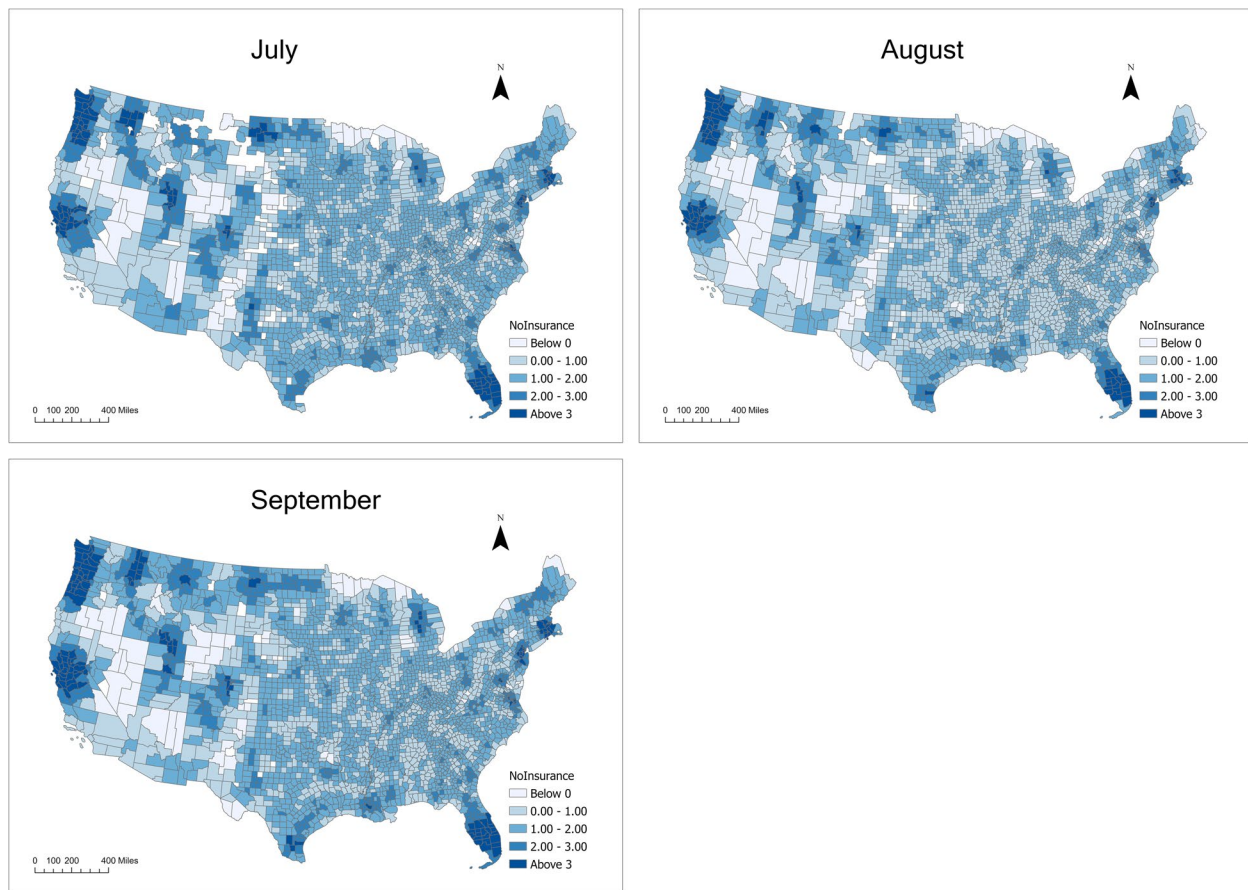
Fig. 4 Local effects of noinsurance factor

cluster warrants further investigation to understand the confluence of factors that contributed to the observed decrease in COVID-19 mortality in Texas during the study period.

In assessing the local effects of mobility on the epidemiological landscape, Fig. 7 provides a temporal choropleth representation. A clear temporal transition is evident, with mobility levels varying significantly across regions. In July, most counties exhibit low mobility levels, with darker shades concentrated in specific areas, suggesting stricter adherence to mobility restrictions or potentially less need for movement. As the summer progresses, the map for August reveals an increased mobility range, with a substantial number of counties shifting towards lighter shades of blue, particularly in the Midwest and Southern regions. This shift could indicate a relaxation of restrictions or an adaptation to new norms of travel and movement. By September, the distribution of mobility levels becomes more homogeneous, with many regions returning to near-normal mobility levels, indicated by the

prevalence of the 'above 0' category in the legend. The increase in mobility could be correlated with various factors, including the reopening of schools and businesses and an increase in travel and economic activity.

5 Discussion

This study marks a significant advancement in understanding the spatiotemporal dynamics of the COVID-19 pandemic by investigating its relationship with six crucial explanatory factors, including mobility, heart disease prevalence, diabetes prevalence, median household income, number of older adults, and insurance coverage, applying the MGWR model and GTWR model. While GTWR has been utilized in several studies, our research distinctively contributes by integrating these specific health and mobility factors into a spatial–temporal framework, thereby shedding light on how these variables collectively influence the pandemic's dynamics over time and across different regions. The MGWR and GTWR models improve the diagnostic information (AICc and R^2) from OLS,
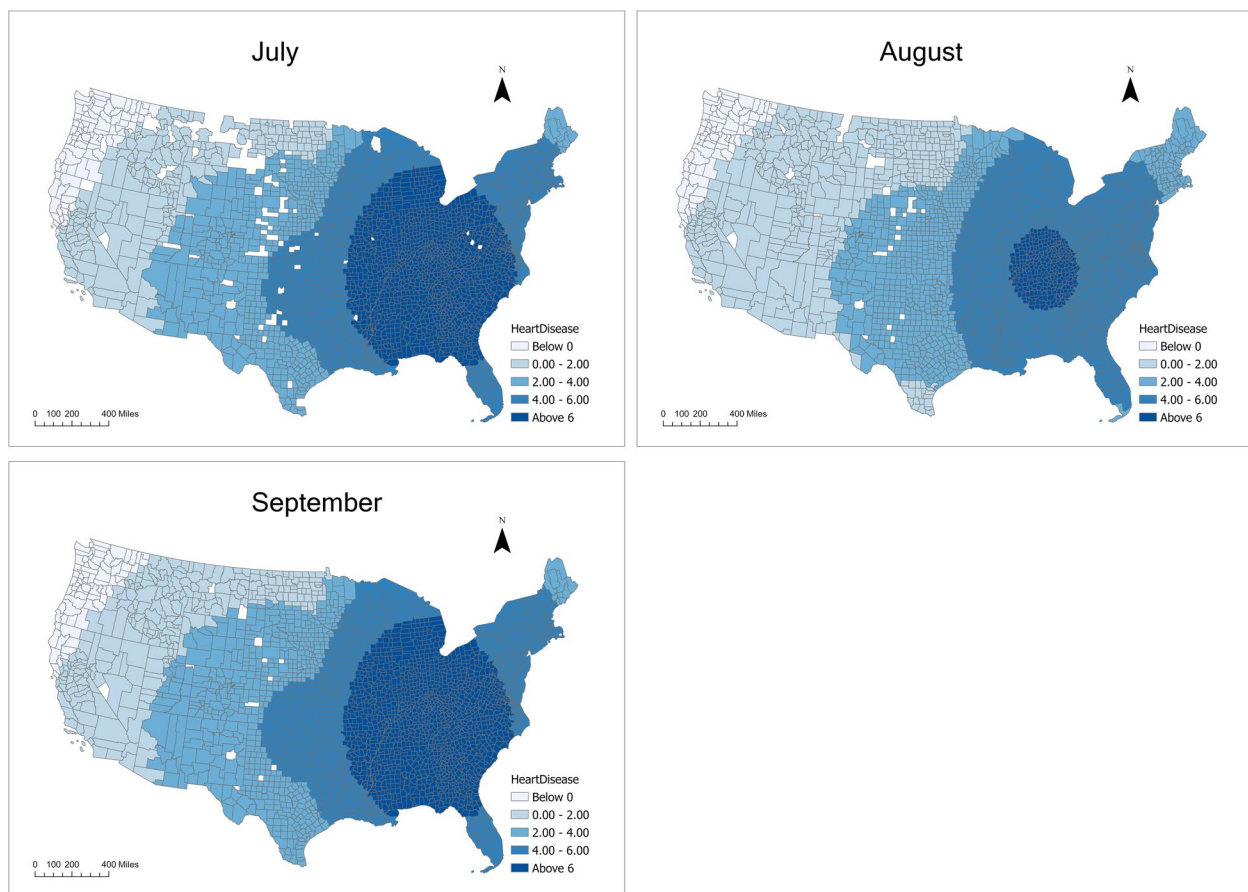
Fig. 5 Local effects of heart disease factor

so they perform better on our dataset. They can better explain the total variations in this issue. The direction of variables is in line with the expectation.

The positive associations of diabetes prevalence, heart disease prevalence, and the number of older adults with COVID-19 mortality align with existing literature that highlights the vulnerability of individuals with pre-existing conditions to the pandemic (Iyanda et al. 2020; CDC COVID 2021; Gu et al. 2020). However, the diminishing influence of the older adult population on mortality over time warrants a nuanced exploration. These results could suggest a successful implementation of protective measures for this demographic or a shift in the pandemic's impact on other age groups.

Contrarily, the negative association of median household income with mortality suggests socio-economic status as a buffer against the pandemic's worst outcomes, potentially due to better access to healthcare resources or the ability to adhere to protective measures such as social distancing (Ehlert 2021). However, the uniformity of this pattern across the country suggests underlying systemic factors at play that transcend local variations. Moreover, the fluctuating significance of the mobility factor raises questions about its role in the spread of COVID-19. While generally negatively related to mortality, indicating the effectiveness of stay-at-home measures, its varying sensitivity across regions points to the complex interplay of local policies, community compliance, and perhaps even cultural attitudes towards mobility and social interaction. There are some data limitations in this research. For example, only one temporal data might limit the performance of the GTWR model, while the results seem fine. Another is the limitation of mobility data types. Comparing different mobility indexes in the same model can improve the results. Due to the health data availability, we cannot access temporal health data to enhance the results.
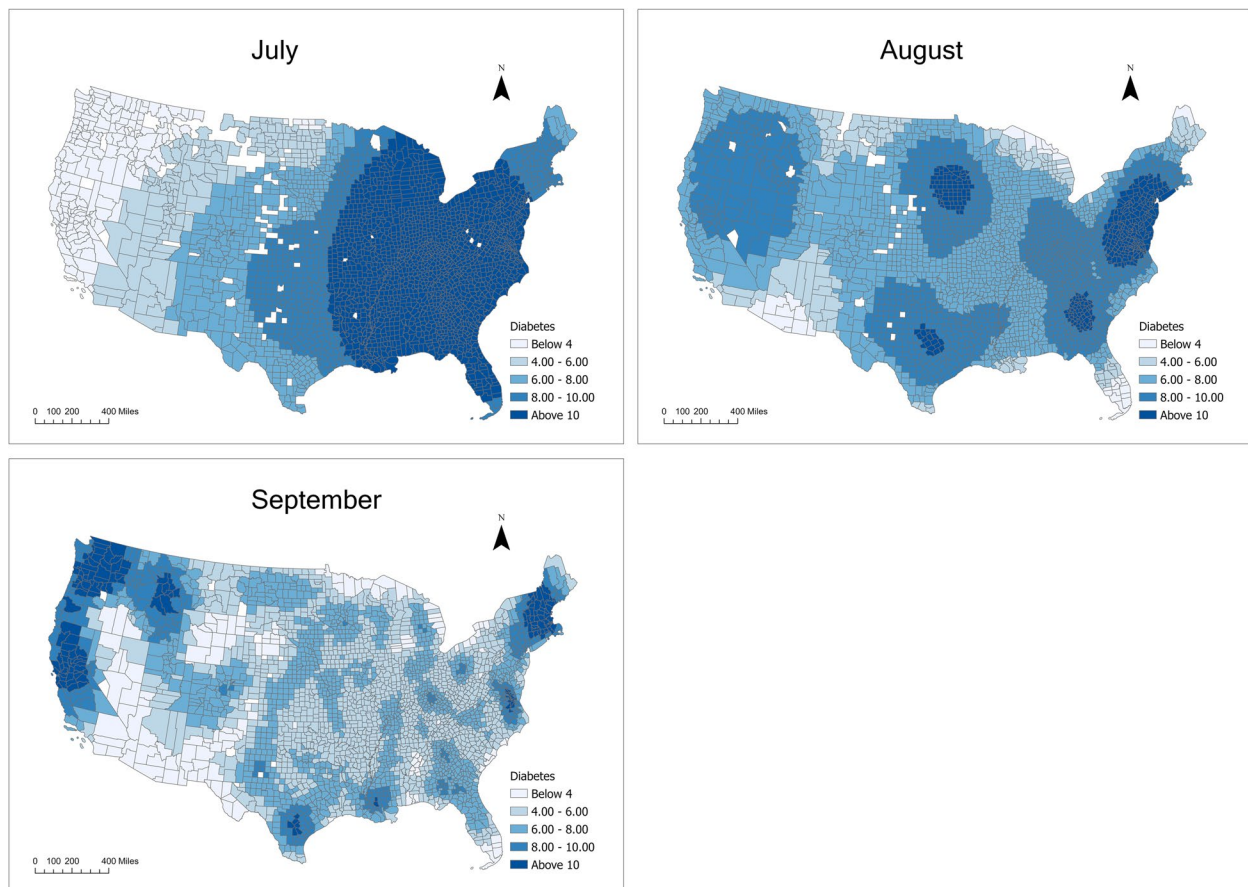
Fig. 6 Local effects of diabetes factor

6 Conclusion

The study investigated the significant factors contributing to the mortality rate of COVID-19 in the United States. Employing the MGWR and GTWR models, the study analyzed six variables: health, socioeconomic, and mobility. The results revealed that the GTWR and MGWR models could explain up to 91% and 88% of the variations in COVID-19 mortality, respectively. The findings highlight the critical role of mobility and health conditions related to COVID-19 in controlling the prevalence of the pandemic. The study indicates that policymakers should pay close attention to individuals with coronary heart disease and diabetes and their mobility. Moreover, the results show that the mean home-dwelling time in most counties negatively impacts the spread of COVID-19, and this impact decreased from July to August and then increased from August to September. The study also revealed that higher median household income could reduce mortality.

Additionally, the number of older adults is essential in addressing this issue, while the role of this factor needs to be more critical in the study. While the importance of addressing the needs of older adults is recognized, future research should explore the nuanced roles of various demographics in pandemic response. The positive correlation between lack of insurance and increased mortality underscores the need for a policy focusing on healthcare access and coverage. The study recommends that policymakers focus on individuals without insurance, as their lack of access to healthcare is positively related to the mortality rate.
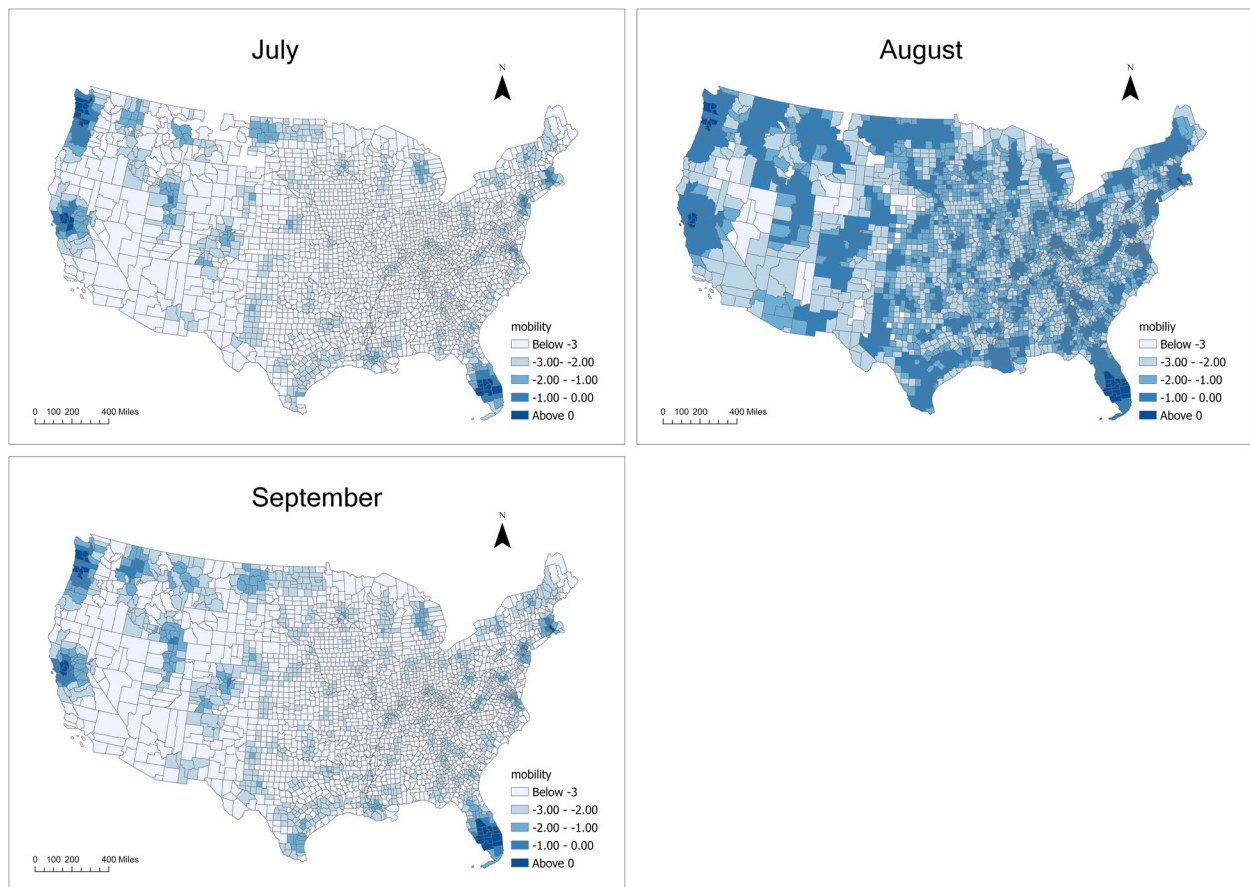
Fig. 7 Local effects of mobility factor

The GTWR and MGWR models are crucial tools for analyzing spatiotemporal COVID-19 issues. They can aid federal and state agencies in resource allocation and applying lockdown policies. This study provides valuable insights into the factors contributing to COVID-19 mortality and the results can help governors make evidence-based policies for different regions.

Authors' contributions
Conception or design of the work (Nanzhou Hu, Zhe Zhang, Bahar Dadashova, Nicholas Duffield); Data Collection (Nanzhou Hu, ziyi Zhang), Data Analysis (Nanzhou Hu), Drafting the article (Nanzhou Hu and Zhe Zhang), Critical Revision of the Article (Nanzhou Hu, Zhe Zhang, Bahar Dadashova, Jason Wu, Xinyue Ye, Siyu Yu, Daikwon Han), Final approval of the version to be published (all).

Availability of data and materials
Data used in this article is available to access via the Safegraph website, CDC Atlas data, New York Times COVID-19 data, and NHGIS website. Source are introduced in the Methods section.

Declarations

Consent for publication
This paper does not contain any individual person's data in any form.

Competing interests
The authors have no conflicts of interest to declare.

References
Abdi, A., Jalilian, M., Sarbarzeh, P. A., & Vlaisavljevic, Z. (2020). Diabetes and COVID-19: A systematic review on the current evidences. *Diabetes Research and Clinical Practice, 166*, 108347.

Ali, K., Partridge, M. D., & Olfert, M. R. (2007). Can geographically weighted regressions improve regional analysis and policy making? *International Regional Science Review, 30*(3), 300–329.

Hu *et al. Computational Urban Science*   (2024) 4:6

Page 13 of 13

Bacha, C. J. (2003). The determinants of reforestation in Brazil. *Applied Economics, 35*(6), 631–639.

Batisani, N., & Yarnal, B. (2009). Urban expansion in Centre County, Pennsylvania: Spatial dynamics and landscape transformations. *Applied Geography, 29*(2), 235–249.

Berry, I., Soucy, J. P. R., Tuite, A., & Fisman, D. (2020). Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *Cmaj, 192*(15), E420.

Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician), 47*(3), 431–443.

CDC, Atlas (2018). Interactive Atlas of Heart disease and Stroke. Retrieved from https://nccd.cdc.gov/DHDSPAtlas/?state=County .(Accessed April 2022)

CDC, COVID (2021). Retrieved from the https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-withmedical-conditions.html. (Accessed April 2022)

CDC, spread (2021). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/faq.html#Spread .(Accessed July 2022)

Covid Github (2020). Coronavirus (Covid-19) Data in the United States (Archived). Retrieved from https://github.com/nytimes/COVID-19-data. (Accessed April 2022)

Cui, W., Hu, N., Zhang, S., Li, D., Martinez, L., Goldberg, D., Güneralp, B., & Zhang, Z. (2022). Analyzing spatial variations of heart disease and type-2 diabetes: A multi-scale geographically weighted regression approach. *Computational Urban Science, 2*(1), 34.

Deller, S. C., & Lledo, V. (2007). Amenities and rural Appalachia economic growth. *Agricultural and Resource Economics Review, 36*, 107–132. 1203-2016-95353.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases, 20*(5), 533-534.

Ehlert, A. (2021). The socio-economic determinants of COVID-19: A spatial analysis of German county level data. *Socio-Economic Planning Sciences, 78*, 101083.

Fortaleza, C., Guimaraes, R. B., de Almeida, G. B., Pronunciate, M., & Ferreira, C. P. (2020). Taking the inner route: Spatial and demographic factors affecting vulnerability to COVID-19 among 604 cities from inner Sao Paulo State, Brazil. *Epidemiology and Infection, 148*, e118.

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A, 30*(11), 1905–1927.

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis, 47*(4), 431–452.

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers, 107*(6), 1247–1265.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. Chichester, England: Hoboken, NJ, USA, Wiley.

Fu, Z., & Li, R. (2020). The contributions of socioeconomic indicators to global PM2.5 based on the hybrid method of spatial econometric model and geographical and temporal weighted regression. *Sci Total Environ, 703*, 135481.

Geri, F., Amici, V., & Rocchini, D. (2010). Human activity impact on the heterogeneity of a Mediterranean landscape. *Applied Geography, 30*(3), 370–379.

Gu, T., Chu, Q., Yu, Z., Fa, B., Li, A., Xu, L., Wu, R., & He, Y. (2020). History of coronary heart disease increased the mortality rate of patients with COVID-19: A nested case-control study. *British Medical Journal Open, 10*(9), e038976.

Huang, X., Zhang, R., Li, X., Dadashova, B., Zhu, L., Zhang, K., & Shen, B. (2022). Health-based geographic information systems for mapping and risk modeling of infectious diseases and COVID-19 to support spatial decision-making. *Translational Informatics: Prevention and Treatment of Viral Infections,* 167–188.

Hu, S., Xiong, C., Younes, H., Yang, M., Darzi, A., & Jin, Z. C. (2022). Examining spatiotemporal evolution of racial/ethnic disparities in human mobility and COVID-19 health outcomes: Evidence from the contiguous United States. *Sustainable Cities and Society, 76*, 103506.

Iyanda, A. E., Adeleke, R., Lu, Y., Osayomi, T., Adaralegbe, A., Lasode, M., Chima-Adaralegbe, N. J., & Osundina, A. M. (2020). A retrospective cross-national examination of COVID-19 outbreak in 175 countries: A multiscale geographically weighted regression analysis (January 11-June 28, 2020). *Journal of Infection and Public Health, 13*(10), 1438–1445.

Jiao, J., Chen, Y., & Azimian, A. (2021). Exploring temporal varying demographic and economic disparities in COVID-19 infections in four U.S. areas: Based on OLS, GWR, and random forest models. *Comput Urban Sci, 1*(1), 27.

Johns Hopkins University Center for Systems Science and Engineering (2023). COVID-19 dashboard. Retrieved from https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6 (Accessed February 2023)

Leung, Y., Mei, C.-L., & Zhang, W.-X. (2000). Statistical tests for spatial non-stationarity based on the geographically weighted regression model. *Environment and Planning A, 32*(1), 9–32.

Li, H., Li, H., Ding, Z., Hu, Z., Chen, F., Wang, K., & Shen, H. (2020). Spatial statistical analysis of coronavirus disease 2019 (COVID-19) in China. *Geospatial Health, 15*(1), 1.

Liu, F., Wang, J., Liu, J., Li, Y., Liu, D., Tong, J., Li, Z., Yu, D., Fan, Y., Bi, X., Zhang, X., & Mo, S. (2020). Predicting and analyzing the COVID-19 epidemic in China: Based on SEIRD, LSTM and GWR models. *Plos One, 15*(8), e0238280.

Mahanty, C., Kumar, R., & Mishra, B. K. (2021). Analyses the effects of COVID-19 outbreak on human sexual behaviour using ordinary least-squares based multivariate logistic regression. *Quality & Quantity, 55*, 1239–1259.

McMillen, D. P. (1996). One hundred fifty years of land values in Chicago: A nonparametric approach. *Journal of Urban Economics, 40*(1), 100–124.

Mollalo, A., Sadeghian, A., Israel, G. D., Rashidi, P., Sofizadeh, A., & Glass, G. E. (2018). Machine learning approaches in GIS-based ecological modeling of the sand fly Phlebotomus papatasi, a vector of zoonotic cutaneous leishmaniasis in Golestan province, Iran. *Acta Tropica, 188*, 187–194.

NHGIS (2021). Retrieved from https://data2.nhgis.org/main (Accessed April 2022)

Oshan, T. M., Smith, J. P., & Fotheringham, A. S. (2020). Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. *International Journal of Health Geographics, 19*, 1–17.

Rogers, M. Y. (1999). Getting started with Geographic Information Systems (GIS): A local health department perspective. *Journal of Public Health Management and Practice, 5*(4), 22–33.

Safegraph (2021). Retrieved from https://marketplace.deweydata.io/#/datasets/SafeGraph .(Accessed April 2022)

Steiger, E., Rass, S., Seidel, A., Kroll, L., & Czihal, T. (2021). COVID-19 Vaccination in medical practices in Germany. *Deutsches Ärzteblatt International, 118*(44), 756–757.

Wu, X., & Zhang, J. (2021). Exploration of spatial-temporal varying impacts on COVID-19 cumulative case in Texas using geographically weighted regression (GWR). *Environmental Science and Pollution Research International, 28*(32), 43732–43746.

Zhang, Z., Demšar, U., Rantala, J., & Virrantaus, K. (2014). A fuzzy multiple-attribute decision-making modelling for vulnerability analysis on the basis of population information for disaster management. *International Journal of Geographical Information Science, 28*(9), 1922–1939.

Zhang, Z., Zou, L., Li, W., Usery, L., Albrecht, J., & Armstrong, M. (2021). Cyberinfrastructure and intelligent spatial decision support systems. *Transactions in GIS, 25*(4), 1651–1653.

Zhang, Z., Hu, H., Yin, D., Kashem, S., Li, R., Cai, H., & Wang, S. (2020). A cyberGIS-enabled multi-criteria spatial decision support system: a case study on flood emergency management. In *Social Sensing and Big Data Computing for Disaster Management* (pp. 167–184). Routledge

Publisher's Note