Energy Research & Social Science 97 (2023) 102949

Contents lists available at ScienceDirect

Energy Research & Social Science

journal homepage: www.elsevier.com/locate/erss

Original research article

Community vulnerability is the key determinant of diverse energy burdens in the United States

Zhenglai Shen [a], Chien-fei Chen [b,*], Hongyu Zhou [b,c], Nina Fefferman [d], Som Shrestha [a]

[a] Buildings and Transportation Science Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[b] Institute for a Secure and Sustainable Environment, University of Tennessee, Knoxville, TN, USA
[c] Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, USA
[d] Department of Math & Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA

ARTICLE INFO

Keywords:
Energy burden
Energy poverty
Spatial analysis
COVID-19
Community vulnerability

ABSTRACT

Low-income households generally experience a high energy burden; however, the factors influencing energy burdens are beyond socio-economics. This study explores the relationships between the multidimensionality of community vulnerability factors and energy burden across multiple geospatial levels in the United States. Our study found the distribution of energy burden in 2020 showed a great deal of variety, ranging from a minimum of 2.93 % to a maximum of 30.45 % across 3142 counties. The results of non-spatial and spatial regressions showed that the vulnerability ranks of socioeconomic, household composition and disability, minority and language, household type and transportation, and COVID mortality rate are significant predictors of energy burdens at the national level. However, at the regional level, only socioeconomic, minority and language significantly influence energy burdens. Minority and language negatively impact energy burdens except for the South East-Central region. Additionally, our analyses highlight the need to consider community vulnerability indicators' spatial homogeneity and heterogeneity. At the national level, only the epidemiological factors index is a spatially homogeneous predictor; on the regional and state level, the spatially homogeneous predictors such as socioeconomic status, household composition and disability, and household type and transportation vary by region. Such a region-sensitive relationship between energy burden and the predictors indicates spatial heterogeneity. This study suggests policy recommendations through the lens of the multidimensionality of community vulnerability factors. Implementing flexible national energy policies while making particular energy assistance policies for the vulnerable population at the regional or state levels is essential.

1. Introduction

1.1. Background

Energy burden (EB) is the percentage of gross household income spent on energy costs [1], which refers to affordability and/or the lack of access to reliable energy services. It is linked to negative mental and physical health outcomes [2–4]. Higher EB indicates greater chances of falling into poverty [5], making trade-offs between utilities and other necessities (e.g., heat-or-eat dilemma [6,7]). Many who experience a high EB also engage in risky behaviors such as payday lending [8], using unsafe energy alternatives like wood and peat [9,10], and more. In the United States (U.S.), on average, underserved communities, including low-income, Black, Hispanic/Latino, multifamily, women, and renters consume less energy but have higher EB than their counterparts [11,12]. This disparity in high-energy poverty reflects historical racial, class, gender, and housing discrimination patterns, lending and housing policies, healthcare access, and wealth accumulation [2,11,13]. While recent studies have widely recognized the impacts of extreme events, including natural disasters and the COVID-19 pandemic, on the worsened EB of underserved communities [2,11,14], little research has demonstrated the multidimensional determinates of EB at multiple geospatial levels. This study moves beyond extant energy justice research to explain the diverse effects of spatial homogeneity and heterogeneity of EB and analyze the relationship between community vulnerability factors and EB at multiple spatial levels, including state, regional, and national levels.

Understanding how community vulnerability determines EB is a

* Corresponding author.
E-mail address: cchen26@utk.edu (C.-f. Chen).

https://doi.org/10.1016/j.erss.2023.102949
Received 13 July 2022; Received in revised form 27 December 2022; Accepted 9 January 2023
Available online 22 February 2023
2214-6296/Published by Elsevier Ltd.

critical step in developing suitable energy justice policies. To examine the relationship between community vulnerability and EB, this study used the indicators from the U.S. COVID-19 Community Vulnerability Index (CCVI) to analyze the multidimensional determinates of EB, including social economics, minority status and language, housing quality, transportation, and health-related factors. The CCVI is based on the U.S. Center for Disease Control (CDC)'s Social Vulnerability Index [15] but incorporates COVID-specific vulnerability indicators, including epidemiological factors and health system strength, to help identify communities that might be at a greater risk in 2020. Specifically, the CCVI metric ranks each geography (state, county, or census tract) to one another on a 0–1 scale, where zero indicates least vulnerable, and one indicates most vulnerable. The CCVI provides opportunities for unique, action-oriented solutions to policies for vulnerable communities and to assess the equity of current outcomes and responses [16]. Studies using the CCVI have examined the impacts of COVID-19 cases, mortalities, vaccination rates, stay-at-home orders, crowded housing units, economic impacts, employment, transportation, natural disasters, and so on among vulnerable populations [17]. For example, a study [18] found that people in vulnerable communities were 21 % more likely to be diagnosed and 47 % more likely to die from COVID-19, unadjusted for age and comorbidities. Other researchers [19] also discovered that areas with an increase in CCVI index were linked with lower vaccination rates in 2020. Using the CCVI index, many studies have found that high COVID-19 vulnerability is concentrated in the South of the U.S. [18,20,21]. Overall, researchers [21] discovered that most individuals in the U.S. lived in moderate to low vulnerability communities, followed by high and very high vulnerability. However, little research has used the CCVI to examine its relationship with EB. Notably, the relationship between health indicators, including individual health conditions and the health care system and EB is not often analyzed in the EB literature.

Meanwhile, the COVID-19 pandemic exacerbated the LMI income households' EB, predominantly minority households, such as African American and Hispanic households. According to a survey by Pew Research Center [22] in April 2020, the COVID-19 pandemic significantly influenced job or wage loss. The survey showed 61 % of Hispanic American households and 44 % of Black American households experienced a job or way loss which significantly increased from the survey data in March 2020 (before the pandemic). Moreover, the COVID-19 pandemic makes some Americans challenging to pay their monthly bills. The survey indicated that 48 % of Black adults, 44 % of Hispanic adults, and 26 % of white adults can only make partial payments. In addition, a few studies [23–25] showed that the COVID-19 pandemic had deepened the prevalence of energy insecurity among low-income households, and those who require the use of an electronic medical device experience higher rates of energy insecurity. Therefore, the impact of COVID-19-related variables on EB during 2020 cannot be neglected.

In addition, there is a growing recognition of the need to account for multidimensional perspectives in examining EB [11], as EB is linked to the interconnected factors of socioeconomics, household composition, transportation, health system, and so on. Importantly, EB can be considered a manifestation of distributional inequalities, particularly spatial inequalities, where space acts as a backdrop for inequalities and actively constructs and upholds inequalities [26–28]. Furthermore, the effects of predictors on EB are varied spatially and can be separated into spatially homogeneous and spatially heterogeneous variables. The spatially homogeneous variables impact EB in the target region, while the spatially heterogeneous variables have a heterogeneous effect across different regions [29]. As such, researchers have explored the impacts of spatial inequities on EB at a single level, including country [11,29], and city [30–32].

The severity of EB presents a distinct phenomenon when examined at a different level. The patterns of spatial EB depend on the scale analysis employed and the material sites considered [28]. For example, approximately 25 % (30.6 million) of U.S. households face a high EB (pay >6 % of income), and 13 % (15.9 million) U.S. households face a difficult EB situation (pay >10 % of income), whereas 60 % (15.4 million) of low-income households (LIHs) face a severe EB [12]. At the regional level, the East South-Central region (i.e., Alabama, Kentucky, Mississippi, and Tennessee) in the U.S. has the highest percentage of households with high EB (38 %) compared to other regions 2017 [12]. During the 2020 COVID-19 pandemic, the EB in the U.S. demonstrated localized spatial and temporal effects at the regional level [11]. For example, in the Mountain West and Midwest, EB rose from 2014 and 2018 to 2020, and rural low-to-moderate income (LMI) households spent almost a quarter more on monthly utilities than their urban counterparts in 2020. On the city level, census blocks with higher percentages of poverty, household heads with racial/ethnic minority status, and individuals with less than a high school education had higher energy use intensities, which experienced poor energy efficiency and a higher EB. These studies [11,12] highlight how historically institutionalized racial and income segregation has affected the distribution of residential energy disparities.

While it is essential to recognize these EB differences, more research is needed to focus on the spatial effects with a systematic multiple spatial levels analysis and further identify the spatially homogeneous and heterogeneous factors contributing to EB. More importantly, little research has investigated the influence of interconnected factors of comprehensive community vulnerability indexes, such as socioeconomics, household composition, disabilities, minority and language, transportation, and epidemiological aspects, on EB.

1.2. Purpose of the study

This study attempts to answer three research questions: "*What is the distribution of EB across various counties, states, and regions during the 2020 COVID-19 pandemic*? "*What are the key community vulnerability determinants influencing EB*? And "*What are geospatial patterns, i.e., spatial homogeneity and spatial heterogeneity, associated with EB, and does the geospatial pattern have a relation with the regional level*?" Here we analyze the EB situations among the low-to-medium income (LMI) households in the U.S. using several nationally representative data sets (see Method). Our study provides a unique opportunity to examine the spatial homogeneity and heterogeneity associated with EB across nine census regions, 3124 counties, 50 states, and the five highest EB states during the COVID-19 pandemic in 2020.

Additionally, we seek to understand the diverse EB situations that community-level factors, including socioeconomics, household composition, minority status, disabilities status, transportation, epidemiological characteristics, and COVID cases and mortality rates, can explain. The 2020 EB is selected because it is the start of the COVID pandemic, and it is expected the EB will have differences from the previous period. In addition, such public emergency provides a suitable candidate to study their effects on EB due to people's behavior change from either mandatory staying-at-home policy or the nature of risk-avoiding. Finally, this study wants to understand if the community-level factors influence EB still hold despite the influence COVID-19 pandemic.

2. Methodology

2.1. Measures

2.1.1. Dependent variable: annual average energy burden

The 2020 county-level monthly EB was estimated from the 2014–2018 county-level LMI annual income ($\text{Income}_{lm}^{2014-2018}$) and 2020 state-level energy expenditure in electricity, fuel, and natural gas ($\sum_1^q \text{Exp}_{lmq}^{2020}$), based on a previous EB study [11], as shown in Eq. (1). This is mainly because the 2020 EB data was unavailable from the U.S. Department of Energy's Low-Income Energy Affordability Data (LEAD) Tool [22]. Therefore, the 2020 energy expenditure was estimated using the 2014–2018 county-level energy expenditure data (electricity, fuel, and natural gas) and the energy consumption data, along with the 2020 county-level energy consumption data. The 2014–2018 county-level LMI households' annual income was estimated by the energy expenditure data divided by its EB. Meanwhile, the energy expenditure of 2020 was calculated by considering the 2014–2018 energy expenditure, 2014–2018 energy consumption, and 2020 energy consumption. Therefore, the 2020 EB can be estimated as follows:

$$\mathrm{EB}_{lm}^{2020} = \frac{\sum_{1}^{q} \mathrm{Exp}_{lmq}^{2020}}{\mathrm{Income}_{lm}^{2014-2018}}$$
$$= \left(\sum_{1}^{q} \mathrm{Exp}_{lq}^{2014-2018} \times \frac{\mathrm{Exp}_{lmq}^{2014-2018}}{\mathrm{Exp}_{lq}^{2014-2018}} \times \frac{\mathrm{Con}_{lmq}^{2020}}{\mathrm{Con}_{lmq}^{2014-2018}} \right) \Bigg/ \frac{\sum_{1}^{q} \mathrm{Exp}_{lq}^{2014-2018}}{\mathrm{EB}_{l}^{2014-2018}} \tag{1}$$

where $\mathrm{EB}_{lm}^{2020}$was the predicted 2020 EB for the l^{th} county in the m^{th} month for LMI households, $\mathrm{Exp}_{lmq}^{2020}$ was the energy expenditure for the q^{th} source (electricity, fuel, and natural gas) of the l^{th} county in the m^{th} month in 2020, and $\mathrm{Income}_{lm}^{2014-2018}$ was the average income of the l^{th} county in the m^{th} month during 2014–2018 for LMIs. $\mathrm{Exp}_{lmq}^{2020}$ was calculated by using energy expenditure for the q^{th} source of the l^{th} county during 2014–2018 ($\mathrm{Exp}_{lq}^{2014-2018}$), energy expenditure ratio in m^{th} month for the q^{th} source compared with the monthly average ($\mathrm{Exp}_{lmq}^{2014-2018}/\bar{\mathrm{Exp}}_{lq}^{2014-2018}$), and 2020 state-level residential energy consumption ratio for the q^{th} source of the l^{th} county in the m^{th} month compared with its counterpart during 2014–2018 ($\mathrm{Con}_{lmq}^{2020}/\mathrm{Con}_{lmq}^{2014-2018}$). $\mathrm{Income}_{lm}^{2014-2018}$ was calculated by using the sum of energy expenditure for the q^{th} source of the l^{th} county during 2014–2018 divided by EB for the l^{th} county for LMI households ($\mathrm{EB}_{l}^{2014-2018}$). The estimated monthly EB of 2020 was then averaged to obtain the annual average EB. In this study, the used monthly income was assumed to be constant. Therefore, the monthly average 2020 EB was equal to the one calculated using the 2020 annual energy expenditure divided by the 2014–2018 average annual income. Historical EB estimations among LMIs of 2014–2018 were collected from the LEAD Tool [22], while the energy expenditure data was obtained from the U.S. Energy Information Administration (EIA) [34].

2.1.2. Independent variables: Community Vulnerability Index and COVID-related variables

The independent variables include eight composite indexes, including the six themes of the COVID-19 Community Vulnerability Index (CCVIs) and the COVID-19 cases and mortality rates, with >40 variables. Specifically, factors used in this study follow six main themes: socioeconomic status (SE), household composition and disability (HD), minority and language (ML), household type and transportation (HT), epidemiological factors (EF), and healthcare system factors (HS). Table 1 summarizs the variables included in each theme. Readers can find the details of creating the CCVI composite measure on the website CCVI [16].

The two COVID-19-related independent variables included the accumulative percentages of COVID-19 cases (COVID case rate) and COVID-19 mortalities (COVID mortality rate) in 2020. The accumulative annual COVID-19 case and mortality rates calculated the county-level cases and mortality rates divided by the corresponding population. The COVID-19-related variables (COVID-19 case and mortality rates) were considered controlled variables given the extreme event of the COVID-19 pandemic. However, they are not the main focus of the study's aims but could influence the outcomes of EB. Previous research [11,12] had shown that the average EB of 2020 increased after March, compared with its counterparts of the five-year average EB of 2014–2018. This result was consistence with the time that the U.S. government issued the Stay-at-Home order (lockdown) on March/19th' 2020. The increased EB in our analysis might attribute to (1) the Stay-at-Home order influencing extra operational household energy (e.g., electricity and natural gas) by running HVAC and other necessary equipment at home; (2) people who lost their income during the pandemic [11]. This result is partly due to the COVID-19 case and mortality rates, where the higher the case and mortality rates, the more people stay at home or work from home, and more people lose their income.

Table 1
Six main themes of the COVID-19 community vulnerability index (CCVI) [16].

CCVI theme	Variables included
1. Socioeconomic Status (SE)	(1) Below poverty level (2) Unemployment (3) Income (4) Age 25 or older with no high school diploma (5) No health insurance
2. Household Composition and Disability (HD)	(1) Dependent children <18 years of age (2) Persons aged 65 years and older (3) Civilian with a disability (4) Single-parent households
3. Minority Status and Language (ML)	(1) Race/ethnicity (2) English-language proficiency
4. Housing and Transportation (HT)	(1) Multi-unit structures (2) Mobile homes (3) Access to indoor plumbing (4) Householdsld without a vehicle (5) Crowding (more people than rooms) (6) In institutionalized group quarters
5. Epidemiological risk factors (EF)	(1) Cardiovascular conditions (2) Respiratory conditions (3) Immune-compromised (4) Obesity (5) Diabetes (6) Population density (7) Influenza and pneumonia death rates
6. Healthcare system factors (HS)	(1) The number of hospital beds per 100,000 (2) Percent of the population with a primary care physician (3) Health spending per capita

2.1.3. Definition of regions

We study EB on the national, regional, and state levels using the county level (total of 3142 counites) data. At the national level, this study includes all the US states and the mainland of the U.S. except for Alaska and Hawaii, that is, the Contiguous US. The regional level consists of the nine census regions, i.e., New England, Middle Atlantic, East North-Central, West South-Central, South Atlantic, East South-Central, West South-Central, Mountain, and Pacific. Similar to the national-level analysis, we also included a connected Pacific region in the U.S. mainland (Contiguous Pacific region). Table 2 lists the states included in

Table 2
Name of the states within each census region [12].

Regions	States
New England	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
Middle Atlantic	New Jersey, New York, Pennsylvania
East North Central	Illinois, Indiana, Michigan, Ohio, Wisconsin
West North Central	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
South Atlantic	Delaware, DC, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia
East South Central	Alabama, Kentucky, Mississippi, Tennessee
West South Central	Arkansas, Louisiana, Oklahoma, Texas
Mountain	Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming
Pacific	Alaska, California, Hawaii, Oregon, Washington
Contiguous Pacific	California, Oregon, Washington

each region. On the state level, five states having the highest EB were considered. Counties in Alaska and Hawaii were excluded from spatial analysis because it is hard to test the influence of proximity; these states do not border other U.S. states, and in Hawaii, no county borders another [35].

2.1.4. Data sources

This study uses publicly and nationally representative available data. It includes the COVID-19 CCVI [36], the U.S. Department of Energy's Low-Income Energy Affordability Data (LEAD) [33], energy expenditure from the U.S. Energy Information Administration (EIA) [34], John Hopkins' COVID-19 data [37], and the social and demographic data from the U.S. Census Bureau American Community Survey (ACS) [38]. Links to the data sources used in Eq. (1) are listed in Table 3. The terms of $\mathrm{Exp}_{lq}^{2014-2018}$ and $\mathrm{Exp}_{lmq}^{2014-2018}$ come from Item (4); the terms of $\mathrm{Con}_{lmq}^{2014-2018}$ and $\mathrm{Con}_{lmq}^{2020}$ come from Items (5), (6), and (7); the term $\mathrm{EB}_{l}^{2014-2018}$ is from Item (3).

2.2. Analysis

A combination of bivariate, multivariate, and non-spatial and spatial analysis methods was employed to examine the relationships between EB and the eight independent variables (predictors). The analysis includes using both correlations (e.g., the distribution of EB, Pearson correlation, and global spatial autocorrelation analysis) and causation methods (non-spatial and spatial analysis). First, statistics characteristics of both EB and predictors were summarized, where EB was also mapped to examine its distribution across different regions. Next, Pearson correlation and global spatial autocorrelation analysis (Moran's *I*) were conducted to test the significance of independent variables and spatial effects in EB. Meanwhile, statistical tests, i.e., multicollinearity, normality, and heteroskedasticity tests, were performed to show the need to consider spatial effects in regression analysis. Lastly, both non-spatial and spatial regression analyses were conducted to analyze the impact of predictors on EB on the national, regional, and state levels and the importance of considering spatial effects. The following sections provide a summary of each analytic approach.

Four different regression models, including the non-spatial regression model, i.e., the ordinary least squares (OLS) regression model, and three spatial regression models, i.e., the spatial lag model [39], geographically weighted regression (GWR) model [40], and multi-scale geographically weighted regression (MGWR) model [41], were used to examine the relationships between EB and the predictors.

Table 3
Data sources for this study with links.

Item	Source
(1) COVID-19 cases and morality rate in 2020	Coronavirus Resource Center (https://coronavirus.jhu.edu/)
(2) CCVI index	Precision For COVID (https://precisionforcovid.org/ccvi)
(3) 2014–2018 5-year EB for LMI	Low-Income Energy Affordability Data (LEAD) Tool https://www.energy.gov/eere/slsc/maps/lead-tool
(4) 2014–2018 LMI household electricity, fuel, and natural gas expenditures ($/month)	Low-Income Energy Affordability Data – LEAD Tool – 2018 Update (https://data.openei.org/submissions/573)
(5) 2014–2018 and 2020 sales of electricity to the residential sector (10^3 MWh/month)	https://www.eia.gov/electricity/data/state/
(6) 2014–2018 and 2020 sales of fuel to the residential sector (10^3 MWh/month)	https://www.eia.gov/dnav/pet/pet_cons_psup_dc_nus_mbbl_m.htm
(7) 2014–2018 and 2020 sales of natural gas to the residential sector (10^3 MWh/month)	https://www.eia.gov/dnav/ng/ng_sum_lsum_dcu_nus_m.htm

The spatial-lag model was used to consider spatially autocorrelated residuals by incorporating a spatially-lagged variable to account for spatial autocorrelation in the model. However, it does not explicitly include spatial non-stationary in the model. GWR and MGWR models allow the estimated coefficients to vary spatially, i.e., create a unique regression equation for each observation in the studied dataset. The GWR mainly considers the non-stationary spatial effects (spatial heterogeneity), which neglects stationary spatial effects (spatial homogeneity). The MGWR was developed based on the GWR by considering the local impact of each predictor, and it is suitable to examine both spatial homogeneity and heterogeneity. Unlike GWR and MGWR, the OLS and spatial-lag models estimate a single, global regression model with constant coefficients for the considered determinants, i.e., the estimated coefficients are the same for all the observations. In this study, EB was log-transformed in all the regression analyses to ensure the data were normally or symmetrically distributed. Additionally, all the independent variables were standardized to make the results comparable in each studied region.

2.2.1. Spatial-lag model

A spatial-lag model was a spatial regression method to account for spatial autocorrelation of the EB, and it added a spatially lagged EB to the multiple linear regression model. It can be expressed as [42]:

$$y_i = \rho \sum_{k=1,k\neq i}^{n} w_{ik} y_k + \sum_{j=0}^{l} b_j x_{ij} + \varepsilon_i \tag{2}$$

Where ρ was a spatial autoregressive coefficient; w_{ik} was the spatial weights of the i^{th} observation due to the remaining k^{th} observation, which continuity-based weights can establish (e.g., Queens method) or distance-based weights (e.g., distance band weights). $\sum_{k=1,k\neq i}^{n} w_{ik} y_k$ was the spatial-lagged term that can be used to identify potential spatial effects. This study used the Queens method to generate the weights matrix for spatial-lag regressions.

2.2.2. GWR model

GWR model was formulated by assuming the estimated coefficients were a function of the location (u_i, v_i) (usually at the centroid coordinates of a block) in an OLS model, as shown in Eq. (3) [40]. In such a way, the estimated regression coefficients were unique to each location.

$$y_i = \sum_{j=0}^{l} b_j(u_i, v_i) x_{ij} + \varepsilon_i \tag{3}$$

The estimate $b_j(u_i, v_i)$ involved the calculation of a weights matrix by using a distance-based scheme. A kernel function was applied to the distance between observations and calibration points, where higher weights were assigned to closer observations than those farther away. Three kernel functions, i.e., Gaussian, exponential, and bi-square, were usually used to estimate the weights matrix. The adaptive bi-square kernel function was adopted because Gaussian and exponential kernel functions led to non-zero weight for observations even far away. In addition, the adaptive bi-square kernel function ensured each group of regression coefficients was estimated using the same bandwidth, i.e., the same number of nearest neighbors for each observation. The number of nearest neighbors, or bandwidth, was determined by the optimal corrected Akaike Information Criterion (AICc). The GWR model used a constant bandwidth to consider the non-stational effects of independent variables (i.e., predictors).

2.2.3. MGWR model

The MGWR model extended the GWR model by considering different bandwidths and predictors, as shown in Eq. (4). It created an optimal bandwidth for each independent variable, allowing some to operate on a global scale while others to operate on a local scale [41]. Therefore, the estimated regression coefficient, $b_{bwj}(u_i, v_i)$, was a function of the bandwidth of each independent variable. Similar to the GWR model, the adaptive bi-square kernel function was used to estimate the weight matrix, and the bandwidth was determined by optimal corrected AICc.

$$y_i = \sum_{j=0}^{l} b_{bwj}(u_i, v_i)x_{ij} + \varepsilon_i \tag{4}$$

3. Results

3.1. Distribution of energy burden

The distribution of EB in 2020 varied spatially. Fig. 1 (a) shows the spatial distribution of the annual average EB for 2020. Hot spots (dark green indicated severe EB) and cold spots (light green indicated low to high EB) scattered over the U.S. and its nine census regions, indicating spatial homogeneity and heterogeneity. Generally, the census regions in the south part had a higher EB than their counterparts in the north. For example, the southeast region experienced the highest average EB of 11.48 % (the East South-Central region).

In comparison, the east-north region had a relatively lower EB of 9.82 % (the East North-Central region). This result indicated spatially uneven distribution, i.e., spatial heterogeneity, of EB across different census regions. Such an uneven distribution of EB was more noticeable when examining the spatial-lagged EB, see Fig. 1 (b). The spatial-lagged EB was spatially weighted and used to find potential spatial clusters (cold and hot spots) in the data.

The spatially uneven distribution of EB was significant at the state level. Fig. 2 presents each state's annual average EB for 2020. The five states having the highest EB were Mississippi (MS) at 13.48 %, Maine (ME) at 13.02 %, Alabama (AL) at 12.97 %, Montana (MT) at 12.69 %, and Georgia (GA) at 12.63 %. On the other hand, the five states having the lowest EB included New Jersey (NJ) at 6.95 %, Washington (WA) at 7.05 %, Utah (UT) at 7.23 %, Hawaii (HI) at 7.35 %, and Wyoming (WY) at 7.71 %, indicate a 6 % EB reduction compared with the five highest states. The estimated EB was consistent with a recent report from the U. S. Department of Energy [43] and another recent study [11]. The spatially uneven distribution of EB was even more drastic at the county

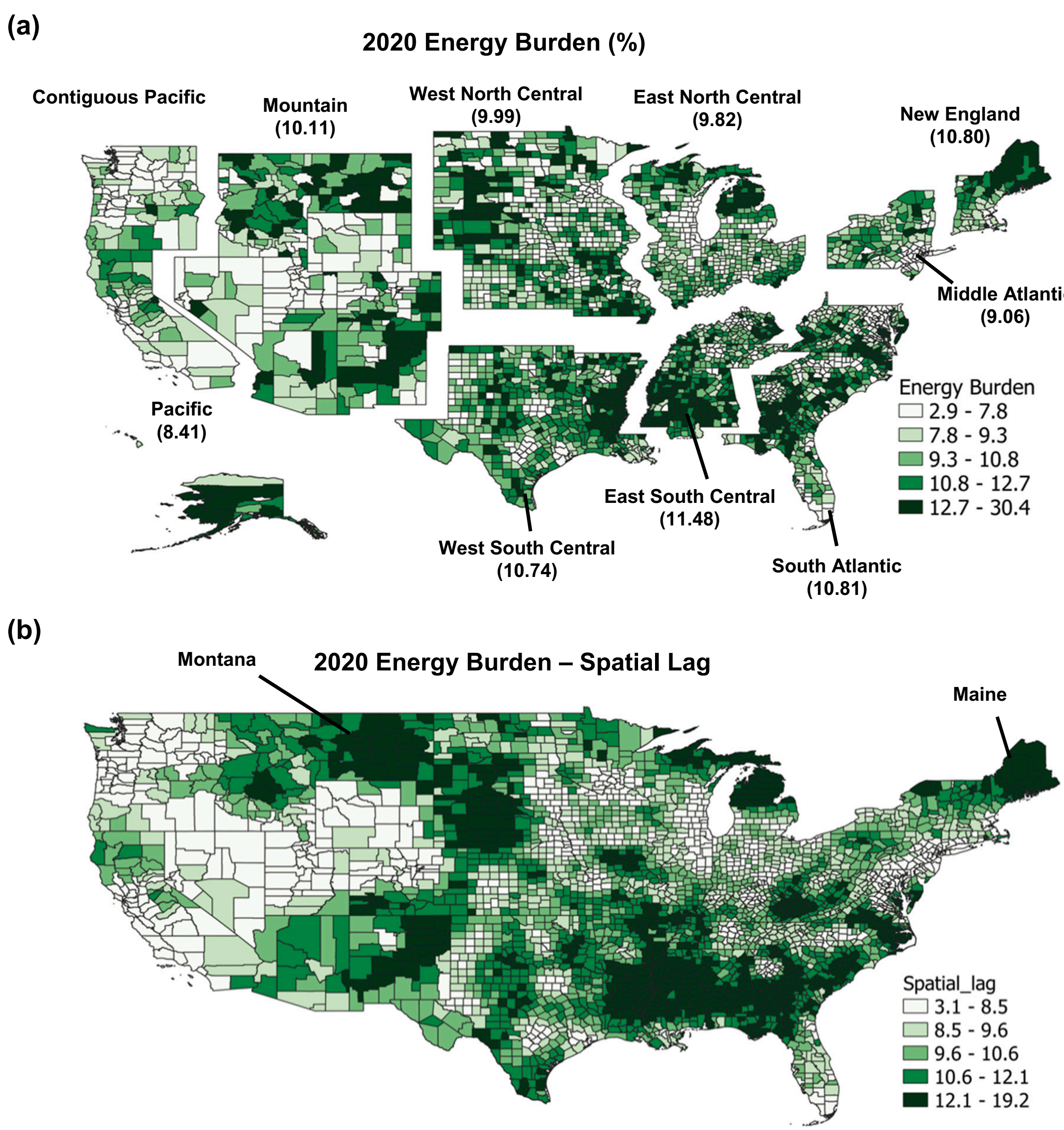

Fig. 1. Spatial distribution of (a) energy burden; and (b) spatially lagged energy burden.

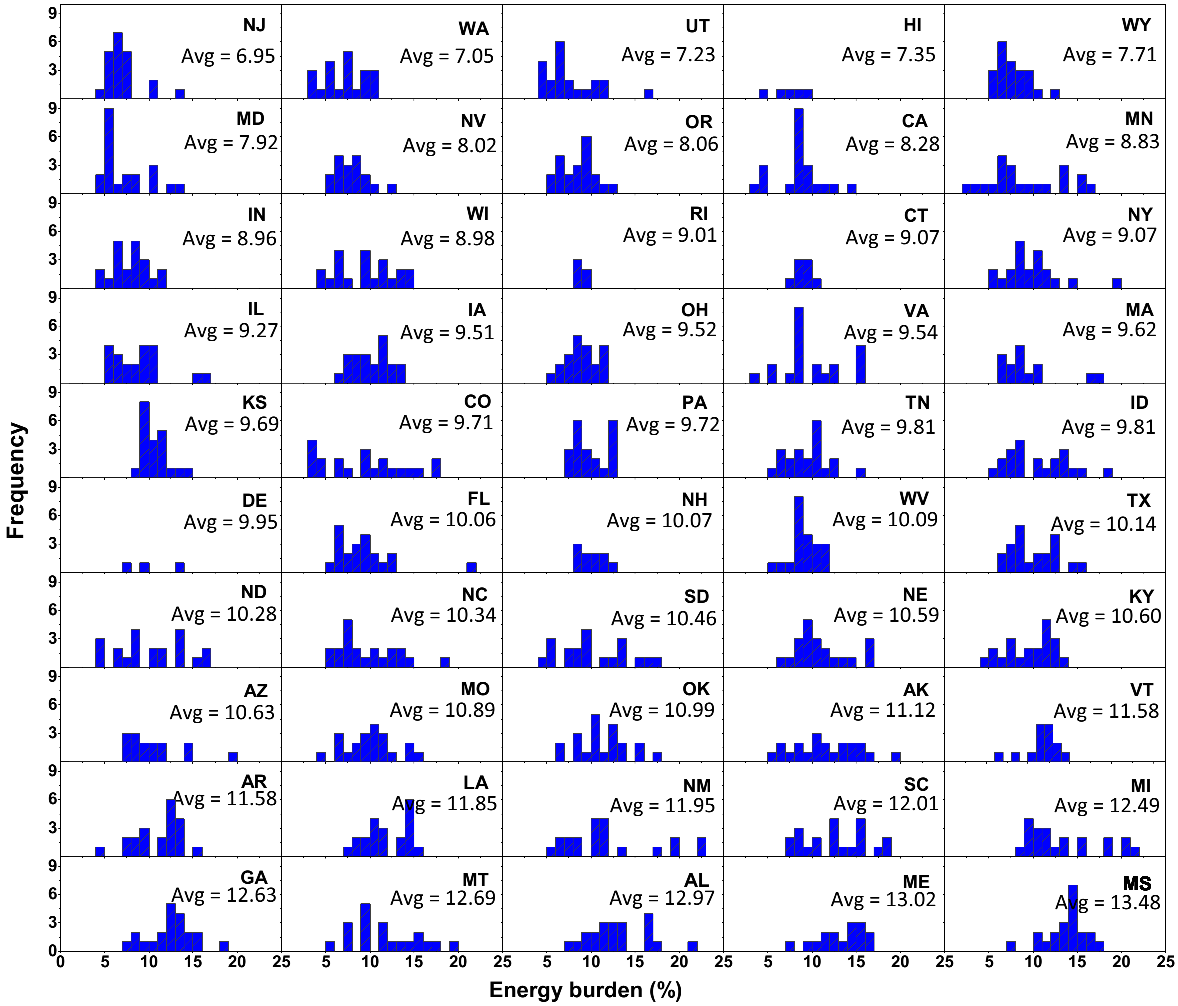

Fig. 2. Annual average energy burden counties scale in each state (The histogram of 50 states, except for Washington, DC, was shown in the figure. The abbreviation of each state was used to identify it; e.g., NJ represents New Jersey. The states follow the order from the minimum yearly average energy burden to the maximum yearly average energy burden. In addition, each state's annual average energy burden was added to it.)

level. For example, in New Jersey (NJ), the highest EB was 13.49 %, and the lowest was 4.18 %.

On the other hand, in Alabama, the highest EB was 22.49 %, and the lowest was 5.89 %. This result represented the diverse EB pattern with a significant level of variety in the U.S. Meanwhile, the diverse EB at the state level did not follow any specific statistical distribution, and each state is unique. However, one interesting pattern observed from the distribution was that a few counties in most states had a relatively higher EB than others. Mostly, these counties showed a relatively low income and high energy expenditure compared with their neighbor counterparts. For example, the LMI household in Decatur County, Tennessee, has a monthly income of about \$1350 with a monthly utility expenditure of approximately \$215, resulting in an EB of roughly 16 %. On the other hand, the neighboring Henderson County has a much lower EB, about 10 %, thanks to the relatively higher monthly income of about 1500 and lower monthly utility expenditure of about \$150.

Table 4 summarizs the mean values of the EB and independent variables. Green and orange shaded cells represented their annual average EB, smaller and more prominent than the national average values. The SE and HD factors followed the EB trend well in most of the regions, except for New England, Maine, and Montana. However, the relationships between EB and other independent variables were region-sensitive, indicating spatial heterogeneity. For example, the Pacific region had a relatively low EB but high levels of ML and HT.

3.2. *Global spatial autocorrelation, multicollinearity, normality, and heteroskedasticity*

Table 5 lists the results of the global spatial autocorrelation, Moran's *I* index. The obtained Moran's *I* index was significant (p-value 0.001) except for the states of Maine and Montana, which indicates the global spatial effects were not negligible. This result was not surprising because the spatially lagged EB (see Fig. 1 (b)) showed the existence of spatial clusters (hot spots and cold spots). The spatially non-significant of Maine and Montana (at the state level) was because of the small sample number of counties, which were 15 and 55, respectively. However, they became hot spots and were spatially significant on the national and census region levels (see Fig. 1 (b)). This result highlighted the importance of conducting multiple-level regional and spatial analyses on EB.

The multicollinearity, normality, and heteroskedasticity were tested using PySAL [39], and the results were also included in Table 5. The multicollinearity number was lower than 10, indicating that there were no two highly correlated independent variables. The Jarque-Bera test was used to examine the normality of the distribution of the errors. It tested the combined effects of skewness and Kurtosis. The results showed that the non-normal distribution of the errors existed in regions on the national level (U.S. and U.S. mainland) and census region level (such as Middle Atlantic).

Additionally, the Koenker-Bassett test was conducted to test the heteroskedasticity, i.e., the variance of the error term, and the results

Table 4
Mean values of EB and independent variables across regions.

	EB (%)	SE	HD	ML	HT	EF	HS	COVID case rate (%)	COVID mortality rate (%)
U.S. (50 states)	10.30	0.50	0.50	0.50	0.50	0.50	0.50	6.88	0.12
Contiguous U.S. (49 states)	10.30	0.50	0.50	0.50	0.50	0.50	0.50	6.70	0.12
New England (6 states)	10.80	0.23	0.26	0.37	0.35	0.41	0.16	2.83	0.07
Middle Atlantic (3 states)	9.06	0.37	0.32	0.47	0.41	0.50	0.65	4.21	0.11
East North Central (5 states)	9.82	0.41	0.45	0.32	0.38	0.53	0.57	7.06	0.12
West North Central (7 states)	9.99	0.23	0.47	0.30	0.53	0.42	0.38	8.06	0.11
South Atlantic (9 states)	10.81	0.61	0.47	0.65	0.48	0.52	0.69	5.66	0.11
East South Central (4 states)	11.48	0.75	0.66	0.41	0.49	0.72	0.58	7.41	0.12
West South Central (4 states)	10.74	0.64	0.63	0.72	0.59	0.57	0.52	7.06	0.14
Mountain (8 states)	10.11	0.42	0.44	0.57	0.54	0.25	0.25	6.78	0.10
Pacific (5 states)	8.41	0.47	0.43	0.73	0.69	0.28	0.38	7.88	0.05
Contiguous Pacific (3 states)	7.86	0.49	0.45	0.71	0.65	0.28	0.45	3.97	0.05
Mississippi	13.48	0.84	0.72	0.63	0.59	0.70	0.40	7.78	0.20
Maine	13.02	0.37	0.46	0.18	0.49	0.46	0.18	1.43	0.02
Alabama	12.97	0.75	0.63	0.57	0.41	0.67	0.82	7.45	0.12
Montana	12.69	0.31	0.39	0.26	0.52	0.28	0.05	7.63	0.14
Georgia	12.63	0.73	0.55	0.68	0.56	0.51	0.78	6.59	0.15

*EB = energy burden, SE = socioeconomic status, HD = household composition and disability, ML = minority and language, HT = household type and transportation, EF = epidemiological factors, HS = health system factors. The green (orange) shaded cells present their mean values are lower (higher) than their corresponding national mean values.

Table 5
Results of global Moran's *I*, multicollinearity, normality, and heteroskedasticity across regions.

Regions	Moran's *I*		Multicollinearity number	Normality (Jarque-Bera)		Heteroskedasticity (Koenker-Bassett)	
	Value	Sig		Value	Sig	Value	Sig
The U.S.	–		2.9	202.6	***	91.7	***
The contiguous U.S.	0.51	***	2.9	225.4	***	103.2	***
New England	0.54	***	6.0	1.2		3.9	
Middle Atlantic	0.58	***	4.5	75.2	***	8.9	
East North Central	0.53	***	3.4	6.8	*	33.9	***
West North Central	0.40	***	2.5	5.3		29.0	***
South Atlantic	0.59	***	3.9	12.0	**	17.9	*
East South Central	0.54	***	3.0	2.4		6.0	
West South Central	0.40	***	2.4	80.8	**	5.9	
Mountain	0.43	***	3.3	8.1	*	29.5	**
Pacific*	–		6.1	7.1	*	10.7	
Contiguous Pacific	0.44	***	3.8	6.5	*	6.6	
Mississippi	0.21	***	3.0	1.8		3.9	
Maine	0.14		7.1	1.3		7.9	
Alabama	0.29	***	3.3	0.5		6.2	
Montana	−0.04		3.4	1.2		5.4	
Georgia	0.57	***	3.8	4.9		6.8	

-Not available due to Alaska and Hawaii being disconnected from the mainland.
* ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$).

showed that the variance of the error terms was not a constant on the national and census region levels. On the other hand, the non-normal distribution of errors and the variance of error terms highlighted the importance of considering spatial effects in the regression analysis on the national and census region levels. Finally, Pearson's correlations found that the relationship between EB and the independent variables was significant, as shown in Table 6.

Table 6
Pearson's correlation coefficient.

CDC themes			Covid-19		
Variable	Pearson's correlation coeff		Variable	Pearson's correlation coeff	
SE	0.50	***	COVID-19 case	0.04	*
HD	0.49	***	COVID-19 death	0.21	***
ML	−0.16	***			
HT	0.31	***			
EF	0.21	***			
HS	0.17	***			

***($P < 0.001$), **($P < 0.01$), *($P < 0.05$).

3.3. Regression results of independent variables on EB

3.3.1. Goodness of fit

To analyze the relationship between independent variables and EB, both non-spatial and spatial regression analyses were performed, whereas OLS and spatial lag regression analyses were conducted in PySAL [44]. In addition, GWR and MGWR analyses were conducted in MGWR2.2 [45]. The regression results, summarized by the goodness of fit metrics (adjusted R^2 (adj. R^2) and corrected Akaike Information Criteria (AICc)), are shown in Table 7. Relatively high adj. R^2 was obtained for all the studied regions, and a notable improvement in model fit, i.e., higher adj. R^2 and lower AICc values were observed from non-spatial (OLS) to spatial regression analysis (spatial lag regression, GWR, and MGWR). This result highlighted the importance of spatial effects in predicting EB. Furthermore, the regional-level regression analysis usually had a higher adj. R^2 value than its counterpart at the national level because EB was more uniform in a small region.

MGWR has a relatively higher adj. R^2 value compared to spatial lag regression and GWR on the national and census region levels. This result was because MGWR considers spatial homogeneities, i.e., the similarity of EB at a location and its surrounding average, and spatial heterogeneity, i.e., spatial non-stationary. In addition, the local adj. The R^2 values of GWR and MGWR at the national and census region levels were compared in Fig. 3. They shared a similar pattern: the East South-Central and West North-Central regions had the highest and lowest adj. R^2 values, respectively. This difference may be explained by analyzing the EB distribution shown in Fig. 1. The East South-Central region showed significant spatial homogeneity (the whole area as a hot spot with minor spatial variations). In contrast, the West North-Central region showed a checkboard pattern where the low EB counties were scattered among high EB counties, indicating a high spatial variation. As a result, the spatial lag model had a relatively higher adj. R^2 value because a small region had more probability of showing spatial homogeneity.

3.3.2. Impacts of independent variables on energy burden

This section describes the impacts of our independent variables on EB. Table 8 summarizes the regression coefficients for the Contiguous U. S. and Appendix I (Tables I-1–I-16) for the remaining regions. At the national level, the OLS results showed that the coefficients of all eight independent variables are significant (*p-value* < 0.05) except for the COVID case rate (see Table II-1). All the estimated coefficients except for minority and language (ML) and COVID case rates were positive, indicating that higher values were associated with higher EB. SE has the most considerable impact on EB, followed by household composition and disability (HD), household type and transportation (HT), and COVID mortality rate. At the regional scale, we observed similar OLS results for most regions, with some exceptions. For example, ML has a positive coefficient in the East South-Central region. In the Middle Atlantic region, EF, HS, and COVID cases and mortality rates had a negative coefficient but were not statistically significant. This result highlighted the regional differences when using the CCVIs and COVID-19-related variables to predict EB. At the state scale, similar variations were observed as at the regional scale. The regression coefficients of the spatial lag model share a similar trend with the OLS model. All the coefficients of spatial weights are positive and significant except Maine (negative). For Maine, it had a high and uneven (like checkboard) EB distribution (see Fig. 1 (a) and Fig. 2), which leads to the non-significant spatial autocorrelation (see Table 5).

To examine how the spatial regression coefficients vary, we further explored the local regression coefficients of MGWR. The local MGWR regression coefficients maps vary in magnitude and significance across space; see Fig. 4 for socioeconomic status (SE) and Figs. II-1–II-8 for the intercept and other independent variables. The local coefficient maps for SE, HD, ML, epidemiological factors (EF), and COVID mortality rates were significant predictors of EB on the national level. On the regional level, the local coefficient for SE (except New England region) and ML (except for a small area in New England region, Middle Atlantic, the East South-Central, and the West South-Central) were significant across all regions. In addition, local regression coefficients of HD, HT, EF, healthcare system factors (HS), and COVID case and mortality rates varied by region. It is noted that EB was first log-transformed and then normalized along with other independent variables in each studied area. Therefore, the regression intercept presents the mean value of the log-transformed EB when setting the independent variables to their means before normalization. On the national level, the negative intercept coefficients of the Pacific, East North Central, and Middle Atlantic regions demonstrated low mean EB values in these regions.

To interpret the regression results, our positive regression coefficients indicate the increase of independent variables leading to the rise in EB. The negative regression coefficients mean the increase of the independent variables leading to decreased EB. It is expected that the rise in CCVI score will increase EB since an area with a higher CCVI score means the community is more vulnerable. For example, a higher SE coefficient in the CCVI (i.e., lower social and economic status) will increase EB and vice versa. On the other hand, the COVID case rate has a negative coefficient, indicating the areas with higher COVID-19 case rates had a lower EB, which does not mean COVID-19 cases reduced EB. For example, New York City in 2020 had a significantly higher rate of COVID-19 cases. Still, the city had a lower EB than other areas, which shows that many higher social-economic areas had a high rate of COVID cases in 2020. COVID mortality rate had a positive coefficient, meaning the areas with a higher COVID-19 death rate (generally have poor social-economic situations and healthcare systems) lead to higher EBs. COVID-19 case and mortality rates were included in the regression analysis of EB mainly because (1) they are not the same as CCVI and directly count the effects of COVID-19; (2) they are considered critical controlled variables for researchers to study similar extreme events on EB. However, COVID-19 related variables are not the central focus of this study but were controlled because they could influence the outcomes of EB due to staying-at-home orders or people who lost jobs.

Table 7
Summary of model fit results[a]

Region	Metric	OLS	Spatial lag	GWR	MGWR
The U.S.	dj. R^2	0.370	–	–	–
	ICc	7476.1	–	–	–
The contiguous U.S.	Adj. R^2	0.379	0.536	0.621	0.653
	AICc	7350.2	–	6224.0	5880.6
New England	Adj. R^2	0.618	0.685	0.689	0.741
	AICc	134.0	–	126.8	116.0
Middle Atlantic	Adj. R^2	0.559	0.618	0.587	0.666
	AICc	311.7	–	309.3	287.1
East North Central	Adj. R^2	0.538	0.646	0.647	0.700
	AICc	911.7	–	816.0	756.8
West North Central	Adj. R^2	0.274	0.445	0.424	0.513
	AICc	1564.9	–	1462.3	1371.7
South Atlantic	Adj. R^2	0.547	0.641	0.708	0.762
	AICc	1211.5	–	1051.6	926.5
East South Central	Adj. R^2	0.588	0.699	0.722	0.747
	AICc	719.3	–	615.2	581.8
West South Central	Adj. R^2	0.411	0.502	0.535	0.628
	AICc	1093.7	–	1031.5	943.7
Mountain	Adj. R^2	0.258	0.508	0.527	0.597
	AICc	722.6	–	628.8	588.3
Pacific	Adj. R^2	0.358	–	–	–
	AICc	408.6	–	–	–
Contiguous Pacific	Adj. R^2	0.468	0.622	0.639	0.715
	AICc	302.3	–	265.1	234.8
Mississippi	Adj. R^2	0.581	0.634	0.629	0.581
	AICc	169.9	–	167.2	170.9
Maine	Adj. R^2	0.819	0.952	–	–
	AICc	22.8	–	–	–
Alabama	Adj. R^2	0.611	0.668	0.650	0.716
	AICc	135.2	–	135.1	123.9
Montana	Adj. R^2	0.224	0.307	–	–
	AICc	152.9	–	–	–
Georgia	Adj. R^2	0.490	0.705	0.626	0.689
	AICc	352.9	–	323.2	353.9

[a] Adj. R^2 = Adjusted R^2 and AICc = corrected Akaike Information Criterion.

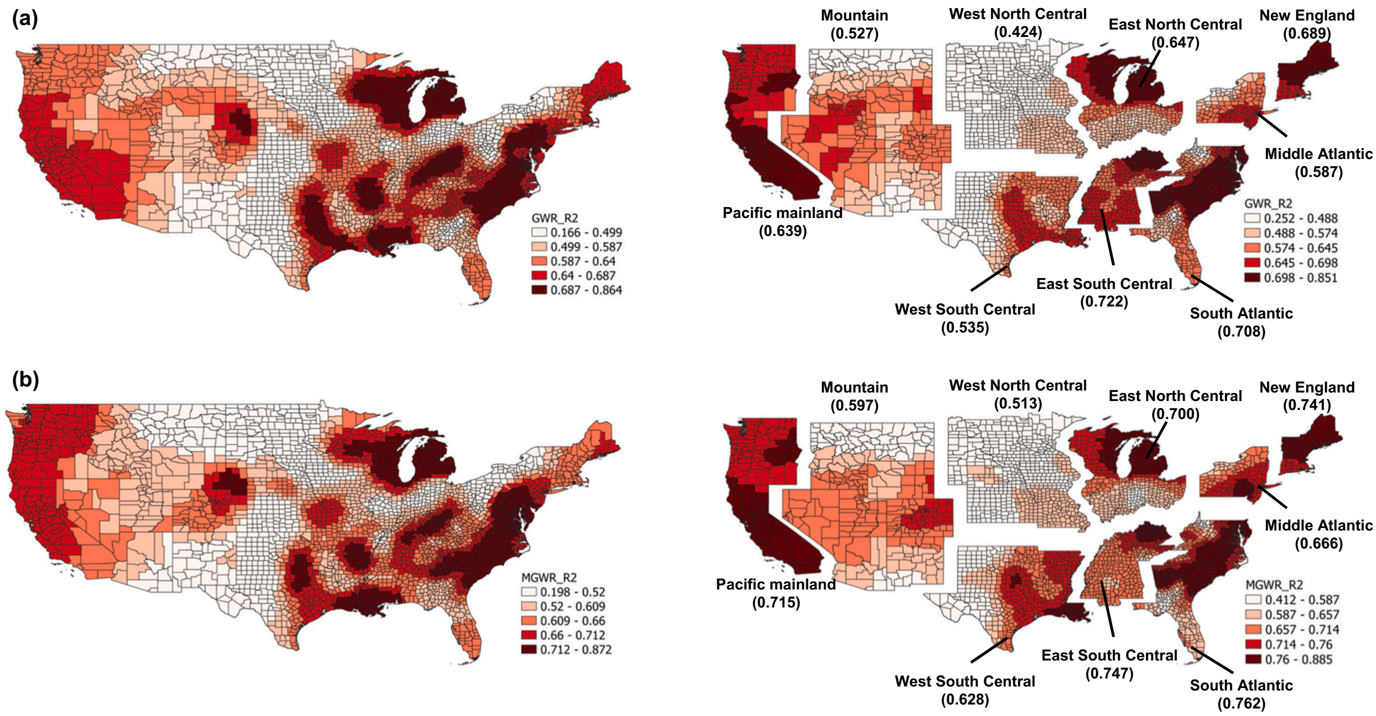

Fig. 3. Local adjusted R^2 values on national and census region levels of (a) GWR model and (b) MGWR model.

Table 8
Model coefficient and significance summary of the Contiguous U.S.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		−0.001		0.045	−1.014	1.680	49	−0.022	−1.452	1.283	61
SE	0.345	***	0.237	***	0.436	−0.146	0.923	83	0.417	0.273	0.505	100
HD	0.178	***	0.146	***	0.139	−0.326	0.681	40	0.112	0.038	0.246	94
ML	−0.256	***	−0.195	***	−0.339	−1.029	0.243	71	−0.283	−0.424	−0.160	100
HT	0.119	***	0.090	***	0.099	−0.331	0.482	34	0.078	−0.507	0.671	28
EF	0.030	*	0.015		0.041	−0.132	0.385	15	0.034	0.029	0.038	100
HS	0.034	*	0.041	**	0.084	−0.532	1.123	35	0.106	−0.154	0.507	52
COVID case rate	−0.100	***	−0.086	***	−0.088	−0.871	0.402	31	−0.114	−0.308	0.020	62
COVID mortality rate	0.153	***	0.120	***	0.070	−0.413	0.717	24	0.070	0.025	0.119	93
w_EB	–	–	0.075	***	–	–	–	–	–	–	–	–

* ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$).

3.3.3. *Results of spatially homogeneous and heterogeneous independent variables on EB*

To examine the influence ranges of the independent variables on EB, the bandwidth of GWR and MGWR were used to determine if they were spatially homogeneous or heterogeneous. Table 9 summarizs the bandwidths of GWR and MGWR. A spatially homogeneous independent variable has a bandwidth close to the entire counties (observations) across the studied region, identified by green shading in the table. On the other hand, a spatially heterogeneous independent variable has a bandwidth of less than the entire county. On the national level, only EF is spatially homogeneous. However, independent variables like SE, HD, ML, HT, HS, COVID case rate, and COVID mortality rate showed different levels of spatially heterogeneous.

At the regional level, the independent variables of all the studied census regions showed some level of spatial homogeneity and heterogeneity. More importantly, the spatially homogeneous and heterogeneous independent variables varied from region to region. First, the total number of homogeneous independent variables is different. The East South-Central and West South-Central regions (total of 3) have the least spatially homogeneous independent variables. In contrast, the West North-Central region has the most spatially homogeneous independent variables (total of 7). The remaining regions have a total of five spatially homogeneous independent variables. Usually, the more spatially homogeneous independent variables, the closer their regression result is to the OLS [30]. However, the 'Intercept,' which represents the effect of locations in MGWR, cannot be neglected in the comparison between spatial and non-spatial regression results; for example, the West North-Central region has seven spatially homogeneous independent variables, its adj. As a result, the R^2 value of OLS is much smaller than the spatial regression models. Second, the homogeneous independent variables are varied for the studied regions. For example, HT, EF, and HS were spatially homogeneous in the East-South-Central. At the same time, SE, ML, EF, COVID cases, and mortality rates were spatially homogeneous in the South Atlantic.

More spatially homogeneous independent variables are expected at the state level due to the reduced spatial scale (see Table 9). Similar to the regional scale results, the spatially homogeneous and heterogeneous independent variables vary from state to state. The differences in spatially homogeneous and heterogeneous independent variables on the regional and state level indicate that policies to relieve high EB of the LMI households must be adjusted across the U.S. This will be discussed in the policy implication section.

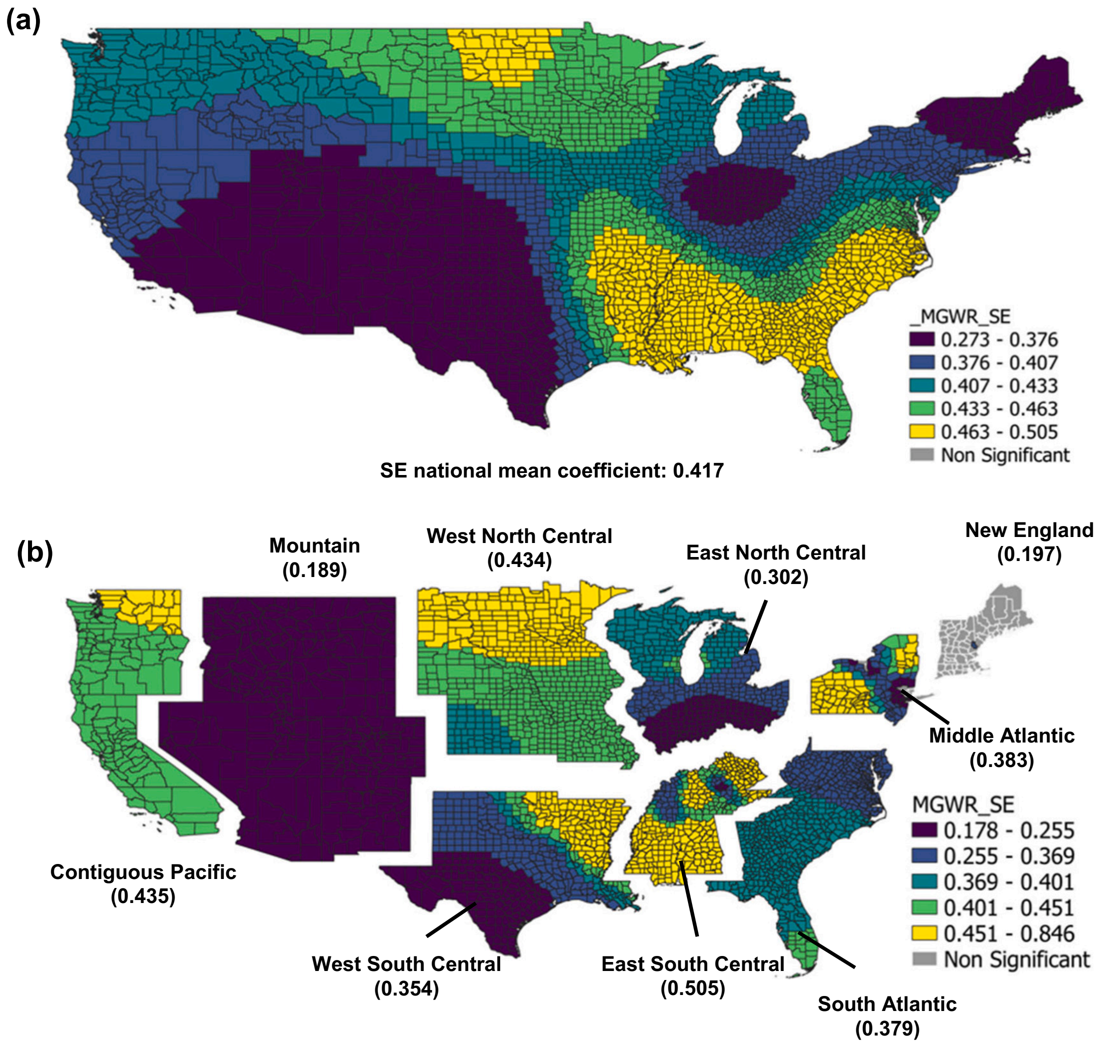

Fig. 4. Local regression coefficients of MGWR for SE at (a) national level; and (b) census region level.

4. Discussion

4.1. Summary of key findings

This study presents a multiple-level spatial analysis of U.S. LMI households' EB during the COVID-19 pandemic. In addition, we use non-spatial and spatial regression methods to better understand the effects of the CCVIs and COVID-19 cases and mortality rates on energy burdens. We have summarized and discussed the following key findings:

1. *The distribution of energy burden is diverse. Our* study suggests the diverse distribution of EB in the U.S. in 2020, ranging from a minimum of 2.93 % (Alexandria County, Virginia) to a maximum of 30.45 % (Quitman County, Georgia) across 3142 counties. Given the similar weather, and electricity prices, the difference in EB within one state across counties could be rather extreme; for example, in Alabama, the difference from the smallest to the largest EB was 16.60 %. The diverse EB was also presented in previous studies [12,30]. In the study of EB at the urban scale, Moore and Webb [30] found that the lowest and highest EBs in Cincinnati, Ohio, was 0.6 % and 19.7 %, respectively. On the other hand, Drehobl et al. [12] analyzed 25 urban areas in the U.S. and found that even the first 25th percentile LIHs have much higher EB than the first 50th percentile LIHs. For example, in San Antonio, Texas, the first 25th percentile LIHs have an EB of 21.7 % while the first 50th percentile LIHs is 7.4 %, where a difference of 14.3 % was observed.
2. *The multidimensionality of community vulnerability influences energy burden.* As mentioned earlier, the causes of EB are not mainly from sociodemographic factors. Our analyses show that EBs are influenced by the multidimensionality of the community vulnerability based on six indexes of CCVI with >40 indicators, including social vulnerability, housing type, transportation access, epidemiological factors, and health system strength. Similar to existing literature [11,29–31]. This study confirms the importance of socioeconomic and race/ethnicity factors contributing to the LMI households' EB. However, different from the existing EB studies, this study discovers the unique relationships between household compositions (e.g., dependent children <18 years of age, persons aged 65 years and older, and single-parent households) and disability, transportation, and housing types (e.g., lack of vehicle access and crowded housing), and EB. Additionally, our study demonstrates EB is positively linked to epidemiological factors (e.g., the prevalence of cardiovascular and respiratory conditions and the share of the population over age 65) and healthcare system factors (e.g., the number of intensive care unit beds per 100,000 people, health spending per capita, and the share of the population with a primary care physician). Therefore, individual factors such as socioeconomic or race/ethnicity are not the only variables influencing EB, and societal infrastructure factors, such as the health care system, are also closely linked to EB. Due to the close

Table 9
Bandwidths of GWR and MGWR.

Region	Model	Bandwidth									Total counties
		Inter cept	SE	HD	ML	HT	EF	HS	COVID case rate	COVID mortality rate	
The contiguous U.S.	GWR						150				3108
	MGWR	44	676	765	977	69	3106	221	527	1555	
New England	GWR						64				67
	MGWR	65	44	65	44	50	65	65	65	44	
Middle Atlantic	GWR						149				150
	MGWR	148	64	148	83	148	148	46	148	107	
East North Central	GWR						292				437
	MGWR	49	322	246	433	436	433	195	436	436	
West North Central	GWR						272				618
	MGWR	43	598	533	593	570	617	617	106	613	
South Atlantic	GWR						100				588
	MGWR	43	587	47	587	91	587	43	496	587	
East South Central	GWR						147				364
	MGWR	151	67	146	117	356	234	335	148	67	
West South Central	GWR						169				470
	MGWR	48	205	469	48	140	469	169	467	70	
Mountain	GWR						128				281
	MGWR	43	280	259	280	50	246	128	239	99	
Contiguous Pacific	GWR						84				133
	MGWR	43	132	123	95	132	132	55	65	121	
Mississippi	GWR						81				81
	MGWR	78	80	80	44	80	59	67	78	63	
Alabama	GWR						66				67
	MGWR	61	65	65	46	65	48	50	61	44	
Georgia	GWR						103				159
	MGWR	44	157	157	157	142	157	155	157	93	

*The results of Maine and Montana are unavailable because the counties in each are less than the minimum number of observations to run GWR and MGWR. However, the green-shaded cells indicate the corresponding independent variables are homogeneous in the studied regions.

relationship of health-related factors with EB in our findings, this study recommends researchers use the CCVI to examine the link between EB, epidemiological factors, healthcare system factors, and people with disabilities.

3. *The key factors influencing energy burden are distinct across regions*. Depending on the region, the effects of community vulnerability factors on EB vary. At the national level, the most significant drivers of EB are the variables relating to socioeconomic status (SE), household composition and disability (HD), minority and language (ML), household type and transportation (HT), and COVID-19 mortality rates. At the regional and state levels, however, the significant positive drivers are social economics. At the same time, minority and language are negative except in the East South-Central region across all the U.S. regions. Moreover, the considerable drivers are varied in different regions. For example, household composition and disabilities positively affect energy burdens in the West South-Central region (e.g., the states of Arkansas, Louisiana, Oklahoma, and Texas) while nearly having no impact on the Mountain and East South-Central regions (e.g., the states of Alabama, Kentucky, Mississippi, Tennessee). On the other hand, household types and transportation positively affect EB in the South Atlantic region (e.g., stretches from the states of Delaware to Florida) but almost no impact on the Middle Atlantic and the West North-Central regions (e.g., the states of New York, Pennsylvania, Minnesota). The analysis indicates that the key factors varied across the spatial scales, and it is beneficial to conduct multiple-spatial scale analyses to identify them across the studied regions. In addition, some of the critical factors that are significant at the national level may show a negligible effect at the regional level and vice versa. For example, the HD, HT, and COVID mortality rate variables are significant at the national level but not in the Middle Atlantic region. In addition, the COVID mortality rate is not significant in the East North-Central region; HT is not statistically significant in the West North-Central and East South-Central regions. This finding reflects the effects of spatial scale and regional differences, which need attentions because the national-level study may mask such regional differences. Therefore, the authors recommend analyzing the large spatial scale first and then going to a small one to pinpoint such differences. Meanwhile, it is noted that such multiple spatial-scale analysis depends on the observations (or the smallest scale) in each region. This study used county-level EB as the smallest scale, and a smaller scale will be needed, such as census tract level EB, if the city-level spatial analysis is desired.
4. *The key factors influencing energy burden show spatial homogeneity and heterogenicity across regions*. On the national level, only the impact of epidemiological factors on energy burden is homogeneous, and all other CCVI factors (e.g., SE, HD, ML, HT) are heterogeneous. However, all the CCVI factors at the regional level show spatially homogeneous and heterogeneous levels but are also distinct by region. For example, the influence of HT, EF, and HS on EB is homogeneous in the East South-Central region (e.g., Alabama, Tennessee). In contrast, the effects of SE, HD, ML, HT, EF, HS, and COVID mortality rates on energy burdens are homogeneous in similar patterns in the West North-Central region (e.g., Iowa, Kansas). The variations of spatially homogeneous independent variables in the different areas are attributed to other climate conditions, energy prices, population distribution, and energy-related policies. Similar spatial homogeneity and heterogeneity were conducted by previous studies [29,30]. For example, Moore and Webb [30] found that the economic factors, i.e., medium household income and poverty, are spatially heterogeneous. In contrast, the social and building physics factors, i.e., non-white, gas heat, and two-family, are spatially homogeneous in Cincinnati, Ohio (city-level). On the other hand, Mashhoodi et al. [29] found that the economic factor, i.e., low income, is spatially homogeneous. In contrast, the building physics factors, i.e., building age, the number of summer days, and the number of frost days, are spatially heterogeneous in the Netherlands (national level). The results from the literature also highlight the importance of considering spatial scales and location differences, where the spatially homogeneous variable at the national level is different from the regional level or city level.

4.2. Limitations and future study

There are several limitations to the present study that could inspire future research. First, the mixed measurement of minority status and language fluency in the CCVI index requires attentions. The quantitative relationship between EB and the ML index varies regionally and presents this study's multiple regional and spatial effects. However, the mixed measurement of these two variables requires further explored. Our study found a negative correlation between minority and language and EB, which contradicts previous research focusing on only minority status. The negative coefficient in our study is mainly because the ML index in the CCVI data is the weighted average scores of minority status and English fluency. The combined measurement of race/ethnicity and language skill into one ML index might bias the analysis of EB. Future research should separate these two variables or use race/ethnicity as the independent variable, similar to some existing studies. For example, Drehobl et al. [12], Chen et al. [11], and others [25,30,46] found that underrepresented minorities (non-white) such as African Americans and Hispanics have much higher EBs than other race/ethnicity. Second, this study only focuses on the effects of household energy expenditure on electricity, fuel, and natural gas; however, future research can investigate other household expenditure burdens connected with EB, such as expenses for rent, medicine, food, transportation, and since LIHs may be forced trade-offs between energy expenditure and other necessities. Third, although this study does not consider particular energy policies and building energy efficiency features, future research should consider adding these variables with national representative data. Finally, this study only investigates the combined effects of the major themes from the CCVI index. Future research should explore the influence of each variable of the CCVI to understand better the independent impact of key variables contributing to EB. For example, it might be helpful for researchers to analyze health-related factors carefully, e.g., how the prevalence of cardiovascular and respiratory conditions and the share of the population over age 65 or the health care system influence EB. In addition, it might be helpful to explore whether poverty, minority status, English skills, lack of transportation, housing type, disability status, or seminar over 65 or connected factors on EBs are more likely to contribute to EB.

On the other hand, the CCVI is critical in understanding how vulnerable communities at census tract, county, state, or regional level suffer from EB and identifying whether energy costs burden a vulnerable community due to people do not have equitable access to health care, transportation, affordable housing, or secure employment, etc. Therefore, CCVI is strongly recommended in EB analysis. Finally, this study did not analyze the role of utilities; future researchers should analyze the impacts of utility policy (e.g., the utility disconnection act) on the EB of LIHs.

5. Policy implications

Our study highlights the importance of considering the multidimensionality of community vulnerability factors and the consideration of spatial homogeneity and heterogeneity in analyzing EB. We propose four critical recommendations to enhance energy burdens in low-income communities:

1. *Improve the accessibility of low-income energy efficiency programs and weatherization.* Our research suggests the highest EB is in the southeastern region, which provides some policy recommendations for the regional level. For example, the low-income home energy assistance program (LIHEAP) and the weatherization assistance program (WAP) at the national level aim to improve residents' EBs through bill payment assistance and energy efficiency measures needed to enhance public awareness regarding program content and actions that customers can take. Experts have also raised the issue of program efficacy [47,48]. The point of distributional equity that relies on a standard federal poverty line and economy food plan to determine the edibility of energy assistance programs requires revaluation by including more low-income residents experiencing energy poverty. Additionally, local government can invest more money to improve current weatherization and energy efficiency retrofit programs funded through utility or federal programs by collaborating with community partners such as city and regional authorities or non-profit organizations to implement existing or create new pilot energy efficiency programs. For example, regional energy-efficiency networks can help to promote energy assistance programs to make them more accessible by leveraging each other's efforts. These networks could include Southeast Energy Efficiency Alliance (SEEA), Southwest Energy Efficiency Project (SWEEP), and South Alliance for Clean Energy (SACE) with partnerships with the U.S. DOE, as well as utilities, third-party program administrators, public officials, advocacy groups, businesses, and foundations can reach to LIHs with a high EB [49]. In addition, the regional energy networks can provide technical assistance to the states in the U.S. southeast and municipalities to support regional efficiency policy development and implementation.
2. *Community co-designed communication and outreach strategies.* Our findings also suggest socioeconomic factors, or race/ethnicity at the county level, have impacted EB; therefore, this study indicates including trustworthy representatives from the local or ethnic communities to combat EB effectively. Importantly, it is desired that these representatives understand local and non-English speaking cultures and their associated EB-related issues and thus integrate into the co-designing utilities' energy efficiency or bill assistant program outreach. Additionally, the design of the program application process and benefits of bill assistance and weatherization programs require utilities to include outreach specialists from the community to perform face-to-face outreach. Therefore, a localized best practice of designing, implementing, and delivering effective energy efficiency programs and weatherization by having community members is much needed.
3. *Set energy burden goals, leverage programs, and evaluate progress.* Our findings suggest a wide range of EB in the U.S. depending on the level of analysis at the city, county, regional, or state levels. Therefore, an average EB at a country or state level might not present the risk of higher EB for s community. Local community utilities and state policymakers should set a realistic goal of improving energy affordability and reducing EBs as their priority in addressing community energy insecurity issues (e.g., no >6 %) [50]. For example, the City of New York has one of the country's highest electricity rates [51]. To reduce its EB, the City of New York launched the State's first energy affordability policy in 2016 with a target of limiting energy costs for LIHs to no >6 % of their pre-tax household income [52]. The city has identified a multi-pronged intervention and initiatives to alleviate the EB to achieve this goal, such as focusing on the policies to impact energy rates and energy consumed, promoting energy efficiency investments, increasing access to low-cost renewable energy, developing policy options to incentivize community solar or low-income solar development, and taking bold steps toward improving the sustainability and resiliency of its energy supply [51]. The city also creates concrete strategies to track progress toward this goal. While leveraging existing programs to improve energy efficacy and affordability for LIHs is essential, periodically evaluating the program design and effectiveness of participation is critical, especially for the harder-to-reach, low-income communities of color.
4. *Develop an interactive tool for measuring energy, housing, and health intersections.* The majority of EB studies in the U.S. focuses on the relationship between socioeconomic and race/ethnicity factors and EB; however, this study has demonstrated the intersectional drivers of community vulnerabilities, including six composite indices of socioeconomics, race/ethnicity, housing condition, and types, disability, transportation accessibility, epidemiological, health care

system, and COVID-19 factors. Our findings contribute to energy justice literature by building a comprehensive quantitative framework from several merged national representative datasets, similar to the U.S. Low-income Energy Affordability Data (LEAD) tool. As a result, policymakers can develop a similar measurable tool for analyzing counties and state and regional EBs by tailoring their own need or expanding the interconnected factors such as sociodemographics, ethnicity, language, transportation, housing conditions, and types, electricity rate, health care system, individual health conditions, heating or cooling degree days, and energy efficiency programs, low-income assistance policies, and so on. Our data could be built into an interactive mapping tool to help the country, region, states, and other stakeholders create better energy strategies by improving their understanding of underserved communities' housing, energy characteristics, and health system. For example, our results demonstrate unique relationships between EBs, poor housing conditions, transportation accessibility, and people with disabilities. Policymakers can develop specific EB assessments and energy efficiency programs for people with disabilities. Further, the local energy policy can incentivize landlords to improve the energy-efficient housing environment for people with disabilities or low-income renters. Energy-efficient tools or policies also need to target the geographic patterns which cause energy poverty rather than offering financial aid to mitigate the overall effects of energy poverty [29].

This study's findings suggest that the influence of the determinants of EB could be spatially homogeneous (i.e., national-level determinants) or heterogeneous (i.e., neighborhood-specific determinants). Therebefore, energy policies needed to accommodate this spatial or location effect by diversifying their target residents, especially those in underserved communities. For example, utility energy efficiency programs and the LIHEAP should consider the policy for no shut-off orders for people with electricity-dependent medical devices or health conditions that are affected by heating and cooling. The WAP should also consider these vulnerable groups as their priority in providing the program assistance. Finally, our findings connecting with health systems and indicators recommend that weatherization program evaluators and outreach personnel could include the benefits of environmental health, housing quality, and energy security in cost-benefit analyzes or outreach materials. The impacts of energy on health are well-documented [53]. Therefore, policymakers should develop health-related policies to improve EB and a healthy built environment for LIHs.

6. Conclusion

This study contributes to energy justice literature and provides a multi-scale and multi-dimensional study of EB. The analysis of the CCVI's themes here (e.g., socioeconomic, minority status and language, housing condition, transportation, epidemiological, health care system, and COVID mortality and cases) shows that different communities across the country are vulnerable to energy burden for various reasons. In supporting vulnerable populations, it is critical to understand what explicitly drives their vulnerability to energy burdens. The connection between vulnerability and energy poverty is complex, and energy poverty does not have the same pattern in every community and exact causes. This study can help researchers and policymakers understand *how* and *where* energy poverty impacts vulnerable populations to prioritize and monitor energy assistance resources. Where available, data on the costs of electricity, transportation, health care, and rent/house value can be combined with the CCVI to quickly identify the interconnected issues of energy, housing, and health during normal and extreme events.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgement

C.-F. Chen, was supported by the Engineering Research Center Program of the U.S. National Science Foundation (NSF) and the Department of Energy under NSF award EEC-1041877 and the CURENT Industry Partnership Program. C.-F. Chen also thanks the support of the U.S. Fulbright Global Scholarship Award 2019-2020. The authors thank Hannah Nelson for researching literature review. This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Appendix I. Model coefficients and significance summary in all the regions except the Contiguous U.S.

Here we present the model coefficient and significance summary of all the studied regions except the Contiguous U.S. in Tables I-1–I-16.

Table I-1
Model coefficient and significance summary of the U.S.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		–	–	–	–	–	–	–	–	–	–
SE	0.350	***	–	–	–	–	–	–	–	–	–	–
HD	0.175	***	–	–	–	–	–	–	–	–	–	–
ML	−0.253	***	–	–	–	–	–	–	–	–	–	–
HT	0.120	***	–	–	–	–	–	–	–	–	–	–
EF	0.029	*	–	–	–	–	–	–	–	–	–	–
HS	0.036	*	–	–	–	–	–	–	–	–	–	–
COVID case rate	−0.011		–	–	–	–	–	–	–	–	–	–
COVID mortality rate	0.108	***	–	–	–	–	–	–	–	–	–	–
w_EB	–	–	–	–	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-2
Model coefficient and significance summary of the New England region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		0.020		−0.111	−0.253	−0.015	45	−0.279	−0.344	−0.212	100
SE	0.219		0.203		0.261	0.171	0.417	34	0.197	−0.036	0.339	1
HD	0.244		0.204		0.237	0.130	0.378	25	0.226	0.119	0.291	36
ML	−0.329		−0.310		−0.454	−0.778	−0.014	60	−0.490	−0.739	−0.210	69
HT	−0.095		−0.120		−0.078	−0.243	0.094	25	−0.018	−0.220	0.097	0
EF	0.055		0.060		0.044	−0.022	0.114	0	0.016	−0.019	0.083	0
HS	0.199	*	0.193	*	0.172	0.046	0.355	52	0.169	0.144	0.228	100
COVID case rate	0.285		0.310	*	0.151	−0.007	0.340	1	0.239	0.192	0.280	49
COVID mortality rate	−0.620	***	−0.548	***	−0.431	−0.856	0.063	60	−0.528	−0.627	−0.354	100
w_EB	–	–	0.054		–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-3
Model coefficient and significance summary of the Middle Atlantic region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		−0.005		0.010	0.089	0.107	0	0.120	0.089	0.157	37
SE	0.354	***	0.299	***	0.329	0.168	0.472	100	0.383	0.168	0.589	93
HD	0.031		−0.037		0.041	−0.054	0.150	0	−0.004	−0.054	0.040	0
ML	−0.529	***	−0.234	*	−0.505	−0.858	−0.310	100	−0.465	−0.858	−0.121	88
HT	0.009		−0.008		0.034	0.009	0.088	0	0.074	0.009	0.119	1
EF	−0.073		−0.057		−0.070	−0.105	−0.040	0	−0.062	−0.105	−0.020	8
HS	−0.055		−0.015		−0.041	−0.356	0.023	0	−0.016	−0.356	0.283	15
COVID case rate	0.027		−0.055		0.061	0.060	0.128	0	0.078	0.060	0.101	0
COVID mortality rate	−0.150		0.007		−0.153	−0.194	−0.131	47	−0.110	−0.194	−0.021	11
w_EB	–	–	0.115	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-4
Model coefficient and significance summary of the East North Central region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		0.003	0.919	−0.035	−0.467	0.403	90	0.063	−0.467	0.880	56
SE	0.243	***	0.162	0.001	0.300	0.214	0.441	100	0.302	0.214	0.409	100
HD	0.156	**	0.144	0.001	0.166	0.070	0.297	80	0.175	0.070	0.334	82
ML	−0.282	***	−0.164	0.000	−0.326	−0.296	−0.205	100	−0.280	−0.296	−0.267	100
HT	0.247	***	0.153	0.000	0.174	0.084	0.307	89	0.096	0.084	0.111	100
EF	0.027		0.027	0.361	0.042	0.013	0.121	20	0.034	0.013	0.071	23
HS	0.027		0.056	0.069	0.057	−0.033	0.091	2	0.090	−0.033	0.236	40
COVID case rate	−0.216	***	−0.103	0.006	−0.132	−0.088	0.015	63	−0.075	−0.088	−0.062	76
COVID mortality rate	0.063		0.046	0.160	0.016	−0.004	0.146	6	0.005	−0.004	0.017	0
w_EB	–	–	0.083	0.000	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-5
Model coefficient and significance summary of the West North Central region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		0.003		0.053	−0.990	0.751	58	0.019	−0.990	1.186	54
SE	0.339	***	0.191	***	0.402	0.388	0.736	94	0.434	0.388	0.483	100
HD	0.170	***	0.127	***	0.191	0.055	0.467	70	0.136	0.055	0.210	92
ML	−0.293	***	−0.154	***	−0.338	−0.346	−0.149	100	−0.292	−0.346	−0.248	100
HT	0.045		−0.008		0.074	−0.044	0.200	25	0.015	−0.044	0.063	0
EF	0.046		0.016		0.059	0.031	0.146	18	0.043	0.031	0.054	0
HS	−0.097	*	0.009		0.039	0.094	0.222	14	0.111	0.094	0.134	100
COVID case rate	−0.096	*	−0.073	*	−0.053	−0.460	0.166	30	−0.068	−0.460	0.217	28
COVID mortality rate	0.108	*	0.112	**	0.037	0.038	0.212	17	0.052	0.038	0.081	27
w_EB	–	–	0.105	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-6
Model coefficient and significance summary of the South Atlantic region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		−0.009		0.093	−0.474	1.294	53	0.028	−0.694	0.834	58
SE	0.398	***	0.336	***	0.414	0.032	0.875	85	0.379	0.361	0.403	100
HD	0.142	***	0.121	***	0.145	−0.199	0.628	36	0.144	−0.336	0.734	47
ML	−0.127	***	−0.100	***	−0.262	−0.786	0.055	66	−0.220	−0.227	−0.210	100
HT	0.134	***	0.104	**	0.129	−0.298	0.494	43	0.089	−0.066	0.283	34
EF	0.088	**	0.081	**	0.097	−0.100	0.359	35	0.084	0.076	0.092	100
HS	0.161	***	0.132	***	0.092	−0.322	0.341	42	0.078	−0.349	0.487	37
COVID case rate	−0.060		−0.062	*	−0.089	−0.478	0.361	26	−0.071	−0.104	−0.032	57
COVID mortality rate	0.163	***	0.129	***	0.109	−0.413	0.444	26	0.093	0.083	0.097	100
w_EB	–	–	0.049	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-7
Model coefficient and significance summary of the East South-Central region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		−0.002		−0.121	−0.744	0.407	90	−0.080	−0.553	0.391	88
SE	0.607	***	0.450	***	0.511	0.511	0.694	100	0.505	0.192	0.846	100
HD	0.040		0.041		0.046	0.046	0.186	15	0.049	−0.045	0.147	12
ML	0.162	***	0.066	*	−0.065	−0.065	0.162	32	−0.039	−0.265	0.212	39
HT	0.010		−0.001		0.053	0.053	0.239	20	0.014	−0.006	0.038	0
EF	−0.005		0.012		0.040	0.040	0.173	13	0.035	−0.021	0.135	20
HS	0.097	**	0.105	***	0.044	0.044	0.214	31	0.054	0.007	0.100	48
COVID case rate	−0.161	***	−0.093	*	−0.071	−0.071	0.084	44	−0.056	−0.202	0.098	37
COVID mortality rate	0.245	***	0.152	***	0.131	0.131	0.324	52	0.151	−0.144	0.374	54
w_EB	–	–	0.087	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-8
Model coefficient and significance summary of the West South-Central region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		0.001		0.079	−0.293	0.382	32	0.066	−0.543	0.579	41
SE	0.322	***	0.238	***	0.426	0.104	0.961	96	0.354	0.187	0.624	100
HD	0.201	***	0.176	***	0.137	−0.093	0.300	57	0.130	0.118	0.146	100
ML	−0.239	***	−0.175	***	−0.318	−0.647	−0.039	78	−0.331	−0.958	0.255	66
HT	0.172	***	0.140	***	0.175	−0.155	0.343	63	0.184	−0.122	0.317	85
EF	0.067		0.057		0.030	−0.140	0.139	2	0.022	0.015	0.031	0
HS	0.042		0.060		0.063	−0.128	0.215	14	0.082	−0.143	0.193	39
COVID case rate	0.009		0.004		0.033	−0.149	0.187	10	0.015	−0.014	0.038	0
COVID mortality rate	0.042		0.035		0.114	−0.134	0.396	33	0.122	−0.233	0.487	30
w_EB	–	–	0.063	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-9
Model coefficient and significance summary of Mountain region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		0.017	0.692	0.011	−0.649	0.367	50	0.011	−0.807	0.637	55
SE	0.325	***	0.178	0.017	0.189	−0.179	0.558	73	0.189	0.178	0.202	100
HD	0.076		0.050	0.369	0.047	−0.274	0.297	46	0.047	0.006	0.087	0
ML	−0.445	***	−0.308	0.000	−0.287	−0.843	0.060	81	−0.287	−0.329	−0.256	100
HT	0.111	*	0.052	0.301	0.144	−0.233	0.518	50	0.144	−0.419	0.848	54
EF	0.035		0.039	0.357	0.051	−0.100	0.249	25	0.051	−0.017	0.146	38
HS	0.097		0.150	0.002	0.254	−0.166	0.585	44	0.254	0.008	0.519	74
COVID case rate	−0.248	***	−0.181	0.000	−0.303	−0.805	−0.122	98	−0.303	−0.391	−0.241	100
COVID mortality rate	0.227	***	0.118	0.030	0.037	−0.177	0.641	42	0.037	−0.175	0.403	20
w_EB	–	–	0.724	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-10
Model coefficient and significance summary of the Pacific region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		–	–	–	–	–	–	–	–	–	–
SE	0.358	***	–	–	–	–	–	–	–	–	–	–
HD	0.043		–	–	–	–	–	–	–	–	–	–
ML	−0.322	***	–	–	–	–	–	–	–	–	–	–
HT	0.236	***	–	–	–	–	–	–	–	–	–	–
EF	−0.105		–	–	–	–	–	–	–	–	–	–
HS	−0.019		–	–	–	–	–	–	–	–	–	–
COVID case rate	0.518	***	–	–	–	–	–	–	–	–	–	–
COVID mortality rate	−0.379	*	–	–	–	–	–	–	–	–	–	–
w_EB	–	–	–	–	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-11
Model coefficient and significance summary of the Contiguous Pacific region.

	OLS		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		−0.018		−0.015	−0.457	0.532	93	0.061	−0.676	0.873	92
SE	0.415	***	0.291	**	0.341	0.099	0.551	56	0.435	0.416	0.455	100
HD	0.124		0.054		0.230	0.012	0.439	44	0.140	0.075	0.214	40
ML	−0.422	***	−0.262	**	−0.477	−0.615	−0.260	100	−0.512	−0.696	−0.355	100
HT	0.054		0.045		0.100	0.034	0.154	0	0.140	0.132	0.151	100
EF	−0.111		−0.067		−0.010	−0.125	0.043	0	0.036	0.018	0.050	0
HS	0.259	***	0.206	**	0.192	0.103	0.340	85	0.090	−0.086	0.358	23
COVID case rate	−0.142		−0.206	*	−0.224	−0.627	0.099	45	−0.306	−0.725	−0.071	50
COVID mortality rate	0.033		0.123		0.144	−0.143	0.423	44	0.154	0.037	0.252	59
w_EB	–	–	0.660	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-12
Model coefficient and significance summary of Mississippi.

	OLS		Categorical		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		–	–	0.012		0.06	0.026	0.138	2	0.121	0.054	0.173	41
SE	0.645	***	–	–	0.604	***	0.64	0.572	0.726	100	0.697	0.656	0.739	100
HD	0.081		–	–	0.054		0.109	0.005	0.223	29	0.075	0.029	0.128	0
ML	0.120		–	–	0.029		0.048	−0.13	0.174	0	−0.012	−0.245	0.329	17
HT	−0.146		–	–	−0.111		−0.124	−0.221	−0.051	33	−0.134	−0.203	−0.066	44
EF	−0.042		–	–	−0.006		−0.047	−0.156	0.05	4	−0.083	−0.277	0.094	30
HS	0.176	*	–	–	0.209	**	0.186	0.115	0.237	94	0.218	0.113	0.348	91
COVID case rate	−0.092		–	–	−0.074		−0.073	−0.193	0.025	18	−0.05	−0.145	0.021	0
COVID mortality rate	0.232	*	–	–	0.108		0.176	0.053	0.293	51	0.155	0.023	0.285	51
w_EB	–	–	–	–	0.622	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-13
Model coefficient and significance summary of Maine.

	OLS		Categorical		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		–	–	−0.033		–	–	–	–	–	–	–	–
SE	0.308		–	–	0.425	***	–	–	–	–	–	–	–	–
HD	0.117		–	–	0.010		–	–	–	–	–	–	–	–
ML	−0.450	*	–	–	−0.445	***	–	–	–	–	–	–	–	–
HT	0.211		–	–	0.310	***	–	–	–	–	–	–	–	–
EF	0.194		–	–	0.212	***	–	–	–	–	–	–	–	–
HS	0.298		–	–	0.143		–	–	–	–	–	–	–	–
COVID case rate	−0.054		–	–	−0.280		–	–	–	–	–	–	–	–
COVID mortality rate	−0.098		–	–	−0.053		–	–	–	–	–	–	–	–
w_EB	–	–	–	–	−0.604	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-14
Model coefficient and significance summary of Alabama.

	OLS		Categorical		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		–	–	0.001		0.049	−0.021	0.14	0	0.049	−0.021	0.14	0
SE	0.739	***	–	–	0.749	***	0.695	0.62	0.748	100	0.695	0.62	0.748	100
HD	−0.166		–	–	−0.239	*	−0.127	−0.218	−0.05	15	−0.127	−0.218	−0.05	24
ML	−0.030		–	–	−0.104		0.018	−0.072	0.133	0	0.018	−0.072	0.133	22
HT	0.144		–	–	0.033		0.145	0.063	0.209	9	0.145	0.063	0.209	78
EF	0.042		–	–	0.058		0.048	−0.004	0.089	0	0.048	−0.004	0.089	6
HS	−0.012		–	–	−0.006		−0.018	−0.088	0.083	0	−0.018	−0.088	0.083	3
COVID case rate	−0.012		–	–	0.027		−0.014	−0.093	0.119	0	−0.014	−0.093	0.119	0
COVID mortality rate	0.195	*	–	–	0.144	*	0.138	0.033	0.216	43	0.138	0.033	0.216	27
w_EB	–	–	–	–	0.536	**	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-15
Model coefficient and significance summary of Montana.

	OLS		Categorical		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		–	–	0.010		–	–	–	–	–	–	–	–
SE	0.145		–	–	0.149		–	–	–	–	–	–	–	–
HD	−0.102		–	–	−0.093		–	–	–	–	–	–	–	–
ML	−0.083		–	–	−0.075		–	–	–	–	–	–	–	–
HT	−0.257		–	–	−0.256		–	–	–	–	–	–	–	–
EF	−0.008		–	–	0.000		–	–	–	–	–	–	–	–
HS	0.173		–	–	0.175		–	–	–	–	–	–	–	–
COVID case rate	−0.283		–	–	−0.293		–	–	–	–	–	–	–	–
COVID mortality rate	0.445	*	–	–	0.418	*	–	–	–	–	–	–	–	–
w_EB	–	–	–	–	0.152		–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Table I-16
Model coefficient and significance summary of Georgia.

	OLS		Categorical		Spatial lag		GWR				MGWR			
Variables	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Mean	Min.	Max.	%Sig.	Mean	Min.	Max.	%Sig.
(Intercept)	0.000		–	–	0.001		0.049	−0.021	0.14	0	0.049	−0.021	0.14	0
SE	0.739	***	–	–	0.749	***	0.695	0.62	0.748	100	0.695	0.62	0.748	100
HD	−0.166		–	–	−0.239		−0.127	−0.218	−0.05	15	−0.127	−0.218	−0.05	24
ML	−0.030		–	–	−0.104		0.018	−0.072	0.133	0	0.018	−0.072	0.133	22
HT	0.144		–	–	0.033		0.145	0.063	0.209	9	0.145	0.063	0.209	78
EF	0.042		–	–	0.058		0.048	−0.004	0.089	0	0.048	−0.004	0.089	6
HS	−0.012	*	–	–	−0.006	**	−0.018	−0.088	0.083	0	−0.018	−0.088	0.083	3
COVID case rate	−0.012		–	–	0.027		−0.014	−0.093	0.119	0	−0.014	−0.093	0.119	0
COVID mortality rate	0.195	*	–	–	0.144		0.138	0.033	0.216	43	0.138	0.033	0.216	27
w_EB	–	–	–	–	0.536	***	–	–	–	–	–	–	–	–

* (P < 0.05), ** (P < 0.01), *** (P < 0.001).

Appendix II. Local regression coefficients of MGWR for all the independent variables except SE

Here we present the local coefficient maps of MGWR for Intercept, HD, ML, HT, EF, HS, Covid case rate, and Covid mortality rate in Figs. II-1–II-8.

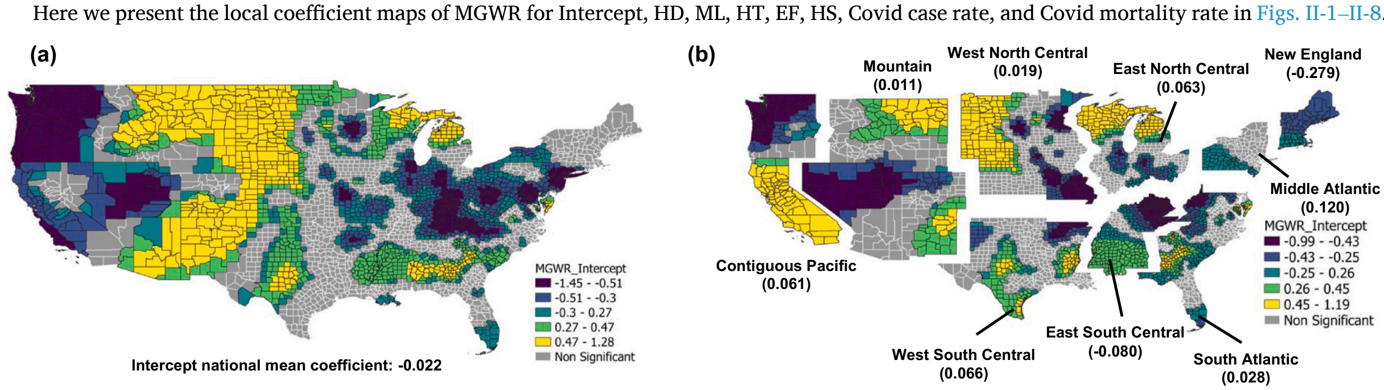

Fig. II-1. Local regression coefficients of MGWR for Intercept at: (a) national level; and (b) census region level.

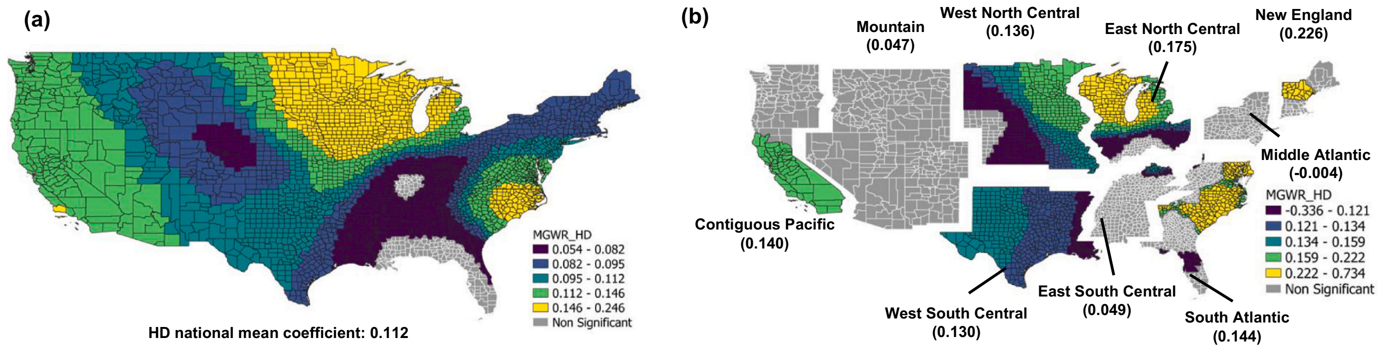

Fig. II-2. Regression coefficients of MGWR for HD at: (a) national level; and (b) census region level.

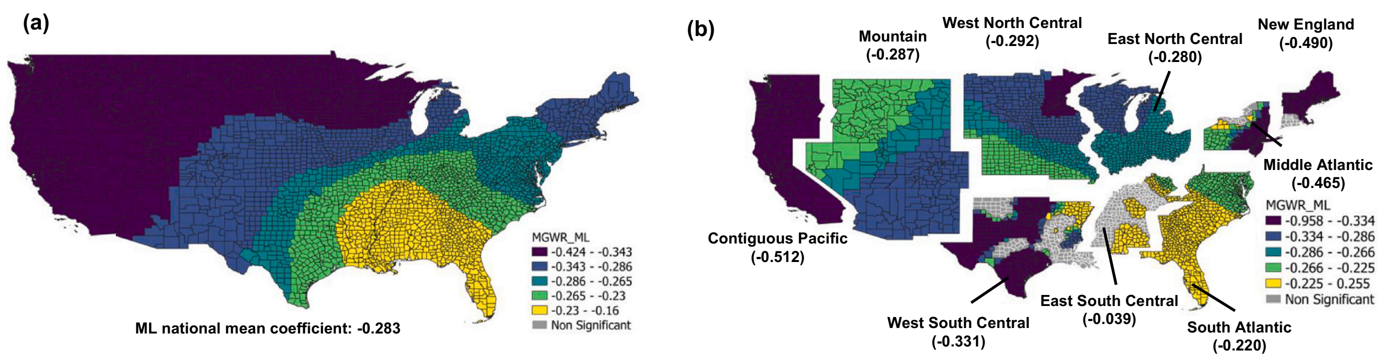

Fig. II-3. Regression coefficients of MGWR for ML at: (a) national level; and (b) census region level.

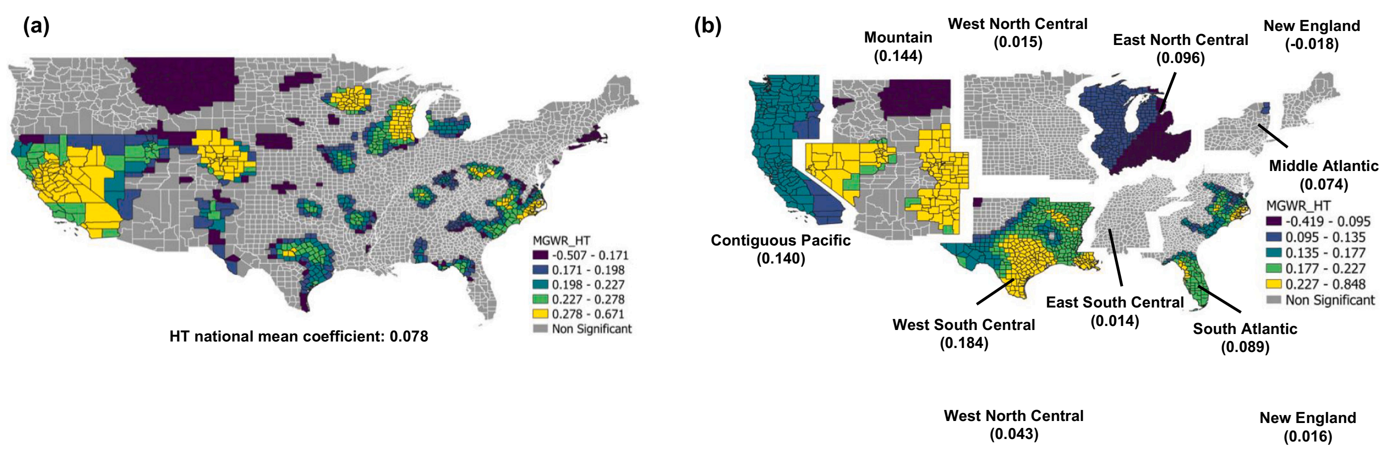

Fig. II-4. Regression coefficients of MGWR for HT at: (a) national level; and (b) census region level.

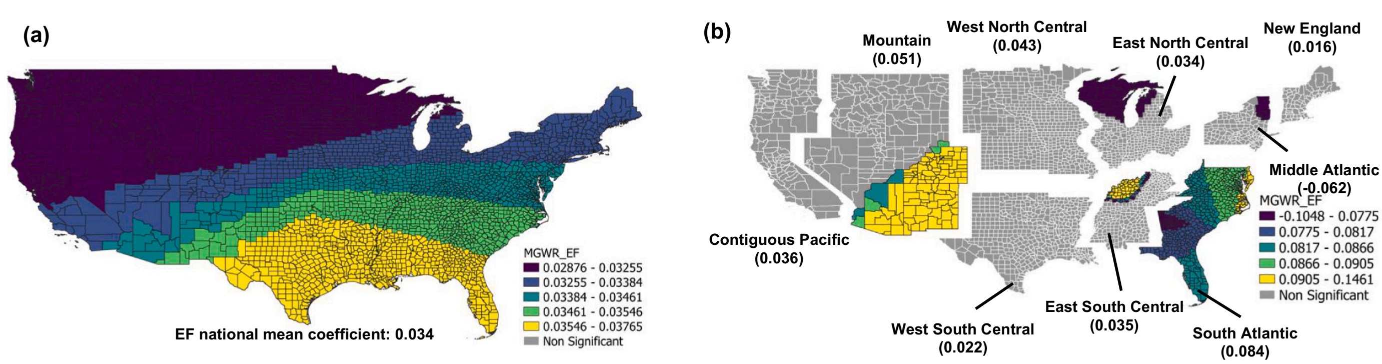

Fig. II-5. Regression coefficients of MGWR for EF at: (a) national level; and (b) census region level.

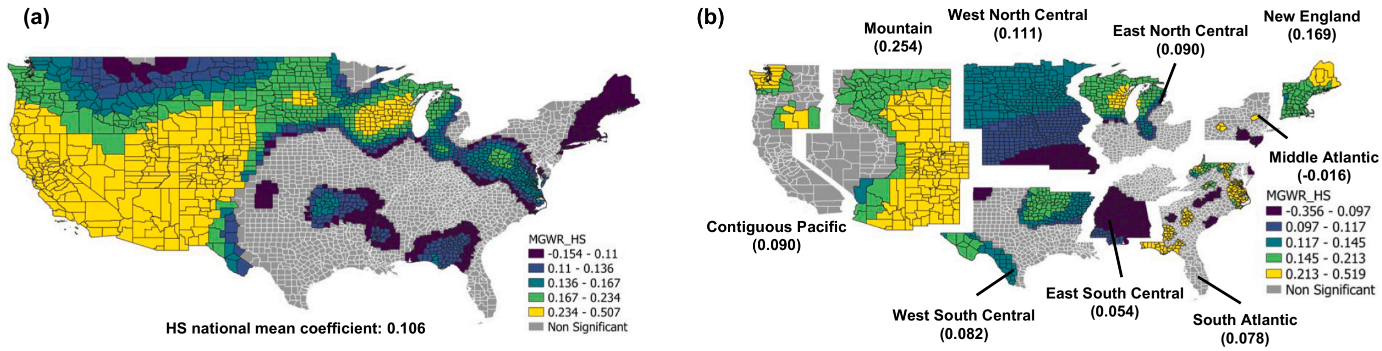

Fig. II-6. Regression coefficients of MGWR for EF at: (a) national level; and (b) census region level.

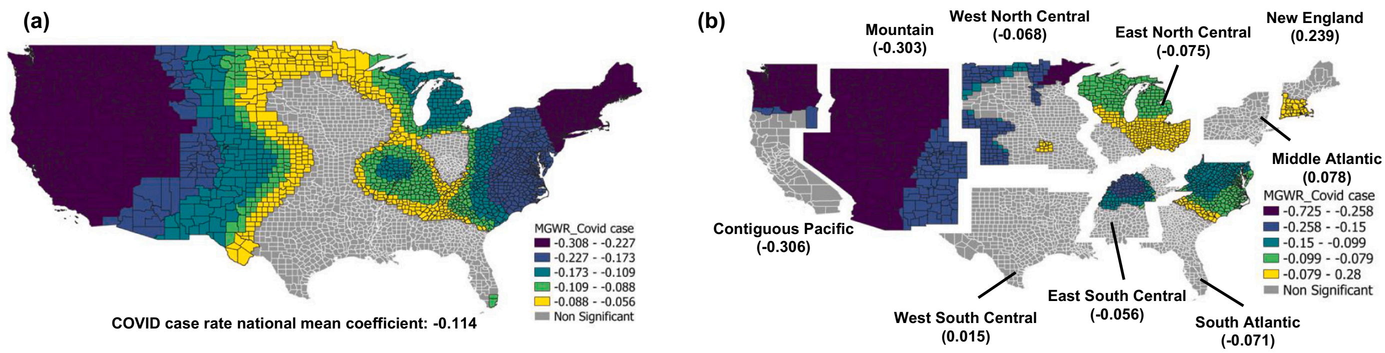

Fig. II-7. Regression coefficients of MGWR for COVID case rate at: (a) national level; and (b) census region level.

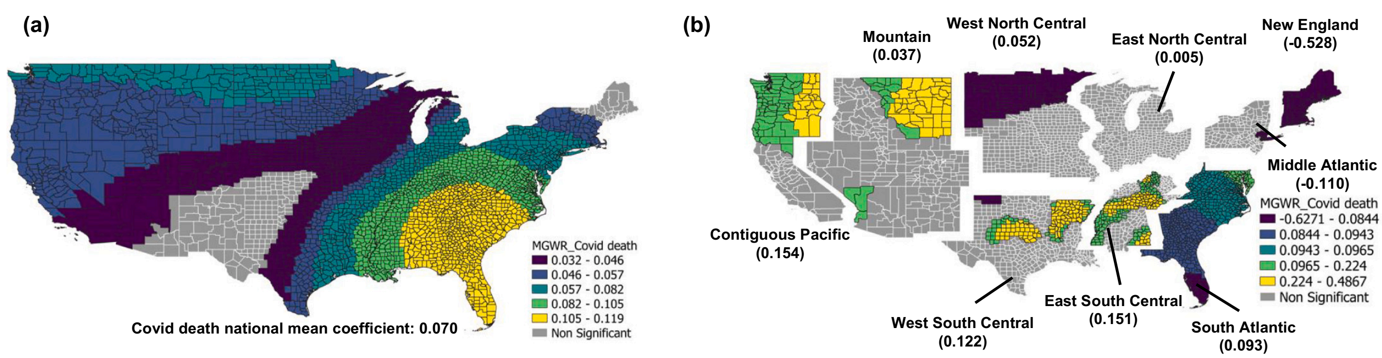

Fig. II-8. Regression coefficients of MGWR for COVID mortality rate at: (a) national level; and (b) census region level.

References

[1] U.S. Department of Energy, Low-Income Community Energy Solutions, (n.d.) 1–6. https://www.energy.gov/eere/slsc/low-income-community-energy-solutions (accessed September 26, 2022).

[2] N. Fefferman, C.F. Chen, G. Bonilla, H. Nelson, C.P. Kuo, How limitations in energy access, poverty, and socioeconomic disparities compromise health interventions for outbreaks in urban settings 24 (2021), 103389, https://doi.org/10.1016/j.isci.2021.103389.

[3] G.O. Boateng, M.R. Balogun, F.O. Dada, F.A. Armah, Household energy insecurity: dimensions and consequences for women, infants and children in low- and middle-income countries, Soc. Sci. Med. 258 (2020), 113068, https://doi.org/10.1016/j.socscimed.2020.113068.

[4] T.G. Reames, D.M. Daley, J.C. Pierce, Exploring the nexus of energy burden, social capital, and environmental quality in shaping health in US counties, Int. J. Environ. Res. Public Health 18 (2021) 1–13, https://doi.org/10.3390/ijerph18020620.

[5] J. Bohr, A.C. McCreery, Do energy burdens contribute to economic poverty in the United States? A panel analysis, Soc. Forces. 99 (2020) 155–177, https://doi.org/10.1093/sf/soz131.

[6] S. Jessel, S. Sawyer, D. Hernández, Energy, poverty, and health in climate change: a comprehensive review of an emerging literature, Front. Public Health 7 (2019), https://doi.org/10.3389/fpubh.2019.00357.

[7] J. Bhattacharya, T. DeLeire, S. Haider, J. Currie, Heat or eat? Cold-weather shocks and nutrition in poor American families, Am. J. Public Health 93 (2003) 1149–1154, https://doi.org/10.2105/ajph.93.7.1149.

[8] M. Graff, S. Carley, COVID-19 assistance needs to target energy, Nat. Energy 5 (2020) 352–354, https://doi.org/10.1038/s41560-020-0620-y.

[9] G. D'Amato, S.T. Holgate, R. Pawankar, D.K. Ledford, L. Cecchi, M. Al-Ahmad, F. Al-Enezi, S. Al-Muhsen, I. Ansotegui, C.E. Baena-Cagnani, D.J. Baker, H. Bayram, K.C. Bergmann, L.P. Boulet, J.T.M. Buters, M. D'Amato, S. Dorsano, J. Douwes, S. E. Finlay, D. Garrasi, M. Gómez, T. Haahtela, R. Halwani, Y. Hassani, B. Mahboub, G. Marks, P. Michelozzi, M. Montagni, C. Nunes, J.J.W. Oh, T.A. Popov, J. Portnoy, E. Ridolo, N. Rosário, M. Rottem, M. Sánchez-Borges, E. Sibanda, J.J. Sienra-Monge, C. Vitale, I. Annesi-Maesano, Meteorological conditions, climate change, new emerging factors, and asthma and related allergic disorders. A statement of the world allergy organization, World Allergy Organ. J. 8 (2015) 25, https://doi.org/10.1186/s40413-015-0073-0.

[10] V. Castán Broto, J. Kirshner, Energy access is needed to maintain health during pandemics, Nat. Energy 5 (2020) 419–421, https://doi.org/10.1038/s41560-020-0625-6.

[11] C.F. Chen, J. Feng, N. Luke, C.P. Kuo, J.S. Fu, Localized energy burden, concentrated disadvantage, and the feminization of energy poverty IScience25 (2022), 104139, https://doi.org/10.1016/j.isci.2022.104139.

[12] A. Drehobl, L. Ross, R. Ayala, How high are household energy burdens? An assessment of National Metropolitan Energy Burdens across the U.S., Washington DC, U.S. https://www.aceee.org/research-report/u2006, 2020.

[13] R.S. Liévanos, Racialized structural vulnerability: neighborhood racial composition, concentrated disadvantage, and fine particulate matter in California, Int. J. Environ. Res. Public Health 16 (2019) 1–24, https://doi.org/10.3390/ijerph16173196.
[14] A. Maxim, E. Grubert, Anticipating climate-related changes to residential energy burden in the United States: advance planning for equity and resilience, Environ. Justice 15 (2022) 139–148, https://doi.org/10.1089/env.2021.0056.
[15] CDC Social Vulnerability Index, Agency Toxic Subst. Dis. Regist, 2020.
[16] Surgo Ventures, The U.S. COVID Community Vulnerability Index (CCVI), 2021.
[17] Surgo Ventures, Vulnerable communities, and COVID-19: the damage done, and the way forward. https://surgoventures.org/resource-library/report-vulnerable-communities-and-covid-19, 2021.
[18] P. Smittenaar, N. Stewart, S. Sutermaster, L. Coome, A. Dibner-Dunlap, M. Jain, Y. Caplan, C. Campigotto, S.K. Sgaier, A COVID-19 Community Vulnerability Index to Drive Precision Policy in the US, MedRxiv, 2021, https://doi.org/10.1101/2021.05.19.21257455.
[19] C.C. Brown, S.G. Young, G.C. Pro, COVID-19 vaccination rates vary by community vulnerability: a county-level analysis, Vaccine 39 (2021) 4245–4249, https://doi.org/10.1016/j.vaccine.2021.06.038.
[20] S.C. Melvin, C. Wiggins, N. Burse, E. Thompson, M. Monger, The role of public health in COVID-19 emergency response efforts from a rural health perspective, Prev. Chronic Dis. 17 (2020) 1–6, https://doi.org/10.5888/PCD17.200256.
[21] A. Tiwari, A.V. Dadhania, V.A.B. Ragunathrao, E.R.A. Oliveira, Using machine learning to develop a novel COVID-19 vulnerability index (C19VI), Sci. Total Environ. 773 (2021), 145650, https://doi.org/10.1016/j.scitotenv.2021.145650.
[22] M.H. Lopez, L. Rainie, A. Budiman, Financial and health impacts of COVID-19 vary widely by race and ethnicity, Paw Res. Cent, 2020. https://www.pewresearch.org/fact-tank/2020/05/05/financial-and-health-impacts-of-covid-19-vary-widely-by-race-and-ethnicity/.
[23] C.F. Chen, H. Nelson, X. Xu, G. Bonilla, N. Jones, Beyond technology adoption: examining home energy management systems, energy burdens and climate change perceptions during COVID-19 pandemic, Renew. Sust. Energ. Rev. 145 (2021), 111066, https://doi.org/10.1016/j.rser.2021.111066.
[24] C.F. Chen, J. Greig, H. Nelson, F. Li, When disadvantage collides: the concentrated effects of energy insecurity and internet burdens in the United States, Energy Res. Soc. Sci. 91 (2022), 102713, https://doi.org/10.1016/j.erss.2022.102713.
[25] T. Memmott, S. Carley, M. Graff, D.M. Konisky, Sociodemographic disparities in energy insecurity among low-income households before and during the COVID-19 pandemic, Nat. Energy 6 (2021) 186–193, https://doi.org/10.1038/s41560-020-00763-9.
[26] M. Dikeç, Justice and the spatial imagination, Environ. Plan. A Econ. Sp. 33 (2001) 1785–1805, https://doi.org/10.1068/a3467.
[27] E.W. Soja, Seeking Spatial Justice, U of Minnesota Press, 2013.
[28] S. Bouzarovski, N. Simcock, Spatializing energy justice, Energy Policy 107 (2017) 640–648, https://doi.org/10.1016/j.enpol.2017.03.064.
[29] B. Mashhoodi, D. Stead, A. van Timmeren, Spatial homogeneity and heterogeneity of energy poverty: a neglected dimension, Ann. GIS 25 (2019) 19–31, https://doi.org/10.1080/19475683.2018.1557253.
[30] D. Moore, A.L. Webb, Evaluating energy burden at the urban scale: a spatial regression approach in Cincinnati, Ohio, Energy Policy 160 (2022), 112651, https://doi.org/10.1016/j.enpol.2021.112651.
[31] T.G. Reames, Targeting energy justice: exploring spatial, racial/ethnic and socioeconomic disparities in urban residential heating energy efficiency, Energy Policy 97 (2016) 549–558, https://doi.org/10.1016/j.enpol.2016.07.048.
[32] D.J. Bednar, T.G. Reames, G.A. Keoleian, The intersection of energy and justice: modeling the spatial, racial/ethnic and socioeconomic patterns of urban residential heating consumption and efficiency in Detroit, Michigan, Energy Build. 143 (2017) 25–34, https://doi.org/10.1016/j.enbuild.2017.03.028.
[33] US Department of Energy, Low-Income Energy Affordability Data (LEAD) tool. https://www.energy.gov/eere/slsc/low-income-energy-affordability-data-lead-tool, 2021. (Accessed 21 November 2021).
[34] U.S. Energy Information Administration, U.S. monthly energy review. https://www.eia.gov/%0Atotalenergy/data/monthly/, 2020.
[35] J.A. Hipp, N. Chalise, Spatial analysis and correlates of county-level diabetes prevalence, 2009–2010, Prev. Chronic Dis. 12 (2015), https://doi.org/10.5888/pcd12.140404.
[36] Surgo Ventures, The U.S. CCVI vulnerability — how well a community handles the repercussions of a COVID-19 outbreak. https://precisionforcovid.org/ccvi, 2021.
[37] Johns Hopkins University, COVID-19 United States cases by county. https://coronavirus.jhu.edu/, 2021.
[38] U.S. Census Bureau, Explore Census data. https://data.census.gov/cedsci/, 2021.
[39] L. Anselin, S.J. Rey, Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL, 2014.
[40] C. Brunsdon, A.S. Fotheringham, M.E. Charlton, Geographically weighted regression: a method for exploring spatial nonstationarity, Geogr. Anal. 28 (1996).
[41] A.S. Fotheringham, W. Yang, W. Kang, Multiscale geographically weighted regression (MGWR), Ann. Am. Assoc. Geogr. 107 (2017) 1247–1265, https://doi.org/10.1080/24694452.2017.1352480.
[42] L. Anselin, Spatial Econometrics: Methods and Models, Springer, 1988.
[43] U.S. Department of Energy, Low-income household energy burden varies among states — efficiency can help in all of them. https://www.energy.gov/sites/prod/files/2019/01/f58/WIP-Energy-Burden_final.pdf, 2018.
[44] S.J. Rey, L. Anselin, X. Li, R. Pahle, J. Laura, W. Li, J. Koschinsky, Open geospatial analytics with PySAL, ISPRS Int. J. Geo Inf. 4 (2015) 815–836, https://doi.org/10.3390/ijgi4020815.
[45] T.M. Oshan, Z. Li, W. Kang, L.J. Wolf, A. Stewart Fotheringham, MGWR: a python implementation of multi-scale geographically weighted regression for investigating process spatial heterogeneity and scale, ISPRS Int. J. Geo Inf. 8 (2019), https://doi.org/10.3390/ijgi8060269.
[46] C.E. Kontokosta, V.J. Reina, B. Bonczak, Energy cost burdens for low-income and minority households: evidence from energy benchmarking and audit data in five U.S. cities, J. Am. Plan. Assoc. 86 (2020) 89–105, https://doi.org/10.1080/01944363.2019.1647446.
[47] E. Scheier, N. Kittner, A measurement strategy to address disparities across household energy burdens, Nat. Commun. 13 (2022) 1–11, https://doi.org/10.1038/s41467-021-27673-y.
[48] I. Faiella, L. Lavecchia, Energy poverty. How can you fight it, if you can't measure it? Energy Build. 233 (2021) 1–11, https://doi.org/10.1016/j.enbuild.2020.110692.
[49] R. Cluett, J. Amann, S. Ou, Building better energy efficiency programs for low-income households. https://aceee.org/sites/default/files/publications/researchreports/a1601.pdf, 2016.
[50] American Council for an Energy-Efficient Economy, Energy burden report: low-income, Black, Hispanic, and Native American households face high energy burdens. https://www.aceee.org/energy-burden, 2020. (Accessed 9 July 2022).
[51] NYC Mayor's Office of Sustainability, Understanding and alleviating energy cost burden in New York City, New York. https://energyefficiencyforall.org/sites/default/files/Liftingthe, 2019.
[52] New York State Executive Chamber, NYS public service commision expands energy discount programs to provide more than $248 million in savings. https://www3.dps.ny.gov/pscweb/webfileroom.nsf/ArticlesByCategory/379F49841174087E85257FB80063D106/$File/gov.051916.pdf, 2016. (Accessed 9 July 2022).
[53] C.F. Chen, T. Dietz, N.H. Fefferman, J. Greig, K. Cetin, C. Robinson, L. Arpan, M. Schweiker, B. Dong, W. Wu, Y. Li, H. Zhou, J. Wu, J. Wen, J.S. Fu, T. Hong, D. Yan, H. Nelson, Y. Zhu, X. Li, L. Xie, R. Fu, Extreme events, energy security and equality through micro- and macro-levels: concepts, challenges and methods, Energy Res. Soc. Sci. 85 (2022), 102401, https://doi.org/10.1016/j.erss.2021.102401.