communications earth & environment

Article

https://doi.org/10.1038/s43247-024-01337-3

Urban form and structure explain variability in spatial inequality of property flood risk among US counties

Check for updates

Junwei Ma & Ali Mostafavi

Understanding the relationship between urban form and structure and spatial inequality of property flood risk has been a longstanding challenge in urban planning and emergency management. Here we explore eight urban form and structure features to explain variability in spatial inequality of property flood risk among 2567 US counties. Using datasets related to human mobility and facility distribution, we identify notable variation in spatial inequality of property flood risk, particularly in coastline and metropolitan counties. The results reveal variations in spatial inequality of property flood risk can be explained based on principal components of development density, economic activity, and centrality and segregation. The classification and regression tree model further demonstrates how these principal components interact and form pathways that explain spatial inequality of property flood risk. The findings underscore the critical role of urban planning in mitigating flood risk inequality, offering valuable insights for crafting integrated strategies as urbanization progresses.

Climate change has increased the frequency and intensity of flooding, elevating concern about the threat it poses to millions of people[1–4], including extensive property damages[5–7]. While flood risk is often examined with a focus on natural factors, such as hydrology and land topography[8–10], the importance of urban form and structure in shaping the spatial distribution of flood risk in cities is increasingly being recognized[11,12]. In particular, understanding the relationship between urban form and structure and spatial distribution of flood risk may hold the key to integrated urban design strategies to effectively address flood risk as cities continue to grow and develop.

Urban form and structure refer to the spatial configuration and organization of cities, such as development patterns, facility distribution, and economic activity[13–15]. Urban form and structure capture characteristics such as population distribution[16], development density[17], human mobility[18], and the centralization of infrastructure[19] that can shape the spatial distribution of flood risk. Yet data-driven insights are missing to inform about the relationship between urban form and structure and spatial distribution of flood risk. This limitation has hindered development and implementation of integrated urban design strategies to inform growth and development of cities while addressing flood risks to people and properties.

Of particular importance is uncovering the extent to which urban form and structure explain variation in spatial inequality of property flood risk in cities. Spatial inequality of property flood risk captures the extent to which properties located in different areas of a city have similar levels of flood risk. The risk of property flooding is not distributed equally across all areas and communities in a city[20,21]. Studies have shown that low-income communities and communities of color are often more vulnerable to flooding than wealthier, white communities[22–25]. Spatial inequality of property flood risk is in part influenced by the patterns of growth and development in cities[26]. A critical and yet unanswered question is the extent to which features of urban form and structure (such as development density, economic activity, centrality, and segregation) can explain the variation in spatial inequality of property flood risk among cities.

To address this research gap, this study investigates the extent to which urban form and structure explain variability in spatial inequality of property flood risk among US counties. The main research questions guiding the study are twofold: (1) What is the extent of spatial inequality in property flood risk in US counties? (2) To what extent are spatial inequality of property flood risk explained by different features of urban form and structure? To answer these questions, we use rich datasets related to property flood risk to quantify spatial inequality of property flood risk in US counties and delineate features of urban form and structure using high-resolution human mobility and facility distribution data. We begin by evaluating spatial inequality of property flood risk using the metric of spatial Gini index (SGI), a measure of spatial inequality, for 2567 counties in the United States, identifying notable variations in spatial inequality of property

Urban Resilience.AI Lab, Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77840, USA.
e-mail: amostafavi@civil.tamu.edu

flood risk across counties. We then explore how urban form and structure may be shaping this spatial inequality of property flood risk, by examining eight distinct urban features (i.e., population density, point of interest (POI) density, road density, and minority segregation. income segregation, urban centrality index (UCI), gross domestic product (GDP), and human mobility index (HMI)) to assess their potential relationships. We use principal component analysis (PCA) to identify three key factors shaping spatial inequality of property flood risk: development density, centrality and segregation, and economic activity. We then develop a classification and regression tree (CART) model to examine ways these factors interact in forming pathways leading to different levels of spatial inequality in property flood risk in US counties. Finally, we discuss integrated urban design strategies for addressing spatial inequality of property flood risk to inform future growth and expansion of cities. Our study provides unique and valuable insights into the intricate relationship between urban form, urban structure, and spatial inequality of property flood risk. These insights carry profound implications for integrated urban design strategies aimed at mitigating property flood risk, particularly as cities undergo further expansion and development.

Results

Variation in spatial inequality of property flood risk among US counties

To explore the variation in spatial inequality of property flood risk among US counties, the SGI was calculated based on the Flood Factor dataset from First Street Foundation (see Methods for detail). Probability density function and complementary cumulative distribution function of the SGI were plotted in Fig. 1a, b to examine the distribution characteristics. The results show that the probability density has an unimodal distribution with a mean of 0.255 and a standard deviation of 0.135. The complementary cumulative density, on the other hand, exhibits a heavy-tailed distribution, indicating the existence of a number of US counties with substantial spatial inequality of property flood risk. The 20% of counties with the highest SGI values account for 36% of the property flood risk, while the 20% of counties with the lowest SGI values only account for 7% of the property flood risk. These results suggest a notable variation in spatial inequality of property flood risk across US counties, highlighting the importance of examining the underlying factors shaping this disparity.

To compare the spatial distribution of SGI and overall flood risk, we created county-level visualizations of SGI and overall flood risk in the United States. The results show a notable variation in the SGI across US counties, with some counties exhibiting extremely high levels of spatial inequality in property flood risk (top panel of Fig. 1c). Interestingly, the high flood-risk property areas shown in the overall flood risk visualization (bottom panel of Fig. 1c), such as counties in the Gulf Coast region, do not necessarily correspond to high levels of spatial inequality in our analysis (see highlighted areas in Fig. 1c). On the contrary, some counties in California and the US Northeast have a much higher SGI than the counties in Texas, yet they have lower overall flood risk. The findings indicate that the spatial inequality of property flood risk is not solely determined by the extent of overall flood risk.

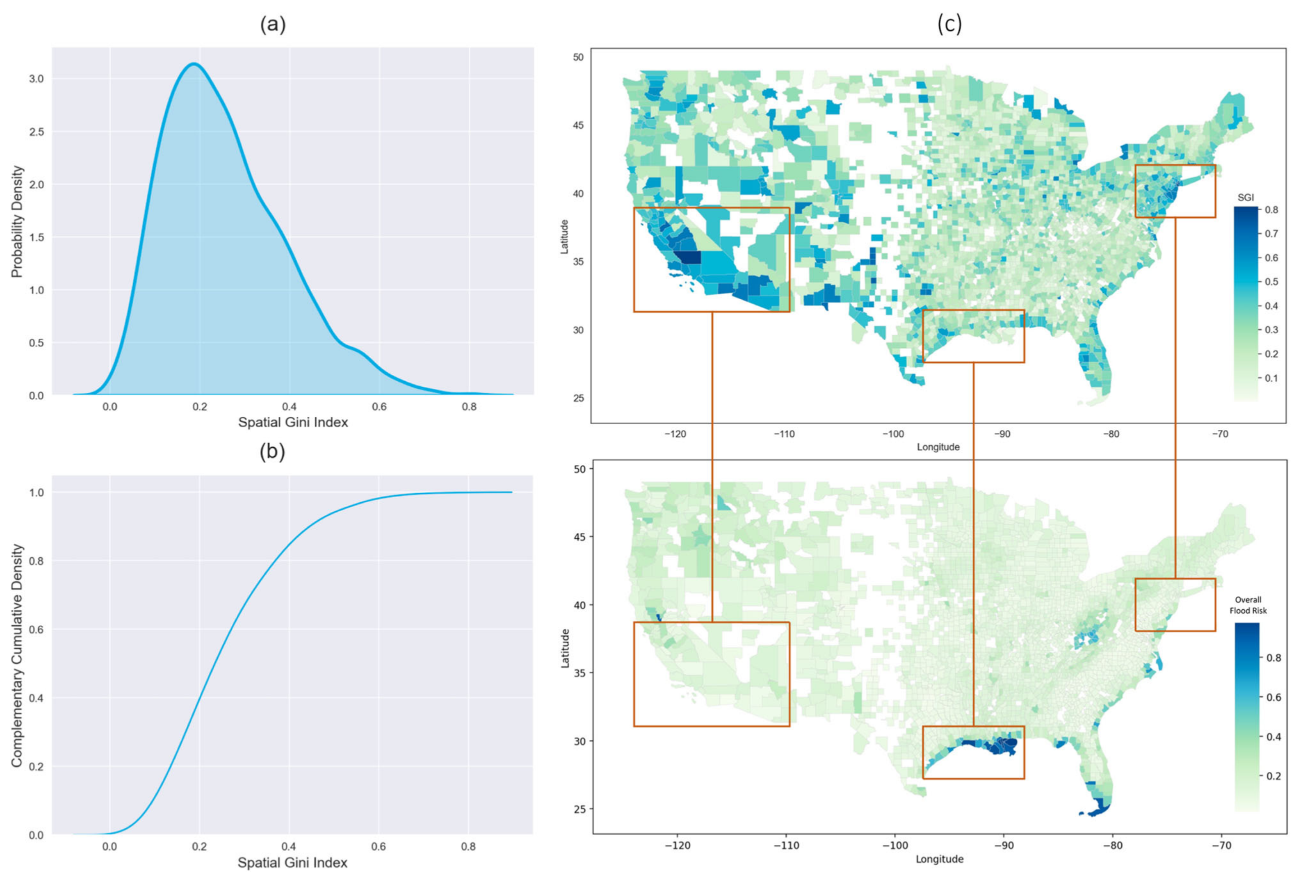


Fig. 1 | Disparities of spatial inequality in property flood risk among US counties. **a** Probability density function of SGI. **b** Complementary cumulative distribution function of SGI. **c** County-level visualized comparison between the SGI and overall flood risk. The overall flood risk was calculated by dividing the number of properties with flood factor risk scores larger than five by the total number of properties in each county (see Methods for detail). The original data comes from the First Street Foundation. We focus on 2567 counties in the continental United States for which all data are available. The top panel of (**c**) uses a green-to-blue color gradient to indicate the range of the SGI across different counties of the United States, with lighter green representing a more equal distribution of property flood risk (lower SGI) and darker blue representing greater spatial disparity of property flood risk (higher SGI). The bottom panel of (**c**) employs a green-to-blue color gradient to show the range of overall flood risk across the United States, with lighter green shades indicating lower flood risk and darker blue shades indicating higher flood risk. It's important to note that the overall flood risk does not reflect the spatial distribution of flood risk as the SGI does.

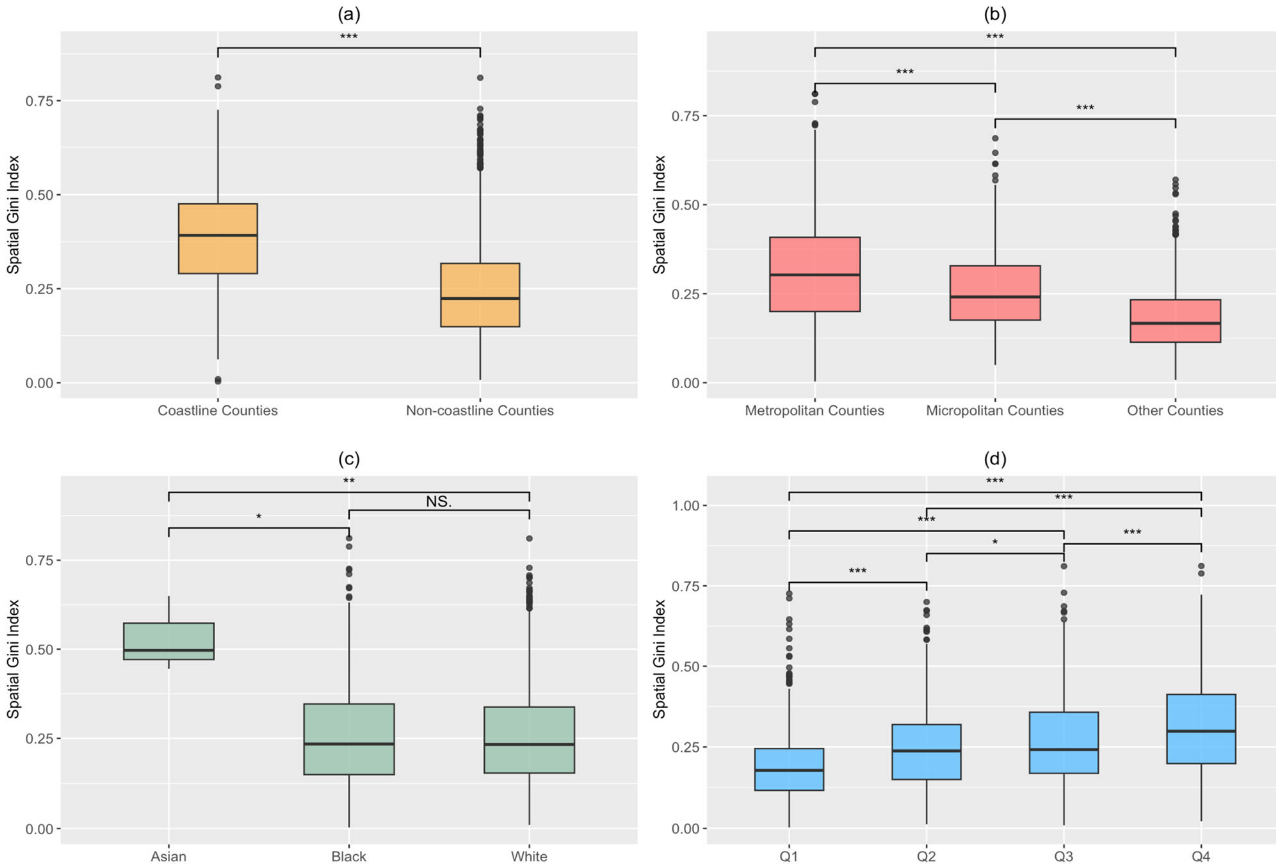


Fig. 2 | Variations in spatial inequality of property flood risk among county groups. a Coastline/non-coastline counties, as defined by the US Census Bureau[60]. We identified 211 coastline counties and 2356 non-coastline counties. **b** Metropolitan/micropolitan/other counties, as identified by US Census Bureau[61]. The numbers are 1101, 601, and 865 respectively. **c** Asian/Black/White counties. The overrepresented race for each county is identified by comparing the proportion of non-Hispanic White, non-Hispanic Black, and non-Hispanic Asian populations to the average for all the counties (See Supplementary Note 1 for more details). We totally have 2025 White-dominated counties, 539 Black-dominated counties, and 3 Asian-dominated counties. We obtained the racial population data from the 2020 race and ethnicity data from the US Census Bureau[38]. **d** We used the median income quantile from 2020 American Community Survey[48] to distinguish counties groups: Q1 (642 counties), Q2 (641 counties), Q3 (642 counties), and Q4 (642 counties). Note: The orange shading represents the boxplots for the SGI in coastline and non-coastline counties. The red shading differentiates the boxplots among metropolitan, micropolitan, and other counties. The green shading is used for the boxplots representing different racial groups: Asian, Black, and White. The blue shading indicates the median income quartiles (Q1, Q2, Q3, Q4) of the SGI distribution. The central line indicates the median, the box edges represent the interquartile range, and the outliers are indicated by individual points. ***$p \leq 0.01$, **$p \leq 0.05$, *$p \leq 0.1$. P values are obtained from t tests.

To further investigate the variations in spatial inequality of property flood risk across US counties, we classified all the counties into different groups: coastline/non-coastline counties, metropolitan/micropolitan/other counties, Asian/Black/White counties, and income Q1/Q2/Q3/Q4 counties. As shown in Fig. 2, the results show notable variations in the SGI across different groups, with some groups exhibiting higher levels of spatial inequality in property flood risk than others. For instance, the boxplot for coastline versus non-coastline counties indicates that the coastline counties tend to have higher SGI, suggesting a higher degree of spatial inequality in property flood risk within these areas. Similarly, metropolitan counties exhibit higher SGI than micropolitan counties and other counties, indicating that metropolitan counties have a more uneven distribution of property flood risk. The spatial visualizations of SGI distribution in these groups can be found in Supplementary Fig. 1. These results highlight the importance of considering differences in urban form and structure that shape the spatial inequality of property flood risk of counties.

Empirical statistics of urban form and structure features

We collected a diverse range of datasets through systematic literature review, enabling us to capture various heterogeneous features related to urban form and urban structure. Urban form and structure are concepts in urban planning and geography that describe the physical layout and organization of cities[13–15]. Urban form pertains to the physical aspects of the urban environment and its various elements. We measured minority segregation, income segregation, population density, and GDP in terms of urban form. Urban structure, on the other hand, deals with the functional aspects of the city and how different locations are connected and used. It includes UCI, POI density, road density, and HMI. The definition of features and the datasets used to develop these features can be found in the Methods section. The literature referenced for the feature screening can be found in Supplementary Table 1.

We divided these features into two aspects as they represent different dimensions of the urban environment that could impact the spatial inequality of property flood risk. Urban form captures social and economic factors that influence the spatial distribution of populations and concentration of economic activities. Urban structure, on the other hand, relates to the structural layout of cities, which can affect land use and development patterns shaping property flood risks. These features could potentially explain the spatial inequality of property flood risk. For example, high levels of minority segregation could be associated with redlining that exacerbates property flood risk in already vulnerable areas, while dense urban centers with a high concentration of POI and human mobility could exacerbate impervious surface and reduce green space.

Our initial analysis involved mapping the eight features to examine variations in terms of urban form and structure among counties. Figure 3 illustrates the distribution for the eight features, revealing that the heterogeneity of urban form and structure among US counties. With the exception of road density and income segregation, all features showed higher values in coastal areas compared to non-coastal areas. Counties in metropolitan areas, such as those in California, Florida, and the Northeastern metropolitan area exhibited particularly high values in the features of population density, POI density, minority segregation, UCI, GDP, and HMI.

In the next step, we examined the urban form and urban structure features for counties with different levels of spatial inequality of property

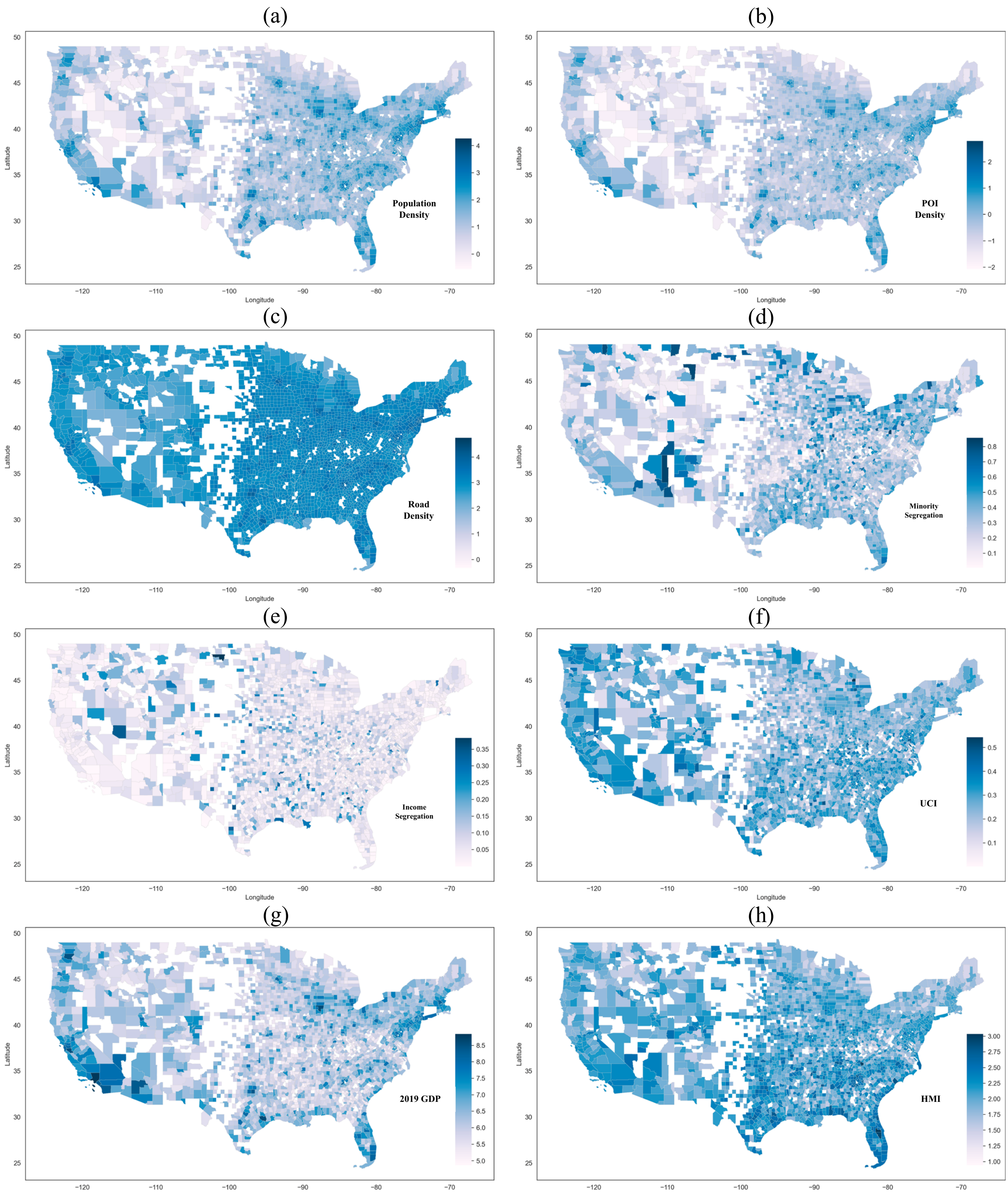


Fig. 3 | Heterogeneity in urban form and structure features among US counties. **a** Population density. **b** POI density. **c** Road density. **d** Minority segregation. **e** Income segregation. **f** UCI. **g** 2019 GDP. **h** HMI. Since population density, POI density, road density, and 2019 GDP have a large scale, we used logarithmic transformation of values. See Methods for metrics definition and data processing.

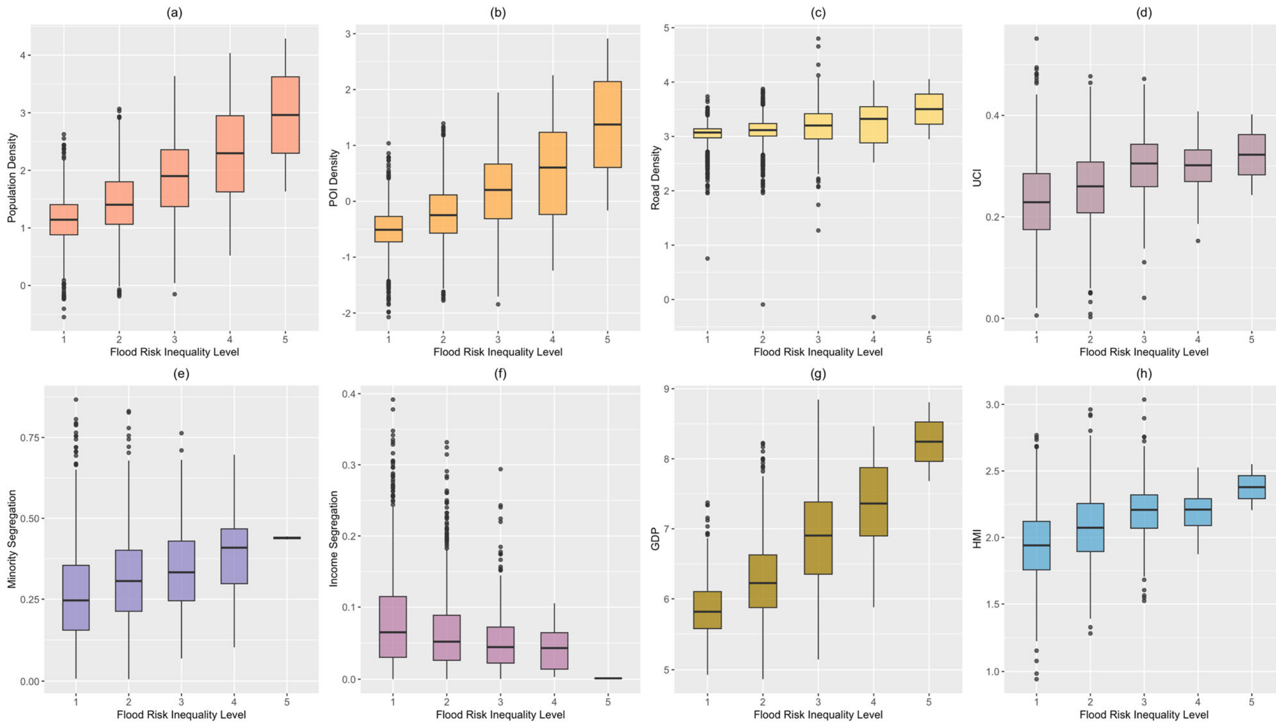


Fig. 4 | Disparities of urban form and urban structure features in different levels of spatial inequality of property flood risk. **a** Population density. **b** POI density. **c** Road density. **d** UCI. **e** Minority segregation. **f** Income segregation. **g** 2019 GDP. **h** HMI. Spatial inequality of property flood risk was categorized into five levels: 0 to ≤20% (level 1, minor inequality), 20% to ≤40% (level 2, moderate inequality), 40% to ≤60% (level 3, major inequality), 60% to ≤80% (level 4, severe inequality), and 80% to ≤100% (level 5, extreme inequality).

flood risk. Figure 4 illustrates that there are notable differences in the features of urban form and urban structure among counties with different levels of spatial inequality in property flood risk. Most features exhibit a positive relationship with the extent of spatial inequality in property flood risk, indicating that higher levels of spatial inequality in property flood risk are associated with a greater extent of these features. Notably, population density, POI density, and GDP show the most pronounced relationship with spatial inequality of property flood risk. This result suggests that counties with greater population density, POI density, and GDP have a greater spatial inequality in property flood risk. Higher population density may lead to a greater concentration of people and property in flood-prone areas, increasing the disparity in property flood risk. Similarly, higher levels of POI density and GDP may indicate greater economic activity and development, which could be associated with denser development that a greater variability in property flood risk. Our observation implies that inequality cannot be simply quantified using only one urban feature due to the complex mechanisms and interaction of urban features. Hidden correlation and interactive pathways of the urban form and structure features causing in inequalities exist and remain underexplored without relying upon further methods. In addition, we ranked the top nine counties in the level five of spatial inequality in property flood risk and showed their statistics for the eight urban form and structure features. (See Supplementary Fig. 2)

Next, we analyzed the correlation between the eight urban forms and structure features and spatial inequality of property flood risk. The results shown in Fig. 5 suggest that seven features were positively correlated with spatial inequality of property flood risk, while income segregation was negatively correlated. The result is quite similar to the boxplots shown in Fig. 4. We also calculated the Kendall coefficient, Pearson coefficient, and Spearman coefficient for each feature to further explore the statistical significance of the correlation (See Supplementary Fig. 3). The detail related to the regression model and statistical test can be found in Methods.

Taking GDP as an example, the distributions of GDP and the spatial inequality of property flood risk measured by SGI are approximately normal, with rugs shown in Fig. 5g. The Kendall rank correlation reaches 0.428, the Spearman rank correlation reaches 0.602, and the Pearson correlation coefficient approaches 0.641. All measures are statistically significant with $p < 0.001$, indicating a strong positive correlation between the GDP and the extent of spatial inequality in property flood risk. Other features such as UCI and population density show a similar trend. The strong correlation between GDP, UCI, population density, and SGI servers as an important indication of the role greater economic activities, denser development, and more centralization of facilities play in spatial inequality of property flood risk in cities. That is a pronounced concentration of economic activities, population density, and facility centralization shape spatial inequality of property flood risk in the United States. The result related to income segregation reveals a reverse relationship compared with the other features. The Kendall rank correlation between income segregation and SGI reaches −0.113, the Spearman rank correlation reaches −0.169, and the Pearson correlation coefficient approaches −0.207. These measures signify a moderate negative correlation between income segregation and spatial inequality of property flood risk. This result suggests that counties with a greater income segregation have a lesser extent of spatial inequality in property flood risks.

Pathways to spatial inequality of property flood risk among US counties

In the next step, we first implemented PCA, a statistical technique used for dimensionality reduction[27], to the eight features to identify the most important components of urban form and structure that contribute to the spatial inequality of property flood risk. The PCA result is shown in Supplementary Fig. 5. The best number of principal components was selected as three, and the cumulative explained variance of 90.59% indicates that these three principal components capture a considerable amount of the variability in the original data and provide a meaningful representation of the urban form and structure.

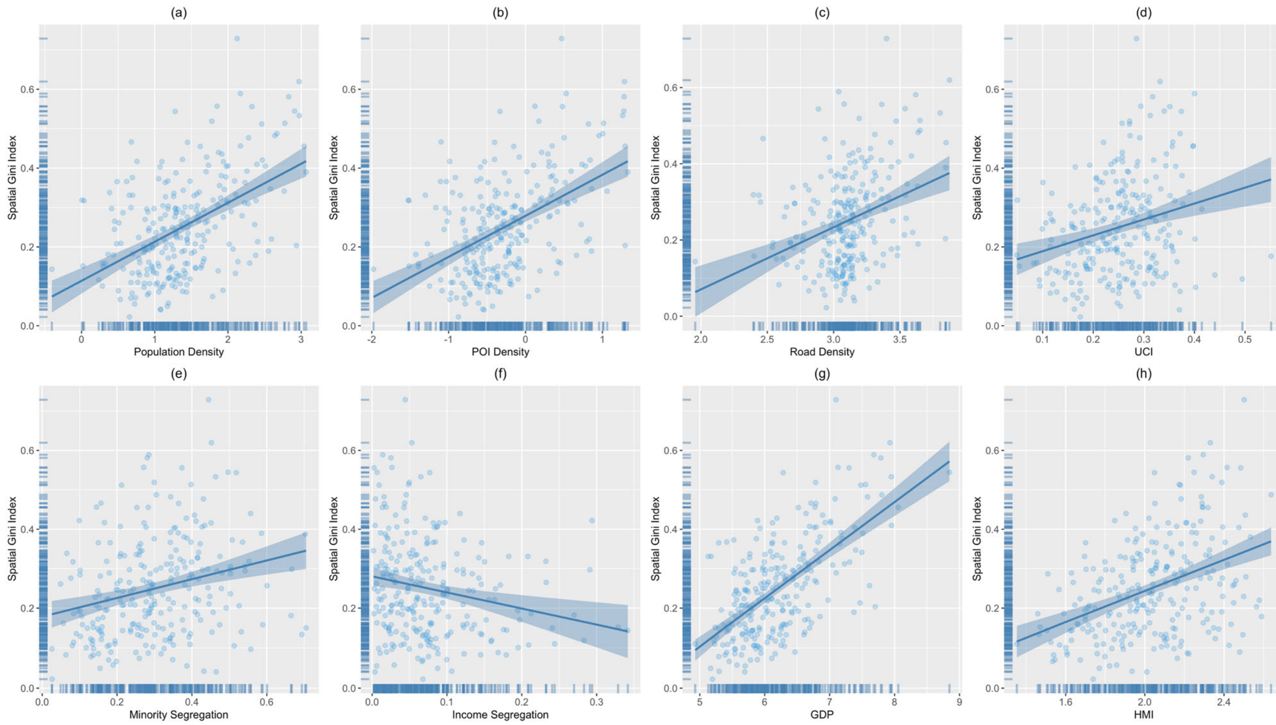


Fig. 5 | Correlation between urban form and structure features and spatial inequality of property flood risk. **a** Population density. **b** POI density. **c** Road density. **d** UCI. **e** Minority segregation. **f** Income segregation. **g** GDP. **h** HMI. The variable distribution was given by rugs. The steel blue points show the distribution of 10% raw data. We used the ordinary least squares (OLS) regression model to fit the data and the confidence interval was set to be 90%. All measures are statistically significant with $^{***}p < 0.001$.

We defined the three principal components in Table 1. The first principal component is named development density, which includes the features of population density, POI density, and road density, explaining 33.41% of the total variance. This component represents the level of urbanization and built environment density in a given area. The second component is defined as centrality and segregation, explaining 27.56% of the total variance and including the features of UCI, minority segregation, and income segregation. This component represents the level of social and economic segregation, as well as the degree of urban centralization in a given area. The third component, economic activity, explains 29.62% of the total variance and includes the features of GDP and HMI. This component represents the level of economic activity and mobility in a given area.

Upon identification and labeling of the principal components, an entropy-based CART model was implemented to identify pathways that could lead to different levels of spatial inequality of property flood risk among counties by involving the three principal components as the predictor variables and SGI levels as the response variable (See Methods for details).

Based on the model training approaches, which involved normalizing the three principal components to a [0, 1] range, performing an 80:20 train-test split, and implementing 10-fold cross-validation, the model achieved strong performance. Specifically, we measured the model's accuracy score (0.8284 for training data and 0.8178 for testing data), precision, recall, and F1 score (see Supplementary Table 3). The results show that the model performs well in accurately predicting spatial inequality of property flood risk in counties based on the pathways with combinations of principal components. In order to show as many pathways as possible and avoid overfitting at the same time, we set the minimum split leaf size to 100 counties and limit the tree depth to 7. The tree graph was generated in Fig. 6a.

In total, we obtained 14 pathways with a single strong majority category. Some tree leaves were combined because of the same pathway structure in Fig. 6a. The spatial visualization of the counties in the 14 pathways is shown in Fig. 7a. Each pathway is composed of no more than 200 counties. We eventually extracted the 14 pathways through the structure of the decision tree, some of which are shown in Fig. 6b. All the pathways can be found in Supplementary Fig. 7. Using Pathway 1 as an example, the decision tree classified 597 counties as having a minor level of spatial inequality of property flood risk. The top leaf of the tree split the principal component of development density to less than 0.442, while the second leaf split the economic activity to less than 0.432. Development density was again split in the third leaf to less than 0.296. Combining these three leaves, we derived the eventual range for Pathway 1 as [0, 0.296] for development density and [0, 0.432] for economic activity.

To classify the 14 pathways based on their levels of spatial inequality in property flood risk, three pathways were identified as minor inequality in the most notable predicted outcomes, while three were classified as moderate, two as major, two as severe, and four as extreme. For pathways with the same level of spatial inequality in property flood risk, we synthesized them and determined the range for the principal components (Fig. 7c). We also created a spatial visualization of the outcomes (Fig. 7b). Results show that the western part of the United States, as well as counties in the Coast Gulf, Florida Peninsula, Northeastern metropolitan areas, and Great Lakes Region, exhibit a higher level of severe and extreme spatial inequality of property flood risk. In contrast, the contiguous counties in the central and eastern areas exhibit relatively low spatial inequality of property flood risk. We also list example counties for all the pathways to spatial inequality of property flood risk, which can be found in Supplementary Table 4. For example, New York County in New York State, Los Angeles County in California, and Harris County in Texas are on the top of the list for the extreme level of spatial inequality in property flood risk.

Our analysis demonstrates that the principal components of development density centrality and segregation have a substantial influence on determining the extreme level of spatial inequality in property flood risk. These components exhibit a range of 0 to 1, indicating their consistent impact on the property flood risk, regardless of the specific values they hold.

Table 1 | Definition of the three principal components and their composed features

Principal component	Name	Proportion of explained variance	Main features
PC1	Development Density (DD)	33.41%	Population density POI density Road density
PC2	Centrality and Segregation (CS)	27.56%	UCI Minority segregation Income segregation
PC3	Economic Activity (EA)	29.62%	GDP HMI

On the other hand, the principal component of economic activity was found to have a range of [0.447, 1]. This observation suggests that economic activity, encompassing GDP and HMI characteristics, plays a crucial role in predicting the extreme level of spatial inequality in property flood risk. This discovery underscores the significance of considering a region's economic activity in evaluating spatial inequality of property flood risk and formulating policies to alleviate the impact of flood on the most vulnerable populations.

The minor level of spatial inequality in property flood risk is another example where we observe a relatively wide range for the three principal components. Specifically, the range for development density is [0, 0.442], which includes features such as population density, POI density, and road density. Meanwhile, the range for economic activity, which includes features such as GDP and HMI, is [0, 0.588]. This suggests that a county can be classified as having a minor level of spatial inequality of property flood risk if it satisfies both ranges for development density and economic activity. It is also important to acknowledge that centrality and segregation play a highly important role in this classification, as their range remains consistent across all counties, spanning from 0 to 1.

Overall, Fig. 7c highlights the complex relationship between different components and their contribution to the spatial inequality of property flood risk. By identifying the specific ranges for each pathway, we can better understand the factors that contribute to the spatial inequality of property flood risk and inform policy-making for more targeted and effective flood risk management.

Discussion

This study explores the relationship between urban form and structure and spatial inequality of property flood risk using rich datasets across the United States. As cities continue to grow and develop, it is essential to understand ways in which features of urban form and structure could shape spatial inequality in flood risk to properties and people. Such understanding is particularly important for devising integrated urban design strategies to address flood risk in conjunction with urban growth and development plans.

The findings from this study advance our understanding of the complex interplay between urban form and structure and spatial inequality of property flood risk in multiple important aspects. First, the findings provide empirical evidence for the presence of notable property flood risk inequality in metropolitan and coastal counties in the United States. The results indicate that the western part of the country, as well as the counties in the Coast Gulf, Florida Peninsula, the Northeastern metropolitan areas, and the Great Lakes Region, exhibit greater levels of spatial inequality in property flood risk. For example, New York County in New York State, Los Angeles County in California, and Harris County in Texas are on the top of the list for the extreme level of spatial inequality in property flood risk.

Second, the findings reveal the principal component factors and pathways related to urban form and structure that shape spatial inequality of property flood risk. Of particular importance is the identification of development density as the prominent principal component factors in explaining variations in spatial inequality of property flood risks. These findings imply that urban growth and development strategies that exacerbate development density in cities could yield serious spatial inequality of property flood risk. Increasing development density in cities exacerbates the existing flood risk hotspots[28] and yields new hotspots[29] all of which would increase spatial inequality of property flood risks. Cities with already substantial development density could primarily address spatial inequality of property flood risk by alleviating development density. Urban development strategies such as zoning regulations[30], mix-use development[31], and prioritization of open area[32] that are shown to alleviate development density in cities could also help address spatial inequality of property flood risk.

Third, the findings also provide data-driven insights regarding the tradeoff between economic development and flood risk in cities. The results of pathway analysis show that cities with moderate development density would have severe or extreme spatial inequality of property flood risk if the level of economic activity is high. The implication of this finding is that, in order for cities to pursue economic development and growth while addressing property flood risk and its spatial inequality, it is important to focus on controlling development density. The combination of dense development and high economic activity would be a pathway for severe or extreme property flood risk in cities.

Fourth, these findings underscore the importance of integrated urban design strategies to address complex urban issues at the intersection of development, growth, and flood risk management. Typically, urban issues are addressed separately using isolated plans and policies. This study shows that urban form and structure which is shaped by various urban growth and development features shape the spatial inequality of property flood risk. Hence, it is important for various urban plans and policies to be integrated to identify strategies that address different vexing urban issues simultaneously. Also, this study shows the importance of data-driven methods for various fields of urban planning, engineering, city science, geography, and environmental sciences in devising integrated urban design strategies based on examining heterogeneous urban features and their interactions to better understand ways different urban features (individually and collectively) shape different outcomes related to sustainability, flood risk, and other urban phenomena.

The contributions of this study inform researchers and practitioners in various fields including urban planning, engineering, city science, and flood risk management about the interplay between urban form and structure features and spatial inequality of property flood risk and also show new avenues for future research directions. For example, based on the findings of this study, future research can investigate causal relationships between features of urban form and structure and property flood risks in cities. For example, causal inference techniques based on spatial deep learning can be adopted. Using such models, future scenarios of urban growth and development and their effects on property flood risk and its spatial heterogeneity could be investigated.

Methods

Definition and data for spatial inequality of property flood risk

Spatial inequality of property flood risk refers to the uneven distribution of flood risk across different geographic areas. It reflects the extent to which certain areas exhibit a higher likelihood of experiencing flood damage to properties than others. This discrepancy arises from a complex interplay of factors, encompassing physical, environmental, and socio-economic elements. While it is widely accepted that various factors, such as hydrology

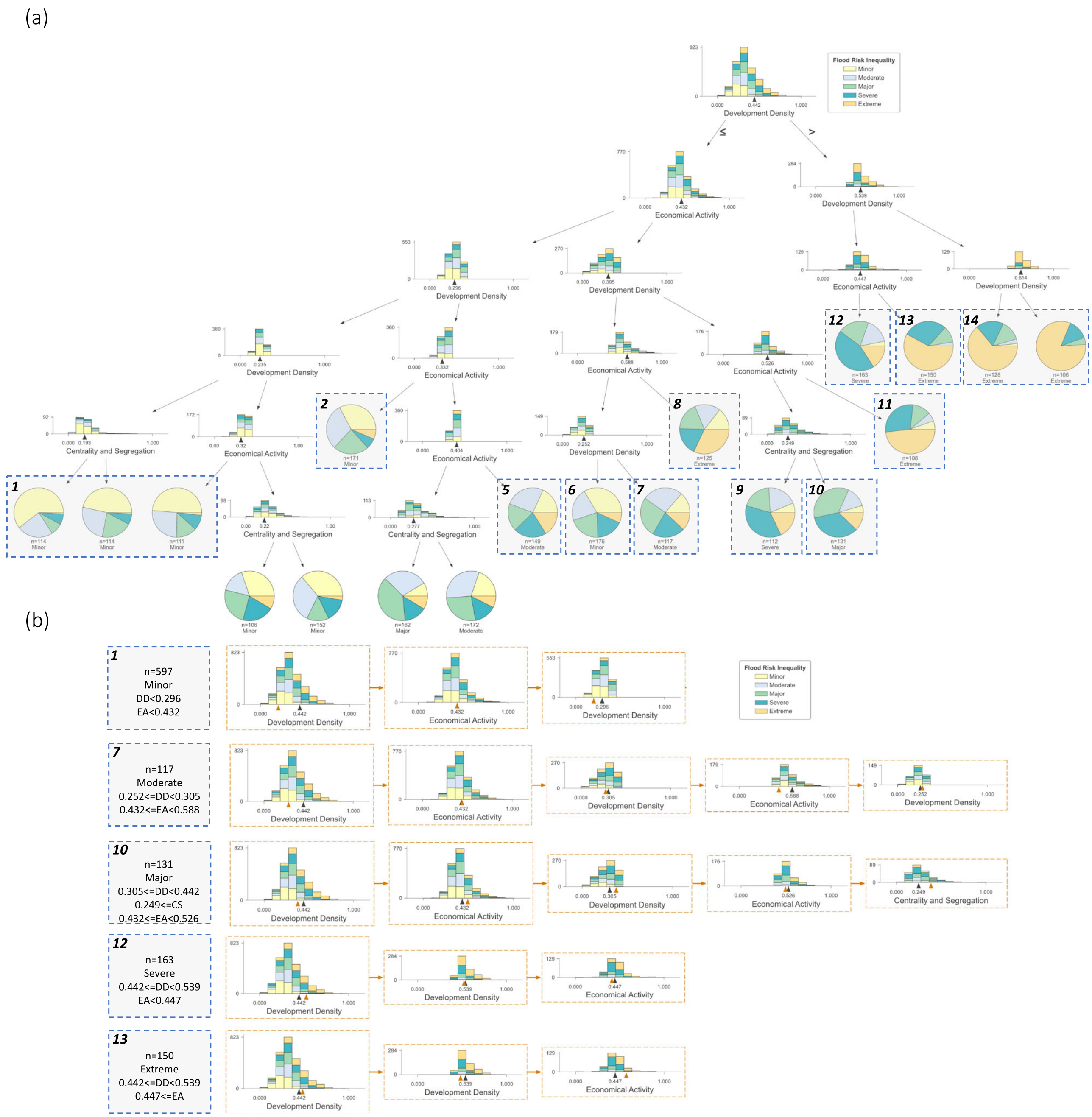
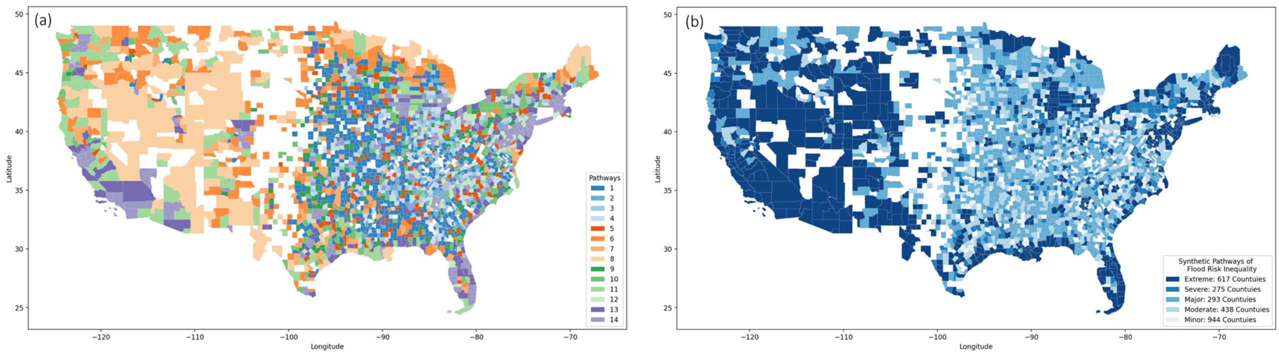


Fig. 6 | CART decision tree in terms of different levels of spatial inequality in property flood risk. a The completed CART decision tree with pathways marked in the terminal leaves. **b** Example pathways for each level of spatial inequality in property flood risk. The histogram shows the feature space distribution of the five levels of spatial inequality in property flood risk. We use a stacked histogram so that overlap is clear in the feature space between samples with different target classes. The feature space of a left child is everything to the left of the parent's split point in the same feature space, similarly for the right child. The height in the Y axis of the stacked histogram is the total number of samples from all classes and multiple class counts are stacked on top of each other. The black wedge highlights the split point and identifies the exact split value. The orange wedge highlights the split direction. The category with the largest area in the pie chart is considered to be the result of the classification prediction. "*n*" represents the number of counties in the pathway. EA, DD, and CS represent economic activity, development density, and centrality and segregation, respectively.

and land topography, influence flood risk[8–10], it is crucial to emphasize that our primary focus in this paper is on how urban form and structure impact this inequality. The significance of urban form and structure in shaping the spatial distribution of flood risk in cities is increasingly gaining recognition[11,12].

The raw dataset of property flood risk was obtained from Flood Factor Score, a model created by First Street Foundation[33]. The Flood Factor model assesses the flood risk of every property in an area and assigns it a risk score from 1 to 10. A score of 1 indicates a low chance of flooding within the next 30 years; a score of 10 indicates a high chance of flooding. The methodology employed by the Flood Factor model involves the integration of property-level data (i.e., elevation, precipitation, environmental changes over time, and community protection initiatives), overlaying building footprints, and applying flood hazard layers to calculate the probability of the maximum depth of floodwater reaching a given property. It's important to note that this dataset encompasses all types of properties in a given area. Also, our dataset has already taken the variations of various types of properties into account during the calculations, incorporating factors such as property type,



(c)

No.	Flood Risk Inequality	Pathways	Principal Componets		
			Development Density	Centrality and Segregation	Economical Activity
1	Minor	1, 2, 6	[0, 0.442)	[0, 1]	[0, 0.588)
2	Moderate	4, 5, 7	[0.252, 0.442)	[0.277, 1]	[0.404, 0.588)
3	Major	3, 10	[0.296, 0.442)	[0, 1]	[0.332, 0.404) [0.432, 0.526)
4	Severe	9, 12	[0.305, 0.539)	[0, 0.249)	[0, 0.526)
5	Extreme	8, 11, 13, 14	[0, 1]	[0, 1]	[0.447, 1]

Fig. 7 | Spatial visualization of the pathways and their range regrading to the principal components. a Spatial distribution for the original fourteen pathways. **b** Spatial distribution for the five synthetic pathways. **c** Range values of the principal components in the five synthetic pathways.

year of construction, structure, height, and whether the property is situated within a floodplain[33].

SGI is a measure of spatial inequality. It is calculated as the area difference between a perfectly equal distribution and the actual distribution[34,35]. An SGI of 0 represents a perfectly equal distribution of the feature of interest and 1 would describe a distribution where different areas have different values of the feature of interest[34]. In this study, SGI captures the spatial heterogeneity of property flood risk in a county. We calculated the percentage of property flood risk for each county by dividing the number of properties with Flood Factor score larger than six to the total number of properties, denoted here by x_i. The entropy-based SGI is given by[36]:

$$SGI = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\left|x_i - x_j\right| + (1 - w_{ij})\left|x_i - x_j\right|}{2n^2\langle x\rangle} \tag{1}$$

where N is the number of neighborhoods, and $\langle x\rangle = \frac{1}{N}\sum_i x_i$ is the mean of the variable of interest. The spatial weight w_{ij} is defined according to the adjacency matrix A where $w_{ij} = 1$ if two areas are neighbors, and 0 otherwise. The diagonal elements $w_{ii} = 0$ as defined in A and W corresponds to the sum of all weights.

Definition and data for urban form features

GDP. To estimate the status of the economic development of the county, we adopted the 2019 data of gross domestic product for each county. The data are provided by the Bureau of Economic Analysis in the US Department of Commerce[37].

Population density. The population size was obtained from the 2020 race and ethnicity data from US Census Bureau[38]. We calculated the population density at the county level by dividing the total population of the county by its land area. Land area data was also obtained from the US Census Bureau[39].

Minority segregation and income segregation. Urban segregation refers to the physical and social separation of different racial, ethnic, and socioeconomic groups within a city[40]. This separation can take many forms, including minority segregation and income segregation. One of the key consequences of urban segregation is that it often leads to unequal distribution of resources, as well as increased exposure to environmental hazards such as flooding[41,42].

In this study, we adopted the Dissimilarity Index (DI) to evaluate minority segregation and income segregation. The DI is a measure of spatial segregation that indicates the extent to which two groups are evenly distributed across different areas, which ranges from 0 (indicating perfect evenness) to 1 (indicating complete separation)[43,44]. We calculated the DI based on the proportion of minority population (for minority segregation) and the proportion of low-income population (for income segregation) at the census tract level relative to the county level[45]:

$$DI = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{x_i}{X} - \frac{y_i}{Y}\right| \tag{2}$$

where x_i is the minority population (or low-income population) in the smaller geographical unit; X is the minority population (or low-income population) in the larger geographical unit. y_i is the reference population in the smaller geographical unit; Y is the reference population in the larger geographical unit. In our study, smaller geographical unit refers to census tract level and the larger geographical unit refers to county level.

For minority segregation, we collected the racial population data from the 2020 race and ethnicity data from the US Census Bureau[38]. The primary racial groups in this study are non-Hispanic White, non-Hispanic Black, and non-Hispanic Asian residents. Non-Hispanic populations are selected because White, Black, or Asian populations can be mutually selective from Hispanic populations. The method of this kind of analysis is consistent with those frequently adopted by high-impact research works[25,46,47]. We considered the non-Hispanic Black, and non-Hispanic Asian as the minority population and non-Hispanic White as the reference population.

For income segregation, we extracted median income data from the 2020 American Community Survey[48]. This study used the 5-year estimates of median income due to the broader coverage of areas, larger sample size, and higher precision, making the data more reliable than 1-year and 3-year estimates. We used the quantile income groups of a county (Q1 to Q4) to indicate income levels with Q1/Q2 representing low-income groups and Q3/Q4 represent high-income groups, respectively.

Definition and data for urban structure features

POI density. To capture the distribution of physical facilities, we adopted the 6.5 million active POI data in the US from SafeGraph[49]. The dataset includes basic information about POIs, such as POI IDs, location names, geographical coordinates, addresses, brands, and North American Industry Classification System (NAICS) codes to categorize POIs. The NAICS code is the standard used by federal statistical agencies in classifying business establishments[50]. In this study, we selected ten essential types of POIs that are closely relevant to human daily lives: restaurants, schools, grocery stores, churches, gas stations, pharmacies and drug stores, banks, hospitals, parks, and shopping malls. We counted the number of POIs in each county and calculated their density as their facility distribution feature.

Road density. To capture the distribution of transportation network, we extracted data from Open Street Map[51] to calculate the density of road segments in counties. We estimated complete road networks from the raw data by assembling road segments. Since the lengths of road segments created by the source were in close proximity, we calculated road density by dividing the number of road segments by the areas of a county.

Urban centrality index. We adopted UCI to characterize the centralization degree of the facilities in a county. UCI is the product of the local coefficient and the proximity index[52]. The local coefficient was computed based on the number of POIs within each census tract; the proximity index was computed based on the number of POIs within each census tract along with a distance matrix that considered the distance between census tracts. The value of UCI ranges from 0 to 1. The values close to 0 indicate polycentric distribution of facilities within a county, while the values close to 1 indicate monocentric distribution of facilities. The indices are formulated as follows[52]:

$$LC = \frac{1}{2}\sum_{i=1}^{N}(k_i - \frac{1}{N}) \tag{3}$$

$$PI = 1 - \frac{V}{V_{\max}} \tag{4}$$

$$V = K' \times D \times K \tag{5}$$

where N is the total number of census tracts in a county; K is a vector of the number of POIs in each census tract; k_i is a component of the vector K; D is the distance matrix between census tracts; V_{max} is calculated by assuming that the total POIs are uniformly settling on the boundary of the county; LC is the local coefficient, which measures the unequal distribution; PI is the proximity index, which resolves the normalization issue; V is the Venables Index.

Human mobility index. To understand the inequality of population activities, we employed mobile phone data from Spectus Inc. to develop the metric of HMI. The data has a wide set of attributes, including anonymized user ID, latitude, longitude, POI ID, time of observation, and the dwelling time of each visit[53]. Prior studies found that Spectus mobile phone data is representative to describe human activities and mobility[54–56]. Hence, the feature generated using the dataset should be representative and valid for our analyses. We extracted the data from April 2019 (28 days) to account for the variation of population activities on weekdays and weekends. Our period is also during regular conditions when no external extreme events perturbed human activities. To develop the HMI, we first assigned each visit point v_i to a defined CBG in a county. Then, we calculated HMI as follows:

$$HMI = \frac{\sum_{i=1}^{n} v_i}{28n} \tag{6}$$

where n denotes the number of CBGs in a county.

We finally mapped the values of HMI to the range from 0 to 1 using min-max scaling. The proximity of HMI values to 0 or 1 indicates the level of human mobility and activity, with values closer to 0 indicating lower activity and values closer to 1 indicating higher activity in a county.

Statistical analysis

Ordinary least squares regression model. We employed an ordinary least squares regression model to capture the relationships between urban form and structure and spatial inequality of property flood risk among counties and to understand the relative importance of each feature[57]:

$$\begin{aligned} y_i \sim \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} \\ + \beta_5 x_{i,5} + \beta_6 x_{i,6} + \beta_7 x_{i,7} + \beta_8 x_{i,8} + \varepsilon_i \end{aligned} \tag{7}$$

where, y_i is the SGI of county i; $x_{i,1}$–$x_{i,8}$ are the features of urban form and structure; β are coefficients; ε_i is the error term.

In the regression, since the values of POI density, population density, road density, and GDP have a much larger scale than other variables, we used logarithmic transformation of values. Three statistical tests, Kendall's tau test, Pearson's correlation test, and Spearman's rank correlation test were then conducted for the correlation analyses to examine statistical significance and determine feature importance.

Classification and regression tree model. The CART model is an unsupervised machine learning algorithm used to build a decision tree by recursively splitting the data based on the predictor variables to minimize the entropy in the response variable[58]. The decision tree consists of a series of nodes, each representing a split in the data based on a particular predictor variable, and terminal nodes representing the predicted response variable for a given combination of predictor variable values. The method to identify the best splits is to minimize the entropy. If the entropy of the two child nodes is not lower than that of a parent node, splitting will be terminated any further. The entropy (E) is given by[59]:

$$E = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{8}$$

where p_i is the fraction of items in the class i.

In this study, we categorized the SGI into five levels: 0 to ≤20% (minor inequality), 20% to ≤40% (moderate inequality), 40% to ≤60% (major inequality), 60% to ≤80% (severe inequality), and 80% to ≤100% (extreme inequality). Then, we implemented a CART classification algorithm using the principal components as the predictor variables and SGI levels as the response variable.

Decision trees were utilized to pinpoint the factors that shape pathways to different levels of spatial inequality of property flood risk among different counties. By analyzing the decision trees and the pathways they present, this study aims to shed light on the contributing factors to the spatial inequality of property flood risk in the United States. Our primary objective is to uncover as many pathways as possible while maintaining good performance, considering the balance between complexity and performance. To achieve this, we set the minimum split leaf size to 100 counties and limit the tree depth to seven, enabling us to generate more pathways while also limiting the tree depth to prevent overfitting.

Data availability

All data were collected through a CCPA- and GDPR-compliant framework and utilized for research purposes. The datasets of POI, human mobility, and Flood Factor scores that support the findings of this study are available from SafeGraph Inc., Spectus Inc., and First Street Foundation respectively, but restrictions apply to the availability of these datasets, which were used under license for the current study. The datasets can be accessed upon request submitted on safegraph.com, spectus.ai, and firststreet.org, respectively. Other data (GDP[37], population data[38], Land area data[39], median income data[48], NAICS code[50], Open Street road segments[51]) used in this study are all publicly available.

Code availability

The code that supports the findings of this study is available from the corresponding author upon request.

Received: 14 June 2023; Accepted: 21 March 2024;
Published online: 01 April 2024

References

1. Hino, M. & Nance, E. Five ways to ensure flood-risk research helps the most vulnerable. *Nature* **595**, 27–29 (2021).
2. Leppold, C., Gibbs, L., Block, K., Reifels, L. & Quinn, P. Public health implications of multiple disaster exposures. *Lancet Pub. Health* **7**, 274–286 (2022).
3. Smith, A. B. & Katz, R. W. US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Nat. Hazards* **67**, 387–410 (2013).
4. Hong, B., Bonczak, B. J., Gupta, A. & Kontokosta, C. E. Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nat. Commun.* **12**, 1870 (2021).
5. McCaughey, J. W., Daly, P., Mundir, I., Mahdi, S. & Patt, A. Socio-economic consequences of post-disaster reconstruction in hazard-exposed areas. *Nat. Sustain.* **1**, 38–43 (2018).
6. Nohrstedt, D., Mazzoleni, M., Parker, C. F. & Di Baldassarre, G. Exposure to natural hazard events unassociated with policy change for improved disaster risk reduction. *Nat. Commun.* **12**, 193 (2021).
7. Gourevitch, J. D. et al. Unpriced climate risk and the potential consequences of overvaluation in US housing markets. *Nat. Clim. Ch.* **13**, 250–257 (2023).
8. Tate, E., Rahman, M. A., Emrich, C. T. & Sampson, C. C. Flood exposure and social vulnerability in the United States. *Nat. Hazards* **106**, 435–457 (2021).
9. Edmonds, D. A., Caldwell, R. L., Brondizio, E. S. & Siani, S. M. Coastal flooding will disproportionately impact people on river deltas. *Nat. Commun.* **11**, 4741 (2020).
10. Hauer, M. E. et al. Assessing population exposure to coastal flooding due to sea level rise. *Nat. Commun.* **12**, 6900 (2021).
11. Hao, H. & Wang, Y. Disentangling relations between urban form and urban accessibility for resilience to extreme weather and climate events. *Landsc. Urb. Plan.* **220**, 104352 (2022).
12. Puzyreva, K. et al. Professionalization of community engagement in flood risk management: Insights from four European countries. *Int. J. Disaster Risk Reduct.* **71**, 102811 (2022).
13. Balland, P.-A. et al. Complex economic activities concentrate in large cities. *Nat. Hum. Behav.* **4**, 248–254 (2020).
14. Niu, T., Chen, Y. & Yuan, Y. Measuring urban poverty using multi-source data and a random forest algorithm: a case study in Guangzhou. *Sustain. Cities Soc.* **54**, 102014 (2020).
15. Wang, J., Kuffer, M., Roy, D. & Pfeffer, K. Deprivation pockets through the lens of convolutional neural networks. *Remote Sens. Environ.* **234**, 111448 (2019).
16. Esmalian, A., Wang, W. & Mostafavi, A. Multi-agent modeling of hazard–household–infrastructure nexus for equitable resilience assessment. *Comput. Aided Civ. Infrastruct. Eng.* **37**, 1491–1520 (2022).
17. Pan, W., Ghoshal, G., Krumme, C., Cebrian, M. & Pentland, A. Urban characteristics attributable to density-driven tie formation. *Nat. Commun.* **4**, 1961 (2013).
18. Zhang, X. & Li, N. Characterizing individual mobility perturbations in cities during extreme weather events. *Int. J. Disaster Risk Reduct.* **72**, 102849 (2022).
19. Xu, Y., Olmos, L. E., Abbar, S. & González, M. C. Deconstructing laws of accessibility and facility distribution in cities. *Sci. Adv.* **6**, eabb4112 (2020).
20. Patrascu, F. I. & Mostafavi, A. Spatial model for predictive recovery monitoring based on hazard, built environment, and population features and their spillover effects. *Environ. Plan. B.* **51**, 39–56 (2023).
21. Esmalian, A., Coleman, N., Yuan, F., Xiao, X. & Mostafavi, A. Characterizing equitable access to grocery stores during disasters using location-based data. *Sci. Rep.* **12**, 20203 (2022).
22. Collins, T. W., Grineski, S. E., Chakraborty, J. & Flores, A. B. Environmental injustice and Hurricane Harvey: a household-level study of socially disparate flood exposures in Greater Houston, Texas, USA. *Environ. Res.* **179**, 108772 (2019).
23. Smiley, K. T. et al. Social inequalities in climate change-attributed impacts of Hurricane Harvey. *Nat. Commun.* **13**, 3418 (2022).
24. Qiang, Y. Disparities of population exposed to flood hazards in the United States. *J. Environ. Manag.* **232**, 295–304 (2019).
25. Liu, T. & Fan, C. Impacts of disaster exposure on climate adaptation injustice across US cities. *Sustain. Cities Soc.* **89**, 104371 (2023).
26. Kubal, C., Haase, D., Meyer, V. & Scheuer, S. Integrated urban flood risk assessment–adapting a multicriteria approach to a city. *Nat. Hazards Earth Syst. Sci.* **9**, 1881–1895 (2009).
27. Abdi, H. & Williams, L. J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2**, 433–459 (2010).
28. Jian, W. et al. Evaluating pluvial flood hazard for highly urbanised cities: a case study of the Pearl River Delta Region in China. *Nat. Hazards* **105**, 1691–1719 (2021).
29. Wolff, C., Nikoletopoulos, T., Hinkel, J. & Vafeidis, A. T. Future urban development exacerbates coastal exposure in the Mediterranean. *Sci. Rep.* **10**, 1–11 (2020).
30. Cutter, S. L., Emrich, C. T., Gall, M. & Reeves, R. Flash flood risk and the paradox of urban development. *Nat. Hazards Rev.* **19**, 05017005 (2018).
31. Su, W., Ye, G., Yao, S. & Yang, G. Urban land pattern impacts on floods in a new district of China. *Sustainability* **6**, 6488–6508 (2014).
32. Pallathadka, A., Sauer, J., Chang, H. & Grimm, N. B. Urban flood risk and green infrastructure: who is exposed to risk and who benefits from investment? a case study of three US Cities. *Landsc. Urb. Plan.* **223**, 104417 (2022).
33. First Street Foundation. *First Street Foundation Flood Model* https://firststreet.org/risk-factor/ (2022).
34. Rey, S. J. & Smith, R. J. A spatial decomposition of the Gini coefficient. *Lett. Spat. Resour. Sci.* **6**, 55–70 (2013).
35. Sousa, S. & Nicosia, V. Quantifying ethnic segregation in cities through random walks. *Nat. Commun.* **13**, 5809 (2022).
36. Coleman, N. et al. Energy inequality in climate hazards: empirical evidence of social and spatial disparities in managed and hazard-induced power outages. *Sustain. Cities Soc.* **92**, 104491 (2023).
37. US Department of Commerce. *Gross Domestic Product by County* https://www.bea.gov/news/2020/gross-domestic-product-county-2019 (2021).
38. US Census Bureau. *Hispanic or Latino, and not Hispanic or Latino by race. Census Bureau Data* https://data.census.gov/cedsci/ (2020).
39. US Census Bureau. *USA Counties: 2011* https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html#LND (2011).
40. Li, Q.-Q., Yue, Y., Gao, Q.-L., Zhong, C. & Barros, J. Towards a new paradigm for segregation measurement in an age of big data. *Urb. Inform.* **1**, 5 (2022).
41. Martines, M. R. et al. Spatial segregation in floodplain: an approach to correlate physical and human dimensions for urban planning. *Cities* **97**, 102551 (2020).

42. Moro, E., Calacci, D., Dong, X. & Pentland, A. Mobility patterns are associated with experienced income segregation in large US cities. *Nat. Commun.* **12**, 4633 (2021).
43. Massey, D. S. & Denton, N. A. The dimensions of residential segregation. *Soc. Forces* **67**, 281–315 (1988).
44. Lichter, D. T., Parisi, D., Grice, S. M. & Taquino, M. C. National estimates of racial segregation in rural and small-town America. *Demography* **44**, 563–581 (2007).
45. Kodros, J. K. et al. Unequal airborne exposure to toxic metals associated with race, ethnicity, and segregation in the USA. *Nat. Commun.* **13**, 6329 (2022).
46. Jbaily, A. et al. Air pollution exposure disparities across US population and income groups. *Nature* **601**, 228–233 (2022).
47. Mehta, N. K., Lee, H. & Ylitalo, K. R. Child health in the United States: recent trends in racial/ethnic disparities. *Soc. Sci. Med.* **95**, 6–15 (2013).
48. US Census Bureau. *Income in the past 12 months (in 2020 Inflation-Adjusted Dollars)* https://data.census.gov/cedsci/ (2020).
49. SafeGraph https://www.safegraph.com/ (2023).
50. US Census Bureau. *North American Industry Classification System* https://www.census.gov/naics/ (2023).
51. Open Street Map https://www.openstreetmap.org (2023).
52. Pereira, R., Nadalin, V., Monasterio, L. & Albuquerque, P. Urban centrality: a simple index. *Geogr. Anal.* **45**, 77–89 (2013).
53. Spectus https://spectus.ai/ (2023).
54. Aleta, A. et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4**, 964–971 (2020).
55. Fan, C., Xu, J., Natarajan, B. Y. & Mostafavi, A. Interpretable machine learning learns complex interactions of urban features to understand socio-economic inequality. *Comput. Aided Civ. Infrastruct. Eng.* **38**, 2013–2029 (2023).
56. Wang, F., Wang, J., Cao, J., Chen, C. & Ban, X. J. Extracting trips from multi-sourced data for mobility pattern analysis: An app-based data example. *Transp. Res. Part C* **105**, 183–202 (2019).
57. Craven, B. & Islam, S. M. Ordinary least-squares regression. In *The SAGE dictionary of quantitative management research.* (eds. Hutcheson, G. D. & Moutinho, L. A. M.) 224–228 (SAGE, London, 2011).
58. Lewis, R. An introduction to classification and regression tree (CART) analysis. (2000).
59. Dargin, J. & Mostafavi, A. Dissecting heterogeneous pathways to disparate household-level impacts due to infrastructure service disruptions. *Int. J. Disaster Risk Reduct.* **83**, 103351 (2022).
60. US Census Bureau. *Coastline Counties of the United States by Coastline Region* https://www.census.gov/library/visualizations/2008/demo/coastline-countries-of-the-united-states-by-coastline-region.html (2008).
61. US Census Bureau. *Metropolitan and Micropolitan* https://www.census.gov/programs-surveys/metro-micro.html (2023).

Acknowledgements
This material is based in part upon work supported by the National Science Foundation under CRISP 2.0 Type 2 No. 1832662 grant. The authors also would like to acknowledge the data support from SafeGraph, Spectus, and First Street Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, SafeGraph, Spectus, or First Street Foundation.

Author contributions
J.M.: Conceptualization, Methodology, Data curation, Formal analysis, Writing—Original draft. A.M.: Conceptualization, Methodology, Writing—Reviewing and Editing, Supervision, Funding acquisition.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-024-01337-3.

Correspondence and requests for materials should be addressed to Ali Mostafavi.

Peer review information *Communications Earth & Environment* thanks Seung Kyum Kim and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Martina Grecequet. A peer review file is available.