



EXPLORING FLOOD VULNERABILITY PATTERNS AND RESIDENTIAL FLOOD IMPACTS IN URBAN ENVIRONMENT

Hyun Jeong Koo¹ and Vassiliki Demetracopoulou²

¹Wayne State University, United States of America

²Rutgers University

Abstract

This study analyzes FEMA flood insurance claims and Zillow housing data in Detroit, United States, using ZIP-code-level metrics and damage-based quartiles to identify neighborhood-scale spatial patterns. Results show statistically significant differences in housing price and claim frequency across damage quartiles, with higher-damage ZIP codes characterized by lower housing values and more frequent flood claims. Housing characteristics showed moderate correlations with flood damage but were insufficient predictors in regression models. By jointly examining flood damage and housing characteristics at the neighborhood scale, this study advances understanding of residential flood vulnerability and provides evidence to support more informed planning and mitigation efforts.

Introduction

Urban flooding has increased in frequency and severity in recent decades, driven by climate change, urbanization, and aging infrastructure, with wide-ranging impacts on transportation systems, utilities, and housing. In the United States, older urban centers are particularly vulnerable due to legacy stormwater systems, high impervious surface coverage, and development patterns that predate modern flood management standards (US EPA, 2021; ASCE, 2025). These conditions contribute to recurrent flood damage, including roadway inundation, service disruptions, and residential property losses.

Vulnerable and low-income communities are often disproportionately affected by urban flooding, as housing characteristics such as older structures, basement units, and limited floodproofing intersect with a reduced capacity to afford repairs or mitigation measures (Cutter et al., 2003; Wing et al., 2022). As a result, urban flooding represents not only an infrastructure resilience challenge but also a mechanism through which existing social and economic inequities are intensified.

Despite increasing attention to urban flooding and social vulnerability, empirical research remains limited on how flooding impacts intersect with housing and residential characteristics at the neighborhood scale. This study addresses this gap by examining relationships between flooding indicators and housing characteristics in Detroit, Michigan, using FEMA flood insurance claims and

Zillow housing data. Detroit provides a critical case for analyzing urban flooding dynamics due to the combined effects of aging infrastructure, extensive impervious surfaces, climate-driven precipitation extremes, and persistent socioeconomic vulnerability. By jointly analyzing flooding indicators, such as total flood damage derived from FEMA claims, and housing characteristics, such as average housing price and built year obtained from Zillow, this research advances neighborhood-scale understanding of urban flood vulnerability and supports more equitable flood mitigation and housing resilience strategies.

Literature review

Urban flooding is primarily driven by short-term or continuous rainfall that exceeds the capacity of the drainage system (Liu et al., 2020; Qi et al., 2020). Unlike traditional flood hazards, urban flooding is strongly shaped by land-use patterns, impervious surface coverage, drainage system design, and infrastructure condition. Climate change has further intensified urban flooding by increasing the frequency and magnitude of short-duration, high-intensity precipitation events (IPCC, 2021). Numerous studies have documented rising flood-related damages in U.S. cities, including transportation disruptions, property losses, and repeated service outages (Wing et al., 2022).

A substantial body of literature has demonstrated that flood impacts are unevenly distributed across populations. Social vulnerability theory emphasizes that socioeconomic characteristics, such as income, race, age, and housing tenure influence exposure, sensitivity, and recovery capacity during hazard events (Cutter et al., 2003; Cutter et al., 2008). In the context of flooding, lower-income and marginalized communities often face higher exposure to flood hazards and experience greater difficulty recovering from flood-related losses.

Recent large-scale studies have strong evidence of systematic inequities in flood risk and damage across the United States. Wing et al. (2022) showed that socially vulnerable populations are disproportionately exposed to flood risk, even after accounting for physical flood hazard condition. Kousky (2018) further noted that limited access to insurance, savings, and credit constrains post-flood recovery, leading to cumulative and compounding

losses over time. These inequities are particularly pronounced in cities such as Detroit, where historical disinvestment and infrastructure decline intersect with contemporary climate risks.

Housing characteristics play a critical role in shaping flood vulnerability, yet they remain underexplored relative to broader socioeconomic indicators. Prior research has shown that older housing stock is often more susceptible to flood damage due to aging materials, basement configurations, and limited incorporation of flood-resilient design features (Elliott & Clement, 2017; Tate et al., 2021). Older homes are also more likely to be served by combined sewer systems, increasing exposure to sewer backups and surface flooding during heavy rainfall events.

Housing value has been identified as an important indicator for both structural condition and adaptive capacity. Lower-value housing is often associated with deferred maintenance, limited floodproofing, and reduced access to resources for mitigation or repair (FEMA, 2023; Kousky, 2018). Several studies have documented higher relative flood losses in lower-income or lower-value housing areas, even when controlling for hazard intensity (Wing et al., 2022; Shokry et al., 2022). However, much of this work relies on modeled flood exposure or survey-based vulnerability indices rather than observed damage data.

Despite growing recognition of the role of housing in flood vulnerability, relatively few studies have empirically linked observed flood damage with neighborhood-scale housing characteristics using claims-based data. Existing research often focuses on floodplain exposure, hydrologic modeling, or county-level socioeconomic indicators, which may obscure localized patterns of residential vulnerability within cities (Tate et al., 2021). Moreover, while FEMA’s National Flood Insurance Program (NFIP) data have been used extensively to study flood losses, they are less frequently combined with independent housing datasets to examine structural and economic housing attributes at spatial scales.

This study addresses this gap by examining how observed flood losses relate to housing characteristics in a legacy urban context, using FEMA flood insurance claims and housing data in the City of Detroit.

Research methodology

To achieve the goal of this study, two primary datasets were used: (1) FEMA National Flood Insurance Program (NFIP) residential claims for the City of Detroit, which provide information on total building and contents losses for flood-affected properties, and (2) publicly available housing data from Zillow that include parcel-level characteristics such as built year, estimated housing value, and selected structural attributes. These datasets were selected because they offer complementary perspectives. NFIP claims capture observed flood damage and loss outcomes, while Zillow housing data provides descriptive indicators of neighborhood housing stock, including

housing prices and the year of construction. The Zillow housing data were obtained in 2024.

Both datasets underwent systematic preprocessing before analysis. NFIP records were screened to remove entries with incomplete flood damage information. Housing data were similarly cleaned by excluding parcels with missing year built, estimated value, or location attributes. Duplicate records, unmatched parcels, and entries lacking ZIP code identifiers were removed to ensure consistency and data quality across datasets.

After preprocessing, the two datasets were spatially joined using ZIP code identifiers, enabling aggregation of parcel-level housing attributes with ZIP-level flood damage data (Table 1). This process produced a unified dataset suitable for neighborhood-scale analysis while retaining comparability of housing characteristics (e.g., year built and housing value) across geographic areas. ZIP codes were then grouped into quartiles based on total flood damage, representing increasing levels of flood impact. This quartile-based stratification enabled systematic comparisons across damage-intensity groups and supported subsequent statistical analyses.

Table 1: Zip-code-level variables dataset

Variables	Definition	Source
Total Flood Damage	Sum of building and contents losses for claims in each ZIP Code	FEMA
Number of Losses	Count of residential claims per ZIP code	FEMA
Water Level	Average reported flood water depth associated with residential claims	FEMA
Average Housing Price	Mean Zillow housing value estimate per Zip code	Zillow
Average Built Year	Mean built year of residential parcels per ZIP code	Zillow
Average Frontage Length	Mean frontage length for parcels in each Zip code	Zillow

The study employed a multi-step analytical framework to examine relationships between neighborhood housing characteristics and flood losses. First, correlation analysis quantified linear associations between flood damage and key variables, including water level, housing price, and year built. Second, one-way analysis of variance (ANOVA) tests assessed whether housing characteristics differed significantly across four damage-based quartile groups. Third, regression models were estimated to evaluate the explanatory power of housing attributes in predicting total flood damage. Finally, an unsupervised machine learning approach (i.e., clustering analysis) was applied to identify data-driven neighborhood groupings based on flood damage and housing characteristics.

This integrated methodological approach supported both exploratory and inferential analyses, enabling a structured assessment of the relationships between neighborhood housing characteristics and flood impacts at the neighborhood scale.

Data analysis and results

Total flood damage and the number of NFIP residential claims varied substantially across Detroit ZIP codes, indicating pronounced spatial heterogeneity in flood impacts. Descriptive analyses indicated that ZIP codes with greater flood damage generally contained older, lower-value housing, suggesting that housing age and value may be associated with disproportionate flood impacts. These patterns were subsequently examined using correlation analysis, one-way ANOVA, regression modeling, and K-means clustering.

(1) Correlation analysis

Pearson correlation coefficients were calculated to assess the strength and direction of relationships between total flood damage and housing-related attributes (Table 2). The analysis revealed no strong linear relationship between average water level and total damage ($r = -0.148$), indicating that water depth alone does not explain ZIP-level differences in losses. This finding is consistent with prior research indicating that factors such as housing condition, age of the housing stock, drainage performance, and neighborhood infrastructure can substantially influence flood impacts beyond hydrologic intensity alone (Cutter et al., 2008; Kousky, 2018; Wing et al., 2020).

A moderate negative correlation between housing price and total damage ($r = -0.379$) indicates that lower-value housing areas experienced higher cumulative flood losses. This pattern suggests that neighborhoods characterized by older or lower-value housing may be more susceptible to flood damage, potentially due to differences in building condition, maintenance, or flood mitigation capacity.

In addition, a moderate positive correlation between year built and total damage ($r = 0.326$) indicates that areas with older housing stock tended to experience greater flood losses. Older buildings may be more vulnerable due to aging materials, basement configurations, and legacy infrastructure systems. Many Detroit neighborhoods contain pre-1950 housing and combined sewer systems, which have been associated with recurrent flooding and sewer backups. These findings highlight the importance of neighborhood housing characteristics in shaping flood impacts, alongside hydrologic factors.

Table 2: Correlation analysis results

Variables	1	2	3	4
1 Total Damage	1.00	-0.379*	0.326	-0.149
2 Housing Price		1.00	-0.109	0.330
3 Built Year			1.00	-0.422*
4 Water Level				1.00

*Significant at alpha level of 0.05.

(2) Comparative analysis across damage quartiles

To assess whether neighborhood housing characteristics differed systematically across damage levels, a one-way

analysis of variance (ANOVA) was conducted comparing four ZIP code quartiles defined by total flood damage (Table 3). The results indicated statistically significant differences in average housing price across quartiles ($F(3, 26) = 4.89$, $p = 0.0079$). ZIP codes in the highest-damage quartile exhibited the lowest average housing values, reinforcing the correlation results and suggesting that economically vulnerable neighborhoods experience disproportionately high flood impacts.

There were also statistically significant differences in the number of losses per ZIP code ($F(3, 26) = 8.49$, $p = 0.0004$), with higher-damage quartiles experiencing more frequent claims. This suggests that repeated events have contributed to cumulative neighborhood-level damage patterns.

In contrast, several variables did not exhibit statistically significant variation across damage quartiles. Average water level ($p = 0.86$), year built ($p = 0.24$), and average frontage length ($p = 0.16$) did not differ significantly among groups. The lack of significance for year built, despite the moderate correlation observed earlier, suggests that housing age may be associated with flood damage but does not clearly differentiate neighborhoods when aggregated into damage-based quartiles. The nonsignificant differences in water level further indicate that structural and socioeconomic conditions may play a larger role than hydrologic indicators alone in shaping neighborhood-scale flood impacts. In summary, the ANOVA results highlight pronounced disparities in housing value across neighborhoods experiencing different levels of flood damage. Figures 2 and 3 visually present the flood level and damages per ZIP code across Detroit.

Table 3: One-way ANOVA results

Variables	df	F	p-value
Housing Price	3	4.89	0.0079
Built Year	3	1.49	0.2403
Water Level	3	0.25	0.8558
Average frontage length	3	1.89	0.1560
Number of Losses	3	8.49	0.0004

(3) Regression Modeling

To further examine predictors of total flood damage, a multiple linear regression model was estimated using average housing price, year built, average water level, and average frontage length as independent variables. The results (table 4) indicated that none of the predictors were statistically significant at conventional levels, and the overall model explained a relatively small proportion of the variance in total flood damage ($R^2 = 0.18$, $p = 0.279$). Although average housing price and year built exhibited coefficient signs consistent with earlier correlation patterns, their effects were weak and not statistically significant. These findings indicate that, at the ZIP code scale, housing-related attributes alone provide limited

explanatory power for modeling variation in total flood damage across Detroit neighborhoods.

Table 4: Regression results

Predictor	Coefficient	Std. Error	t-value	p-value
Constant	22,243	.*	0.20	0.84
Housing Price	3.52	5.51	0.64	0.53
Built Year	-13.81	11.72	-1.18	0.25
Water Level	-28.87	63.00	-0.46	0.65
Average frontage length	-0.00	0.00	-0.44	0.66

* The standard error of the intercept is omitted because it does not provide meaningful interpretive value.

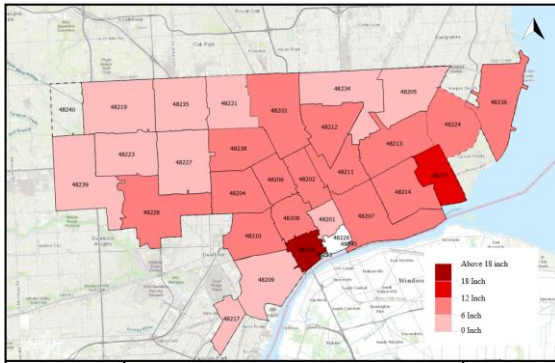


Figure 2: Detroit flood level by zip code

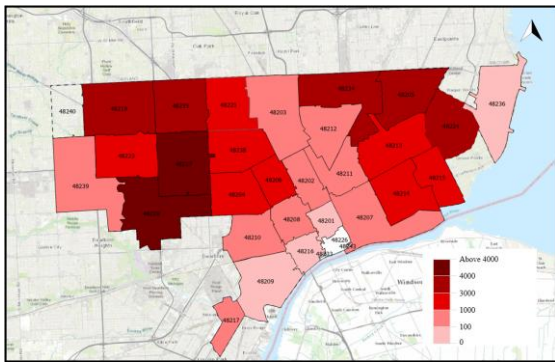


Figure 3: Detroit damage by zip code

(4) K-Means clustering analysis

To complement the quartile-based comparisons, a K-means clustering analysis ($k=3$) was conducted to identify data-driven neighborhood groupings based on flood damage and housing characteristics. Alternative values of k were explored, but $k=3$ yielded the most meaningful and interpretable group distinctions for the study objectives. Six variables were standardized and used in the analysis, to ensure that all variables contributed equally to the clustering process.

Table 5 presents the results of the clustering analysis, which identified three distinct flood-vulnerability profiles. Cluster 1 represents the highest-risk group, characterized by substantially higher total flood damage,

lower housing values, smaller average frontage lengths, higher water levels, and the greatest number of flood loss events. The housing stock in this cluster is also older on average, indicating elevated structural vulnerability.

Clusters 0 and 2 exhibit comparatively lower flood damage and fewer loss events, along with higher housing values and larger average frontage lengths, suggesting relatively lower flood vulnerability. Although the average year built is similar across clusters, the combined economic and physical characteristics clearly differentiate the resulting risk profiles.

Notably, the highest-risk cluster closely aligns with the highest-damage quartile identified in earlier analyses, reinforcing the conclusion that older, lower-value housing areas are disproportionately affected by flooding. Overall, the clustering results demonstrate that unsupervised, data-driven classification methods can effectively identify neighborhood-level flood-vulnerability patterns and support automated flood-risk assessment and resilience planning.

Table 5. Clustering Results

Predictor	Cluster 0	Cluster 1	Cluster 2
Total Flood Damage	425	2,580	354
Housing Price	131	53	85
Built Year	1940	1942	1946
Water Level	6.0	6.5	5.5
Average frontage length	28501	1600	2600
Number of Losses	2.7	3.0	2.5

Discussion and conclusions

This study examined relationships between observed flood losses and neighborhood housing characteristics in Detroit to better understand residential flood vulnerability in a legacy urban context. The combined results of the correlation and comparative analyses indicate that neighborhood housing characteristics are associated with patterns of flood damage in Detroit, even though they do not fully explain flood losses on their own. Moderate correlations and statistically significant differences across damage quartiles show that ZIP codes with lower average housing values experienced higher cumulative flood damage and more frequent loss events. Housing age exhibited a moderate association with flood damage, but did not differ significantly across damage quartiles, suggesting that older housing stock may contribute to flood vulnerability without clearly distinguishing neighborhoods when aggregated at the ZIP-code scale. These findings indicate that economically vulnerable neighborhoods experience disproportionate flood impacts, reinforcing prior evidence that flood risk and losses are unevenly distributed. The absence of significant differences in average water levels across damage quartiles further suggests that hydrologic intensity alone

does not account for observed neighborhood-scale disparities in flood damage.

The lack of statistically significant regression results highlights the multidimensional nature of urban flood vulnerability. Flood impacts are also shaped by factors not captured in parcel-level housing datasets, including sewer system condition, drainage capacity, basement construction, neighborhood infrastructure age, and localized hydrologic pathways. In addition, aggregation at the ZIP-code scale may obscure finer-grained spatial variability in flood exposure, particularly in legacy cities where housing and infrastructure conditions vary substantially at smaller spatial scales. NFIP claims may also underrepresent total flood losses due to incomplete insurance coverage, underreporting, and socioeconomic differences in claim participation. The findings also underscore trends that warrant further exploration, such as the negative correlation between total damage and house prices in recorded flooding events. Future research should therefore explore alternative modeling approaches, including nonlinear and spatial regression models, as well as machine learning techniques that can capture complex interactions among variables. Incorporating additional covariates—such as drainage system condition, impervious surface coverage, elevation, floodplain proximity, and building-level characteristics—may also improve explanatory power. These additions clarify that the current findings are not treated as terminal, but rather as a basis for guiding more comprehensive modeling approaches in future work.

This study provides empirical evidence of neighborhood-level disparities in flood impacts using observed damage data. While the analysis is limited to the City of Detroit, similar dynamics may be present in other cities with aging housing stock, legacy infrastructure, and recurrent urban flooding. A key limitation of this study is the use of Zillow housing value estimates. Prior research has shown that automated valuation models (AVMs), including Zillow estimates, may be less reliable in areas with high housing vacancy rates, limited transaction activity, and heterogeneous property conditions (Elliott & Clement, 2017). In legacy cities, where housing markets are often fragmented and marked by significant disinvestment, estimated housing values may not fully reflect the true structural conditions or market dynamics. As a result, housing prices should be interpreted as a proxy indicator of neighborhood economic conditions rather than a precise measure of property-level value.

In addition, NFIP claims data may introduce systematic bias due to uneven participation in flood insurance programs. Enrollment in the National Flood Insurance Program is not universal and is often lower in socially vulnerable or lower-income communities due to affordability constraints, limited risk awareness, and institutional barriers (Kousky, 2018; Wing et al., 2022). Consequently, observed claims may underrepresent total flood damage in uninsured or underinsured areas, potentially biasing comparisons across neighborhoods. These limitations are particularly relevant when interpreting equity-related findings, as areas identified as

lower-risk based on claims data may in fact experience substantial unreported losses. Therefore, while the results provide important insights into neighborhood-scale flood vulnerability, they should be interpreted considering the context of this work.

Future research should extend this approach to multiple cities and integrate additional variables related to infrastructure condition, drainage performance, and building-level characteristics to improve explanatory power and generalizability. In addition, future research should integrate additional data sources and finer-scale observations to more fully capture the distribution of flood impacts.

The findings presented in this study are relevant to a diverse set of stakeholders involved in urban flood risk management and housing resilience. City officials and urban planners can use the neighborhood-scale insights to prioritize flood mitigation investments, target infrastructure upgrades, and integrate housing vulnerability considerations into stormwater and resilience planning. Housing agencies and community development organizations may apply the results to identify neighborhoods where older, lower-value housing stock faces elevated flood impacts and where rehabilitation, retrofit, or assistance programs could yield the greatest equity benefits. Homeowners and residents can benefit from improved awareness of localized flood vulnerability, supporting more informed decisions regarding property-level mitigation and insurance coverage. In addition, insurers and risk analysts may use the findings to enhance neighborhood-level risk assessment and better understand the limitations of housing characteristics alone in predicting flood losses. Collectively, these results support more targeted, data-informed, and equitable approaches to urban flood mitigation and housing resilience in legacy cities.

References

- ASCE. (2025). Infrastructure Report Card. <https://infrastructurereportcard.org/>
- Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. *Global environmental change*, 18(4), 598–606.
- Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly*, 84(2), 242–261. <https://doi.org/10.1111/1540-6237.8402002>
- Elliott, J. R., & Clement, M. T. (2017). Natural hazards and local development: The successive nature of landscape transformation in the United States. *Social Forces*, 96(2), 851–876.
- Federal Emergency Management Agency (FEMA). (2023). National Flood Insurance Program (NFIP) Overview and Policy Details.

- IPCC Sixth Assessment Report. (2021). Climate Change 2021. <https://www.ipcc.ch/report/ar6/wg1/>
- Kousky, C. (2018). Financing flood losses: A discussion of the National Flood Insurance Program. *Risk Management and Insurance Review*, 21(1), 11–32. <https://doi.org/10.1111/rmir.12090>
- Liu, J., Shao, W., Xiang, C., Mei, C., & Li, Z. (2020). Uncertainties of urban flood modeling: Influence of parameters for different underlying surfaces. *Environmental research*, 182, 108929.
- Qi, M., Huang, H., Liu, L., & Chen, X. (2020). Spatial heterogeneity of controlling factors' impact on urban pluvial flooding in Cincinnati, US. *Applied Geography*, 125, 102362.
- Shokry, G., Anguelovski, I., Connolly, J. J., Maroko, A., & Pearsall, H. (2022). “They didn’t see it coming”: Green resilience planning and vulnerability to future climate gentrification. *Housing Policy Debate*, 32(1), 211-245.
- Tate, E., Rahman, M. A., Emrich, C. T., & Sampson, C. C. (2021). Flood exposure and social vulnerability in the United States. *Natural Hazards*, 106(1), 435-457.
- U.S. EPA. (2021). Flood Resilience and Stormwater Management.
- Wing, O. E., Lehman, W., Bates, P. D., Sampson, C. C., Quinn, N., Smith, A. M., ... & Kousky, C. (2022). Inequitable patterns of US flood risk in the Anthropocene. *Nature Climate Change*, 12(2), 156-162.