nature portfolio

Peer Review File

Urban Form and Structure Explain Variability in Spatial Inequality of Property Flood Risk among US Counties

Web links to the author's journal account have been redacted from the decision letters as indicated to maintain confidentiality

5th Sep 23

Dear Professor Mostafavi,

Your manuscript titled "Urban Form and Structure Explain Variability in Spatial Inequality of Property Flood Risk among US Counties" has now been seen by 3 reviewers, and we include their comments at the end of this message. They find your work of interest, but some important points are raised. We are interested in the possibility of publishing your study in Communications Earth & Environment, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response that takes into account the points raised. In particular, for publication in Communications Earth & Environment, we require that you:

* Make a compelling case about spatial inequality of property flood risk in US counties and the main factors - development density, economic
activity, centrality, and segregation- influencing flood risk inequality.
* Outline your methodology at a level of detail that will allow an independent researcher to replicate your research.
* Clarify your choice of statistical approach over the other methods.
* Discuss your findings in wider context and address limitation of your findings

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the referees' comments (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and the completed checklist:
[link redacted]
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to receive your revised paper within six weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

Please do not hesitate to contact us if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Martina Grecequet PhD

Associate Editor,
Communications Earth & Environment

EDITORIAL POLICIES AND FORMATTING

We ask that you ensure your manuscript complies with our editorial policies. Please ensure that the following formatting requirements are met, and any checklist relevant to your research is completed and uploaded as a Related Manuscript file type with the revised article.

Editorial Policy: Policy requirements (Download the link to your computer as a PDF.)

Furthermore, please align your manuscript with our format requirements, which are summarized on the following checklist:
Communications Earth & Environment formatting checklist

and also in our style and formatting guide Communications Earth & Environment formatting guide .

*** DATA: Communications Earth & Environment endorses the principles of the Enabling FAIR data project (http://www.copdess.org/enabling-fair-data-project/). We ask authors to make the data that support their conclusions available in permanent, publically accessible data repositories. (Please contact the editor if you are unable to make your data available).

All Communications Earth & Environment manuscripts must include a section titled "Data Availability" at the end of the Methods section or main text (if no Methods). More information on this policy, is available at http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf.

In particular, the Data availability statement should include:
- Unique identifiers (such as DOIs and hyperlinks for datasets in public repositories)
- Accession codes where appropriate
- If applicable, a statement regarding data available with restrictions
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Data Availability Statement.

DATA SOURCES: All new data associated with the paper should be placed in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific, community-recognized repositories, where possible and a list of recommended repositories is provided at http://www.nature.com/sdata/policies/repositories.

If a community resource is unavailable, data can be submitted to generalist repositories such as figshare or Dryad Digital Repository. Please provide a unique identifier for the data (for example a DOI or a permanent URL) in the data availability statement, if possible. If the repository does not provide identifiers, we encourage authors to supply the search terms that will return the data. For data that have been obtained from publically available sources, please provide a URL and the specific data product name in the data availability statement. Data with a DOI should be further cited in the methods reference section.

Please refer to our data policies at http://www.nature.com/authors/policies/availability.html.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

The objective of the paper is to explore the relationship between urban form and structure and the variation in property flood risk across different areas. By analyzing various features such as population density, infrastructure, socioeconomic factors, and mobility patterns, the study identifies significant spatial inequality in property flood risk among US counties and demonstrates how key components of urban development and segregation contribute to this inequality.

Overall, the paper is well written, but some of the analytical strategies and framework are not convincing. I have a few major concerns and I would like the authors to consider the following suggestions, which may improve the paper.

1. Data availability and selection:

The availability of used data seems to be good. The data cover the entire study area and have a reasonable resolution. However, the manuscript reveals little about the criteria and reasoning for the indicator selection. The selected eight urban features (i.e., population density, point of interest density, road density, minority segregation, income segregation, urban centrality index, gross domestic product, and human mobility index) seem to be reasonable for the study area, but the selection is not further justified or explained. The indicators seem to focus on socio-economic sensitivity. The authors do not include any physical sensitivity that could be assessed based on e.g., slope, green and gray infrastructure, other physical flood mitigation or adaptation measures which can significantly impact property flood risk. Again, the selection of the eight key features is not further justified or explained. I would suggest the authors justify and explain the selection of indicators thoroughly. In case the limitation to the selected indicators cannot be justified, the authors should amend their analysis with appropriate indicators.

2. Methodological issues

The authors used Principal Component Analysis (PCA) to identify the three key factors. PCA assumes a linear relationship between features, and the algorithm is not well suited to capturing non-linear relationships. However, the data (particularly, population density, POI density, road density, minority segregation) in Fig. 5 appears to be better suited for a non-linear function rather than a linear specification.

3. The assessment does not fully support the claim

The authors argue that “urban growth and development strategies that exacerbate development density in cities could yield significant spatial inequality of property flood risk.” This claim is based on their analysis of Flood Factor data, which evaluates the risk of flooding for properties. However, it is important to consider potential variations in flood risk vertically. If the Flood Factor model solely relies on data for single-family housing, the authors should acknowledge this limitation. If the dataset includes multi-story residential buildings, the authors' claim could be problematic as the flood risk for lower floors of these buildings may differ from that of higher floor units, which can have a significant impact on the assessment between spatial inequality of property flood risk and development density.

4. Data classification

The authors categorized all the counties into different groups to examine the variations in spatial inequality of property flood risk. However, the numbers of counties will be varied depending on threshold identifying the group. For example, only three counties with a predominantly Asian population are shown in Fig. 2. It would be helpful to clarify the threshold used to identify Asian-dominated counties. Furthermore, there are numerous counties along the US-Mexico border that have a significant Hispanic population. Considering the presence of these Hispanic-dominated counties, it could be important to include them in the analysis. Therefore, I would suggest that the authors explain the reason behind excluding the Hispanic county group in this particular analysis.

Reviewer #2 (Remarks to the Author):

The authors aim to understand the spatial inequality in floods based on urban form and structure. The authors use various data sources as attributes explaining urban form and structure and then apply statistical methods – PCA and OLS and Regression/Decision Tree to determine the factors leading to spatial inequality in flood risk. This is an important topic that addresses a crucial challenge to ensure equity in flood risk management practices and also consider flood risk while developing urban areas.
However, I have the following concerns about the conceptualization and formulation of the paper.
1. The resolution of flood maps (source of SGI) is not provided. Since the authors look at inequality in risk within counties, the resolution of the risk map is important. Flood risk models often inherently consider economic and building characteristics. Hence, an explanation of what factors are considered in developing the maps should be provided to the readers.
2. GDP and Population density are at the county-level and the proportion of minority and low-income population is at the census tract level. However, the derived Minority segregation and income segregation data is at the county-level. Similarly, the indicators of urban features are also computed at county-level. The authors aim to explain the intra-county spatial variability in flood risk with these attributes. The analysis is meticulous and very structured. However, I am not convinced that the inferences are useful. I would ask the authors to explain on how their inferences could help urban planning and flood risk management decision making. For example, income segregation being correlated with SGI could represent inequality is flood adaptation infrastructure or higher damages to more expensive buildings during a flood.
3. The methods section of the manuscript is very clearly written. However, the results and discussion sections are difficult to read and also miss specific take away points apart from general inferences.
4. The authors do not provide uncertainty and error metrics from their model fit and also input (flood risk) data.

Reviewer #3 (Remarks to the Author):

The manuscript explores eight key features (population density, point of interest density, road density, minority segregation, income segregation, urban centrality index, gross domestic product, and human mobility index) related to urban form and structure to explain variability in spatial inequality of property flood risk among 2567 US counties. Key conclusions are that development density is a strong influential factor that is exacerbated if economic activity is relatively high. The paper calls for integrated urban design strategies to address their identified influential factors. Overall, the paper is well written and easy to follow. The visuals are clear although the legends are quite small and hard to read in some cases. The topic is of interest to readers and wide audience. The research is interesting from a scientific and technical standpoint and follows similar studies that have been undertaken in various fields that complete a big-data US type analysis to explore specific questions.

One recommendation for the authors would be to define flood risk overall and clarify it as used in their paper (definitions on line 366 are not sufficient). This is important to put their research in context and bracket the analysis and its findings. It is to be expected that there is spatial inequality in flood risk due to a variety of reasons and not just the ones described by the authors. The authors aim to quantify it and explain it using the specific factors they studied as they stated. They also indicate that the findings are useful to planners and practitioners but this is harder to glean (the final parts of discussion are not persuasive) since urban design is very local and almost never global at the US level.

An aspect that needs clarification is the resolution of their data (a table here might be illustrative) and the relative impact of this resolution heterogeneity on findings (aggregation and disaggregation among the questions that emerge).

The authors additionally need to address issues of locational bias. For example, precipitation that exhibits spatial variability across the US. This is critical especially since some of the findings in the paper are predicated on this and well established in the literature already and are not novel. For example, coastal counties are at higher risk of flooding.

It is also important to explain how and why the specific factors studied were chosen? ie, what is the conceptual model whereby all factors are evaluated and down scaled to the selected 8.

Another comment that is important would be to address the findings in the context of focusing on a smaller geographic region, for example, western coast. Would findings change or remain the same. This could be future work but is worthy of commenting on especially in light of locational bias as noted above. Were the 8 factors studied tested for independence? it would seem that some are correlated and not independent. The statistical analyses need to address this issue.

Reviewer #1

The objective of the paper is to explore the relationship between urban form and structure and the variation in property flood risk across different areas. By analyzing various features such as population density, infrastructure, socioeconomic factors, and mobility patterns, the study identifies significant spatial inequality in property flood risk among US counties and demonstrates how key components of urban development and segregation contribute to this inequality.
Overall, the paper is well written, but some of the analytical strategies and framework are not convincing. I have a few major concerns and I would like the authors to consider the following suggestions, which may improve the paper.

1. Data availability and selection
The availability of used data seems to be good. The data cover the entire study area and have a reasonable resolution. However, the manuscript reveals little about the criteria and reasoning for the indicator selection. The selected eight urban features (i.e., population density, point of interest density, road density, minority segregation, income segregation, urban centrality index, gross domestic product, and human mobility index) seem to be reasonable for the study area, but the selection is not further justified or explained. The indicators seem to focus on socio-economic sensitivity. The authors do not include any physical sensitivity that could be assessed based on e.g., slope, green and gray infrastructure, other physical flood mitigation or adaptation measures which can significantly impact property flood risk. Again, the selection of the eight key features is not further justified or explained. I would suggest the authors justify and explain the selection of indicators thoroughly. In case the limitation to the selected indicators cannot be justified, the authors should amend their analysis with appropriate indicators.

<u>Author's Response:</u>
Thank you for your comment and highlighting the importance of providing specific details about the feature selection in our study. This paper employs a variety of data sources as features to elucidate urban form and structure. After conducting a systematic review of the existing literature and taking data availability into account, we meticulously curated a set of eight features to urban environment. These features were subsequently categorized into two distinct aspects, as they represent different dimensions of urban environment that can have an impact on the spatial inequality in property flood risk. The urban form features encompass socio-economic factors, such as GDP, population density, minority segregation, and income segregation. These factors play a pivotal role in influencing the spatial distribution of populations and the concentration of economic activities within urban areas. On the other hand, the urban structure features, including points of interest (POI) density, road density, urban centrality index, and human mobility index, pertain to the structural layout of cities. This aspect can significantly influence land use and development patterns, subsequently shaping property flood risks. We agree with your point that physical factors, like slope and the presence of green and gray infrastructure, can impact property flood risk. To capture physical infrastructure, we have incorporated POI density into our urban structure features. Our original POI dataset encompasses ten essential categories of urban facilities, including a comprehensive range of critical physical infrastructure types found in urban environment (see "POI density" part of the methods section in the revised manuscript for details).
It is also important to note that while we acknowledge the influence of natural factors, such as hydrology and land topography, on property flood risk, our primary focus in this paper is on urban form and structure. As mentioned in the introduction section of our revised manuscript, we underscore that flood risk is often examined with a predominant focus on natural factors, such as rainfall and land topography. However, there is a growing recognition of the role that urban form and structure play in shaping the spatial distribution of flood risk of properties in cities[1, 2]. This gap aligns with our research objective to delve into the realms of development patterns, economic activity, and facility distribution. Therefore, our feature selection is designed to encompass the most relevant indicators of urban form and structure based on a thorough review of the existing literature. We have further clarified and reinforced the scope of our

manuscript within the introduction section by clarifying that we do not encompass natural hydrological factors. Moreover, to enhance the transparency of our feature selection process, we have provided Table S1 in the Supplementary Information. This table offers a clear reference point for the literature sources that informed our feature screening.

2. Methodological issues

The authors used Principal Component Analysis (PCA) to identify the three key factors. PCA assumes a linear relationship between features, and the algorithm is not well suited to capturing non-linear relationships. However, the data (particularly, population density, POI density, road density, minority segregation) in Fig. 5 appears to be better suited for a non-linear function rather than a linear specification.

Author's Response:

Thank you for your comment regarding our use of PCA in the manuscript. We appreciate the opportunity to clarify the rationale for employing PCA in our analysis. The objective of employing PCA was to effectively reduce the dimensionality of the feature set. This dimensionality reduction has a twofold benefit. First, it enhances the interpretability of the feature set, rendering it more manageable for subsequent analyses. Secondly, it plays a crucial role in mitigating potential issues related to overfitting, especially in the context of our decision tree analysis The primary goal of our decision tree analysis is to uncover as many pathways to high property flood risk inequality as possible while maintaining a high level of performance. By enabling us to generate a larger number of pathways while simultaneously preventing overfitting, PCA was considered as a suitable tool in our study.

*Furthermore, we conducted a correlation analysis of our dataset, as shown in Fig 5 and Fig S3 in the Supplementary Information. We applied the ordinary least squares (OLS) regression model to fit the dataset, setting the confidence interval at 90%. The presence of highly significant p-values (***p<0.001) in the correlation analysis demonstrates the existence of strong linear relationships between the features and Spatial Gini Index. Also, we added two more correlation analysis between the eight features and between the three principal components resulting from PCA in the Fig S4, Fig S6 and Table S2 of Supplementary Information. Fig S4 clearly indicates a robust linear relationship between the eight features. The results in Fig S6 and Table S2 clearly show that linear relationship between the three principal components is not statistically significant. These findings underscore the efficacy of PCA in achieving the desired dimensionality reduction. In light of the considerations outlined above, PCA is a well-suited technique for identifying and characterizing our three principal components.*

3. The assessment does not fully support the claim

The authors argue that "urban growth and development strategies that exacerbate development density in cities could yield significant spatial inequality of property flood risk." This claim is based on their analysis of Flood Factor data, which evaluates the risk of flooding for properties. However, it is important to consider potential variations in flood risk vertically. If the Flood Factor model solely relies on data for single-family housing, the authors should acknowledge this limitation. If the dataset includes multi-story residential buildings, the authors' claim could be problematic as the flood risk for lower floors of these buildings may differ from that of higher floor units, which can have a significant impact on the assessment between spatial inequality of property flood risk and development density.

Author's Response:

Thank you for your comment. We acknowledge your input, and as a result of your comment, we have included additional statements regarding property types in the first part of the methods section to enhance the clarity of our approach. The raw dataset of property flood risk in our study was obtained from Flood Factor Score, a model developed by First Street Foundation[3]. The Flood Factor Score model is designed to evaluate flood risk associated with every individual property within a given area and assigns a corresponding risk score, ranging from 1 to 10. A score of 1 signifies a low probability of flooding occurring within the next 30 years, whereas a score of 10 indicates a high likelihood of flooding. The methodology

employed by the Flood Factor Score 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, not limited solely to single-family housing. Also, our dataset has already taken the variations of various types of properties into account during the calculations, incorporating factors such as property type, year of construction, structure, height, and whether the property is situated within a floodplain.
Furthermore, it is worth emphasizing that our analysis is conducted at the inter-county level, and we do not engage in property-level comparisons. As a result, the specific variations in property characteristics do not need to be a consideration when aggregating the data to the county level.

4. Data classification
The authors categorized all the counties into different groups to examine the variations in spatial inequality of property flood risk. However, the numbers of counties will be varied depending on threshold identifying the group. For example, only three counties with a predominantly Asian population are shown in Fig. 2. It would be helpful to clarify the threshold used to identify Asian-dominated counties. Furthermore, there are numerous counties along the US-Mexico border that have a significant Hispanic population. Considering the presence of these Hispanic-dominated counties, it could be important to include them in the analysis. Therefore, I would suggest that the authors explain the reason behind excluding the Hispanic county group in this particular analysis.

Author's Response:
Thank you for your suggestion. The primary focus of this study is racial disparity among major racial groups in the U.S.. Our analysis of racial disparity primarily focuses on 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. This analytical approach aligns with the methodology frequently employed in influential research studies [4-6]. We have added the above statement in the minority segregation and income segregation part of the methods section.
The identification of the overrepresented race for each county is established through a comparative analysis, which measures the proportion of non-Hispanic White, non-Hispanic Black, and non-Hispanic Asian populations against the average proportions for all the counties (2, 567 counties) included in this study. Here is a breakdown of the steps involved in this identification process:

- *Step 1: we calculate the average percentages of non-Hispanic White, non-Hispanic Black, and non-Hispanic Asian populations across all the counties in our analysis, which stand at 60%, 20%, and 5%, respectively.*
- *Step 2: we determine whether a county has a percentage of non-Hispanic White population equal to or exceed 60%. If it meets this criterion, we label the county as a White-dominant county. If not, we proceed to the step 3.*
- *Step 3: we determine whether a county has a percentage of non-Hispanic Black population equal to or exceed 10%. If it meets this criterion, we label the county as a Black-dominant county. If not, we proceed to the step 4.*
- *Step 4: we determine whether a county has a percentage of non-Hispanic Asian population equal to or exceed 5%. If it meets this criterion, we label the county as a Asian-dominant county. If not, we proceed to the step 5.*
- *Step 5: we assess whether any remaining counties have a percentage of non-Hispanic White population equal to or exceed 50%. If it meets this criterion, we label them as White-dominant counties. Counties that do not meet these criteria are labeled as "other" and are excluded from further analysis.*

We have incorporated the above procedure into Supplementary Information Note 1.

Reviewer #2

The authors aim to understand the spatial inequality in floods based on urban form and structure. The authors use various data sources as attributes explaining urban form and structure and then apply statistical methods – PCA and OLS and Regression/Decision Tree to determine the factors leading to spatial inequality in flood risk. This is an important topic that addresses a crucial challenge to ensure equity in flood risk management practices and also consider flood risk while developing urban areas.
However, I have the following concerns about the conceptualization and formulation of the paper.

1. The resolution of flood maps (source of SGI) is not provided. Since the authors look at inequality in risk within counties, the resolution of the risk map is important. Flood risk models often inherently consider economic and building characteristics. Hence, an explanation of what factors are considered in developing the maps should be provided to the readers.

<u>Author's Response:</u>
Thank you for your comment. Our paper is dedicated to the analysis of spatial inequality in property flood risk, specifically at the county level. This led us to conduct Spatial Gini Index (SGI) mapping, encompassing a total of 2,567 counties spanning the continental United States. Each county was assigned a unique SGI value, and varying colors were employed to visually depict these distinctions (see Fig 1 in the revised manuscript). Our primary dataset for SGI is derived from Flood Factor Model by the First Street Foundation [3]. We have made substantial improvements to the description of the data source and definition of spatial inequality in property flood risk and SGI, while also reinforcing the explanation of our data source, the calculation process, and the range of SGI values in the first part of the methods section. These enhancements are aimed at providing a more comprehensive and detailed understanding of our research methodology.
In addition, as part of our research, we calculated a set of eight features related to urban form and structure. These features were developed by following systematic review of literature and a thorough assessment of data availability. These eight features can be categorized into two distinct aspects, each representing different dimensions of urban environment and its potential influence on property flood risk. Urban form features focus on social and economic factors that have an impact on the spatial distribution of populations and the concentration of economic activities. They encompass key features such as GDP, population density, minority segregation, and income segregation. Urban structure features delve into the structural layout of cities, which can significantly shape land use and development patterns, thereby influencing property flood risks. These features include point of interest (POI) density, road density, urban centrality index, and human mobility index. We have augmented the elaboration of these features in the methods section to provide a more comprehensive understanding. Additionally, to enhance the transparency of our feature selection process, we have provided Table S1 in the Supplementary Information. This table offers a clear reference point for the literature sources that informed our feature screening.

2. GDP and Population density are at the county-level and the proportion of minority and low-income population is at the census tract level. However, the derived Minority segregation and income segregation data is at the county-level. Similarly, the indicators of urban features are also computed at county-level. The authors aim to explain the intra-county spatial variability in flood risk with these attributes. The analysis is meticulous and very structured. However, I am not convinced that the inferences are useful. I would ask the authors to explain on how their inferences could help urban planning and flood risk management decision making. For example, income segregation being correlated with SGI could represent inequality is flood adaptation infrastructure or higher damages to more expensive buildings during a flood.

<u>Author's Response:</u>
Thank you for your comment. This paper aims to explore the extent to which variation in spatial inequality of property flood risk among U.S. counties can be explained by heterogenous features of urban form and

structure. As a result, all of our metrics were computed at the county level. We collected census tract-level data for minority and low-income populations specifically for the calculation of Dissimilarity Index (DI), which we used to assess minority segregation and income segregation (two of our eight features). According to the DI's definition and formula (see the "Minority segregation and income segregation" part of the methods section), it is necessary to compute representative data for subgroups at a smaller geographic scale in comparison to that at the larger geographical scale. Therefore, we employed subgroup-specific data (i.e., minority and low-income populations) at the smaller geographic scale (i.e., census tract) to compute minority segregation and income segregation within the larger geographical scale (i.e., county). We have added Table S1 in Supplementary Information for feature resolution and clarified this aspect of data granularity in the "Minority segregation and income segregation" part of the methods section.

Regarding the comment that how our inferences could help urban planning and flood risk management decision making, we condensed the initial eight features into three principal components via PCA model. Subsequently, we assigned a specific interpretation to each of these principal components (see Table 1 in the revised manuscript), and then we employed a CART model to investigate how the three principal components (i.e., development density, centrality and segregation, and economic activity) interact to create pathways leading to varying levels of spatial inequality in property flood risk within U.S. counties. We have illuminated the interactive relationship among these principal components and their role in shaping spatial inequality in property flood risk (see Fig 6 in the revised manuscript and Fig S7 in Supplementary Information). By identifying specific ranges for each pathway (see Fig 7 in the revised manuscript), we gain a deeper understanding of the factors contributing to spatial inequality in property flood risk, which can provide valuable insights for crafting more targeted and effective flood risk management policies.

For a clearer illustration, let's take an example of extreme level of spatial inequality in property flood risk. In this scenario, the range of development density spans from 0 to 1, indicating its consistent impact on property flood risk, irrespective of the specific values it encompasses. This finding underscores the importance of factoring in a city's development density when assessing spatial inequality in property flood risk: the greater the density of cities, the greater the property flood risk inequality. Hence, urban development strategies that can balance urban density could alleviate property flood risk inequality. Cities with high development density can prioritize addressing this issue by reducing development density. Urban development approaches such as zoning regulations [7], mixed-use development [8], and the prioritization of open spaces [9], which have been shown to alleviate development density in urban areas, can also contribute to mitigating spatial inequality in property flood risk. We have provided and strengthened numerous examples and comprehensive rationales to elucidate our findings in the sections of results and discussion.

3. The methods section of the manuscript is very clearly written. However, the results and discussion sections are difficult to read and also miss specific take away points apart from general inferences.

<u>Author's Response:</u>

Thank you for your comment and acknowledging the methods section. As for the revised results section, we have restructured it by condensing it from four parts into three, with the goal of enhancing its readability and logic. In the first part, we conducted an examination of variation in spatial inequality of property flood risk among US counties. We employed Spatial Gini Index (SGI) to measure this inequality. We carried out the analysis through three main steps: (1) calculating the Probability Density Function (PDF) and Complementary Cumulative Distribution Function (CCDF) of SGI (see Fig 1a, 1b), (2) visualizing the spatial distribution of SGI (see Fig 1c), and (3) presenting boxplot distributions among county groups (see Fig 2). These steps allowed us to examine 2,567 US counties and pinpoint significant disparities in spatial inequality of property flood risk. The second part of the results focuses on "Empirical statistics of urban form and structural features". In this part, we firstly derived eight urban environment features related to urban form and structure. We visually represented these features spatially (see Fig 3) and compared their distributions using boxplots among different levels of SGI (see Fig 4). In addition, we delved into the correlation between these eight features and spatial inequality of property flood risk (i.e., SGI) to better

understand the influence of urban form and structure on this inequality (see Fig 5). In the third part of "Pathways to spatial inequality of property flood risk among US counties", we first utilized Principal Component Analysis (PCA) on the eight features to identify three principal components that contribute to the formation of spatial inequality in property flood risk (see Table 1). Subsequently, we implemented a classification and regression tree model 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 Fig 6).
The revised discussion section primarily revolves around the principal components and the pathways analyses. The first paragraph provides a paper overview, while the second paragraph discussed the findings derived from the spatial distribution of SGI. Following this, we present three more findings that focus on principal components, with particular attention to development density, and how these components shape spatial inequality in property flood risk. Our discussion section concludes with a statement of the contributions and future research work. Throughout the revised results and discussion sections, we have put an emphasis on maintaining a logical organization, providing concrete examples, and making comparisons to previous research to enhance the clarity, organization, and readability of both sections.

4. The authors do not provide uncertainty and error metrics from their model fit and also input (flood risk) data.

<u>Author's Response:</u>
Thank you for your comment. The statistical analysis in this study encompasses three models. The first model is Ordinary Least Squares (OLS) regression model. We employed OLS regression to examine the relationships between urban form and structure (i.e., eight urban form and structure features) and the spatial inequality of property flood risk (i.e., Spatial Gini index) at the county level. You can find more details on model settings in the "Ordinary least squares regression model" part of the methods section. The regression results are presented in Fig 5 of the revised manuscript, while the statistical significance of Kendall coefficient, Pearson coefficient, and Spearman coefficient for each feature can be referenced in Fig S3 of Supplementary Information. The correlation heatmap and Pearson statistical significance results among the eight features are available in Fig S4 of Supplementary Information.
The second model involves Principal Component Analysis (PCA), utilized for dimensionality reduction of the eight urban form and structure features. This process aims to identify the most principal components of urban form and structure that contribute to the spatial inequality of property flood risk. The input for PCA comprises the eight urban form and structure features, and the results of PCA, including the scree plot, definitions and proportions of explained variance of the three principal components, and Pearson correlation among the three principal components, can be found in Fig S5 of Supplementary Information, Table 1, and Figure S6 of Supplementary Information, respectively.
The third model is the Classification and Regression Tree (CART) model, which was implemented to identify pathways leading to varying levels of spatial inequality in property flood risk among counties. This model uses the three principal components as predictor variables and SGI levels as the response variable. Specific details on model settings can be located in the "Classification and regression tree model" part of the methods section. Performance results of the CART model are available in Table S5 of Supplementary Information, while visual representations of the decision tree can be found in Figure 6 and Figure S7 of Supplementary Information.

Reviewer #3

The manuscript explores eight key features (population density, point of interest density, road density, minority segregation, income segregation, urban centrality index, gross domestic product, and human mobility index) related to urban form and structure to explain variability in spatial inequality of property flood risk among 2567 US counties. Key conclusions are that development density is a strong influential factor that is exacerbated if economic activity is relatively high. The paper calls for integrated urban design strategies to address their identified influential factors. Overall, the paper is well written and easy to follow. The visuals are clear although the legends are quite small and hard to read in some cases. The topic is of interest to readers and wide audience. The research is interesting from a scientific and technical standpoint and follows similar studies that have been undertaken in various fields that complete a big-data US type analysis to explore specific questions.

Author's Response:

Thank you for your insightful comments on our manuscript. We appreciate your acknowledgment of our contribution to the existing body of knowledge. We have plotted high-resolution version of the pictures and increased the size of the legends in the manuscript and supplementary information as much as possible while still preserving overall aesthetics.

1.One recommendation for the authors would be to define flood risk overall and clarify it as used in their paper (definitions on line 366 are not sufficient). This is important to put their research in context and bracket the analysis and its findings. It is to be expected that there is spatial inequality in flood risk due to a variety of reasons and not just the ones described by the authors. The authors aim to quantify it and explain it using the specific factors they studied as they stated. They also indicate that the findings are useful to planners and practitioners but this is harder to glean (the final parts of discussion are not persuasive) since urban design is very local and almost never global at the US level.

Author's Response:

Thank you for your positive and constructive comments. Spatial inequality of property flood risk, as used in our revised paper, 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. It is important to note that while we acknowledge the influence of natural factors, such as hydrology and land topography, on property flood risk, our primary focus in this paper is on urban form and structure. As mentioned in the introduction section of our revised manuscript, we underscore that flood risk is often examined with a predominant focus on natural factors, such as rainfall and land topography. However, there is a growing recognition of the role that urban form and structure play in shaping the spatial distribution of flood risk in cities [1, 2]. This aligns with our research objective to delve into the realms of development patterns, economic activity, and facility distribution. Therefore, our feature selection is designed to encompass the most relevant indicators of urban form and structure based on a thorough review of the existing literature. We appreciate your insightful suggestion and we have revised the definition of spatial inequality of property flood risk in the first paragraph of methods section. In this paragraph, we have both emphasized the diverse causes of flood risk and limited the scope of our study to the urban form and structure. We hope that these clarifications and adjustments to our definition align with your suggestion and provide a more precise understanding of the scope and focus of our research.

Regarding the issue of the locality of urban design on flood risk, we want to clarify that our research was conducted at the county level due to data availability and computing complexity constraints. The methodology in our paper allows us to provide valuable pathways and practical suggestions at the county level in the U.S.. For instance, in our discussion, we highlight that specific counties, such as New York County in New York State, Los Angeles County in California, and Harris County in Texas, rank at the top *for extreme spatial inequality in property flood risk. By doing so, we aimed to inspire urban planners in these counties to pay closer attention to our findings, particularly in terms of development density, economic activity, and centrality and segregation. Also, we mentioned in the discussion section (paragraph 2) 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. In addition, we have provided example counties for all the pathways to spatial inequality of property flood risk, which can be found in the Table S4 of Supplementary Information. Therefore, our research not only offers general recommendations but also take full account of locality, providing highly specific insights and actionable guidance at different geographical levels.*

2. An aspect that needs clarification is the resolution of their data (a table here might be illustrative) and the relative impact of this resolution heterogeneity on findings (aggregation and disaggregation among the questions that emerge).

<u>Author's Response:</u>
Thank you for your comment and highlighting the importance of providing specific details about datasets used in our study. We have provided the table of data resolution for Spatial Gini Index and the eight urban form and structure features in the Table S1 of Supplementary Information. It is important to clarify that our calculations were ultimately based on the county-level aggregation, with all data being aggregated to this level. Notably, three of the features, namely Spatial Gini Index of property flood risk, population density, GDP, and road density, were originally obtained at the county level, ensuring consistency in data resolution. Minority segregation, income segregation, and UCI required calculations at a smaller geographical unit (census tract level in our study) comparing to the geographical unit (county level) according to their respective definitions. This difference in resolution was intentional and not indicative of resolution heterogeneity. In the case of POI density and HMI, we collected POI information at the user level within a county. The aggregation of user-level POI information to a larger geographical unit, such as the county, is a common practice in previous research studies, often done to protect privacy [1, 10-15]. Although the raw data for certain indicators needed to be sourced or computed from smaller geographic units, it is essential to underscore that all eight features and Spatial Gini Index were ultimately calculated using county-level data. Based on the existing body of research [11, 16], we believe that our approach to data resolution is unlikely to affect the validity of our findings.

3. The authors additionally need to address issues of locational bias. For example, precipitation that exhibits spatial variability across the US. This is critical especially since some of the findings in the paper are predicated on this and well established in the literature already and are not novel. For example, coastal counties are at higher risk of flooding.

<u>Author's Response:</u>
Thank you for putting forward the good question. Our findings offer data-drive evidence for the existence of significant spatial inequality in property flood risk in the metropolitan and coastal counties of the United States. It is important to clarify that our focus is on highlighting the substantial spatial inequality in coastal regions, not merely the acknowledgment of high flood risk, which is widely studied. Our primary focus in this study is on the role of urban form and structure in shaping property flood risk inequality (which is rather understudied).
As we state in the first part of the methods section, we use Spatial Gini index (SGI) to measure this spatial heterogeneity of property flood risk in a county, not flood risk. What we mean by "high spatial inequality" has yet to be thoroughly explored in existing literature. Also, as mentioned in the first response, we agree that flood risk is often examined with a focus on natural factors, like precipitation. However, our research

does not focus on the natural, hydrological, or similar features. Instead, our emphasis lies on the urban environmental dimensions of urban form and urban structure. We appreciate your comment again as it has been instrumental in reinforcing our focus in this paper.

4. It is also important to explain how and why the specific factors studied were chosen? ie, what is the conceptual model whereby all factors are evaluated and down scaled to the selected 8.

Author's Response:
Thank you for your comment and highlighting the importance of providing specific details about the feature selection in our study. This paper employs a variety of data sources as features to elucidate urban form and structure. After conducting a systematic review of the existing literature and taking data availability into account, we meticulously curated a set of eight features to the urban environment. These features were subsequently categorized into two distinct aspects, as they represent different dimensions of the urban environment that can have an impact on the spatial inequality in property flood risk. The urban form features encompass socio-economic factors, such as GDP, population density, minority segregation, and income segregation. These factors play a pivotal role in influencing the spatial distribution of populations and the concentration of economic activities within urban areas. On the other hand, the urban structure features, including points of interest (POI) density, road density, urban centrality index, and human mobility index, pertain to the structural layout of cities. This aspect can significantly influence land use and development patterns, subsequently shaping property flood risks. To enhance the transparency of our feature selection process, we have provided Table S1 in the Supplementary Information. This table offers a clear reference point for the literature sources that informed our feature screening.

5. Another comment that is important would be to address the findings in the context of focusing on a smaller geographic region, for example, western coast. Would findings change or remain the same. This could be future work but is worthy of commenting on especially in light of locational bias as noted above. Were the 8 factors studied tested for independence? it would seem that some are correlated and not independent. The statistical analyses need to address this issue.

Author's Response:
Thank you for your insightful comment. We acknowledge the presence of locational bias as a critical consideration in the study of flood risk. However, as we addressed in response to comments #1 and #3, our research is primarily centered on the examination of spatial disparities in property flood risk across all counties in the continental United States. We do not focus on an analysis of flood risk or related natural factors (such as weather and rainfall features). Furthermore, we conducted an extensive analysis of spatial inequality, categorizing this kind of spatial inequality based on county-level data for all counties, as shown in Fig 7(b) and Table S4 of Supplementary Information. It is worth noting that if we had chosen to focus solely on a specific region (e.g., the West Coast) and utilized smaller geographic units (e.g., census tract), we might have obtained different results. Machine learning-based studies can yield varying outcomes based on changes in the sample size. We have clarified this point in the discussion section in he revised manuscirpt. Regarding the matter of features independence, we conducted a correlation test among the eight features. The correlation heatmap and the Pearson statistical significance results for these eight features can be found in Figure S5 of the Supplementary Information. It is evident from the results that most of the features exhibit significant correlations. Notably, the correlations between income segregation and population density/POI density are not considered significant, as they pertain to different aspects of urban environmental features. However, it's important to clarify that our primary focus in this study is on how these eight features collectively explain Spatial Gini Index (SGI). We are not concerned with the correlations among these eight features. In our analysis of the manuscript, we examined the relationship between each feature and SGI, and each feature is significantly correlated with SGI (see Fig 5 in the revised manuscript for details). Subsequently, we employed Principal Component Analysis (PCA) to downscale

these features. At this stage, we considered the correlation between the three principal components, which is crucial to ensure their independence as we need to define them differently (see Table 1 in the revised manuscript) and conduct CART model based on them. Details regarding the Pearson correlation among the three principal components is located in Fig S6 and Table S2 within the Supplementary Information.

Reference:

1. Hao, H. and Y. Wang, *Disentangling relations between urban form and urban accessibility for resilience to extreme weather and climate events.* Landscape and urban planning, 2022. 220: p. 104352.
2. Puzyreva, K., et al., *Professionalization of community engagement in flood risk management: Insights from four European countries.* International Journal of Disaster Risk Reduction, 2022. 71: p. 102811.
3. *First Street Foundation.* First Street Foundation Flood Model, 2022. https://firststreet.org/risk-factor/.
4. Liu, T. and C. Fan, *Impacts of disaster exposure on climate adaptation injustice across US cities.* Sustainable Cities and Society, 2023. 89: p. 104371.
5. Jbaily, A., et al., *Air pollution exposure disparities across US population and income groups.* Nature, 2022. 601(7892): p. 228-233.
6. Mehta, N.K., H. Lee, and K.R. Ylitalo, *Child health in the United States: recent trends in racial/ethnic disparities.* Social Science & Medicine, 2013. 95: p. 6-15.
7. Cutter, S.L., et al., *Flash flood risk and the paradox of urban development.* Natural Hazards Review, 2018. 19(1): p. 05017005.
8. Su, W., et al., *Urban land pattern impacts on floods in a new district of China.* Sustainability, 2014. 6(10): p. 6488-6508.
9. Pallathadka, A., et al., *Urban flood risk and green infrastructure: Who is exposed to risk and who benefits from investment? A case study of three US Cities.* Landscape and Urban Planning, 2022. 223: p. 104417.
10. Pan, W., et al., *Urban characteristics attributable to density-driven tie formation.* Nature communications, 2013. 4(1): p. 1961.
11. Fan, C., et al., *Interpretable machine learning learns complex interactions of urban features to understand socio-economic inequality.* Computer-Aided Civil and Infrastructure Engineering, 2023.
12. Aleta, A., et al., *Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19.* Nature Human Behaviour, 2020. 4(9): p. 964-971.
13. Wang, F., et al., *Extracting trips from multi-sourced data for mobility pattern analysis: An app-based data example.* Transportation Research Part C: Emerging Technologies, 2019. 105: p. 183-202.
14. Zhang, X. and N. Li, *Characterizing individual mobility perturbations in cities during extreme weather events.* International Journal of Disaster Risk Reduction, 2022. 72: p. 102849.
15. Wang, Q., et al., *Urban mobility and neighborhood isolation in America's 50 largest cities.* Proceedings of the National Academy of Sciences, 2018. 115(30): p. 7735-7740.
16. Niu, T., Y. Chen, and Y. Yuan, *Measuring urban poverty using multi-source data and a random forest algorithm: A case study in Guangzhou.* Sustainable Cities and Society, 2020. 54: p. 102014.

9th Feb 24

Dear Professor Mostafavi,

Your manuscript titled "Urban Form and Structure Explain Variability in Spatial Inequality of Property Flood Risk among US Counties" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment under the open access CC BY license (Creative Commons Attribution v4.0 International License).

We therefore invite you to edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at https://www.nature.com/documents/commsj-file-checklist.pdf .

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication under a [CC BY license](#) (Creative Commons Attribution 4.0 International License). This license allows maximum dissemination and re-use of open access materials and is preferred by many research funding bodies.

For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit https://www.nature.com/commsenv/article-processing-charges

At acceptance, you will be provided with instructions for completing this CC BY license on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:
[link redacted]
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Martina Grecequet, PhD
Associate Editor,
Communications Earth & Environment
@CommsEarth

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

Thank you for giving me an opportunity once again to take a look at this valuable manuscript. After second review, I am happy that authors have addressed the reviewers comments.

Reviewer #3 (Remarks to the Author):

The authors have responded to all reviewer comments. The manuscript is much improved and does not require further revisions. It can be published in its present format.