

Research papers

Flood resilience through crowdsourced rainfall data collection: Growing engagement faces non-uniform spatial adoption

Alexander B. Chen^a, Jonathan L. Goodall^{a,*}, T. Donna Chen^a, Zihao Zhang^b

^a Department of Engineering Systems and Environment, University of Virginia, 151 Engineers Way, Charlottesville, VA 22904, USA

^b School of Architecture, University of Virginia, Campbell Hall, PO Box 400122, Charlottesville, VA 22904, USA¹

ARTICLE INFO

This manuscript was handled by Nandita Basu, Editor-in-Chief, with the assistance of Marc F. Muller, Associate Editor

Keywords:

Crowdsourcing
Rainfall
Flooding
Spatial analysis
Logistic regression
Flood resilience

ABSTRACT

Crowdsourced Personal Weather Stations (PWSs) adoption has been growing rapidly and provides the potential to fill in hyper-local rainfall observation gaps. However, current adoption patterns exhibit spatial biases that must be understood when using the data for modeling and decision-making. Here, we first examine the PWS rainfall spatial representation at HUC-12 watersheds in twelve metropolitan areas in the U.S. Furthermore, by modeling the PWS adoption using socio-economic and flood-related data at census tract level, the results suggest current adoption patterns exhibit spatial biases toward wealthier neighborhoods and flood-prone regions. The findings provide insights to inform how policies could be made to distribute resources to improve the rainfall data collection efforts in PWS-underrepresented regions. As crowdsourced data are increasingly used for decision-making by policymakers, efforts to close the gap in current non-uniform PWS spatial adoption will allow crowdsourced rainfall data to be better positioned to support decision-makers in their flood resilience efforts.

1. Introduction

Flooding causes significant social and economic damage and loss in the United States each year (The National Academies of Science, Engineering, 2019). Hurricane Harvey is an extreme example of bringing unprecedented rainfall across the city of Houston, Texas, leading to catastrophic flooding that impacted more than 100,000 homes and incurred an estimated damage cost of \$125 billion (Van Oldenborgh et al., 2017). With the increase in frequency and intensity of heavy rainfall due to climate change, the impact of flooding is projected to become more severe over time (Cheng and Aghakouchak, 2014; Fowler et al., 2021). In urban watersheds with large portions of impervious surfaces and low-lying areas, rainfall-driven flooding has been causing even more considerable damage, affecting greater numbers of people (Rosenzweig et al., 2018). Therefore, there is a need for improved flood assessments as well as building flood resilience in urban areas to adapt to the increasing severity of flood hazards (Bertilsson et al., 2019).

Flood assessments at high spatial and temporal resolution remain a challenge. An ideal flood assessment will require a spatially and temporally representative dataset, as flood modeling involves complex and nonlinear flow and physical processes (Hu et al., 2019).

Traditionally, flood assessments have mainly been efforts by government agencies. Local government agencies are responsible for identifying flood risks and creating flood forecasting models (Tyler et al., 2019). However, flood-related data available from government agencies needed for enabling high-resolution flood modeling at flood-prone areas are usually insufficient or lacking (Cristiano et al., 2017). Furthermore, increasing attention has been focused on incorporating the public and a wider representation of stakeholders in the decision-making process of flood assessment and management (White et al., 2010). Lack of engagement from communities for locally specific flood knowledge will lead to inadequate flood model calibration and validation (Gebremedhin et al., 2020). In such cases, flood forecasting models generated by government agencies may not be fully accepted and supported by local stakeholders, as they may not reflect adequate representations of floods in certain regions (Gebremedhin et al., 2020; Rosenzweig et al., 2018; Sy et al., 2019). Lack of engagement of local stakeholders may also worsen the disparities of social, economic, and environmental resources across communities (White et al., 2010).

Crowdsourced data can provide a potential solution. Crowdsourcing is a method of collecting information from as many involved contributors from the general public (Muller et al., 2015), which offers a way of

* Corresponding author.

E-mail address: goodall@virginia.edu (J.L. Goodall).

¹ Now with Spitzer School of Architecture, City College of New York, 141 Convent Ave, New York, NY 10031, USA.

obtaining large amounts of flooding-related data cost-efficiently (Assumpção et al., 2018). With the development of inexpensive sensors and communication technology, the amount of data collected by the public has been growing rapidly, generating a wealth of information (Muller et al., 2015). For example, social media such as Twitter data produces geotagged information that can be extracted to map the potential extent of flood events (Wang et al., 2018). Mobile apps such as Google Waze enable users to report floods in a convenient and efficient way, which greatly supports government efforts to identify the location of floods (Prahara et al., 2021). In the case of rainfall monitoring, the adoption of personal weather stations (PWSs) by the general public fills in the rainfall observation gap in locations where government agency data are unavailable (de Vos et al., 2017). PWSs are off-the-shelf weather stations installed and maintained by individuals, and the reported data can be easily shared through websites such as Weather Underground (Gharesifard and Wehn, 2016). In recent years, crowdsourced PWS adoption has been growing rapidly to supplement agency-maintained rain gauges which are usually limited in coverage (de Vos et al., 2017), as well as remote sensing rainfall (e.g. satellite and radar) which requires validation from ground rain gauges (Muller et al., 2015). For example, in Houston, Texas, PWS adoption density has grown from 0.06 to 0.24 PWS per km² from 2016 to 2019. Such exponential growth suggests that in populated areas in the U.S., PWSs could alone provide sufficient spatial resolutions for rainfall observations needed for urban hydrology in a few years (Berne et al., 2004; Chen et al., 2021).

Furthermore, by participating in data collection via crowdsourcing, the public may also become more proactive in engaging in local decision-making processes (Buytaert et al., 2016). In fact, the motivation behind adopting PWSs and sharing the collected data often starts with owners using the data for personal purposes. Simultaneously, growth in PWS adoption has the potential to benefit society at large since the collected data could implicitly create crucial knowledge for local communities (Gharesifard and Wehn, 2016). Therefore, as local stakeholders (e.g. PWS owners) gain better understanding of their environments with time, empowering them to grow from collecting data toward developing models and decision support systems becomes key (Almoradie et al., 2015; Voinov et al., 2016). The ongoing trend of increasing crowdsourcing participation (e.g., rainfall data collection by PWS adoption) from the general public in local communities, therefore, lays the foundation for generating collective knowledge to support flood resilience efforts (Paul et al., 2018).

Despite the rapid growth of PWS adoption for generating a wealth of rainfall observations, current adoption patterns exhibit spatial biases caused by underrepresentation or overrepresentation of certain regions (Muller et al., 2015). Using non-representative and spatially-biased datasets as model inputs could lead to biased modeling results and decision-making (Towe et al., 2020). Furthermore, recent advancements of data-driven techniques have enabled modeling flood at high spatial and temporal resolution (Mosavi et al., 2018; Sadler et al., 2018; Shen et al., 2019; Zahura et al., 2020), which has also highlighted the importance of using representative datasets, as large datasets are not always comprehensive (Torralba and Efron, 2011). Therefore, it is important to ensure that crowdsourced rainfall datasets are analyzed for spatial and temporal representation and variability and, if biases exist, such biases are clearly understood and accounted for when using the crowdsourced data in modeling and decision-making.

Spatial bias in crowdsourced data has been studied across a variety of fields. For example, social media usage is mainly concentrated in populous areas, leading to certain groups not receiving needed assistance during disasters (Fan et al., 2020). Mobile phone applications have the potential to enhance resilience building against disasters, but the lack of user-centered design often results in low uptakes in marginalized and vulnerable groups who are more vulnerable in disaster situations (Craig et al., 2019; Paul et al., 2021). Crowdsourced bicycle ridership data are biased toward recreational riders (who track exercise activity), which requires a correction for better representation of the ridership

patterns of all riders, such as commuter cyclists (Roy et al., 2019). Biodiversity data collected from crowdsourcing exhibited spatial biases toward accessible areas or recreational summer homes (Millar et al., 2019). PWS adoption and other crowdsourced climate and atmospheric data have spatial biases toward populous areas (Muller et al., 2015). In previous studies, regression models using socio-economic and demographic factors have been widely used for assessing the factors that affect the spatial bias pattern. Examples include exploring the effect of socio-economic factors and environmental attitudes on rain barrel spatial adoption patterns (Ando and Freitas, 2011), the presence of hazardous waste sites affecting life expectancy (Kiaghadi et al., 2021), and spatial models to predict electric vehicle ownership choice behavior (Chen et al., 2015). However, limited studies have been focused on the fast-growing crowdsourced PWS network to provide high-resolution rainfall observations and engage communities in supporting flood resilience. Therefore, the research questions that guide this study are: (1) Are PWS providing a spatially representative sample of rainfall data? (2) What are the underlying factors that affect the spatial pattern of PWS adoption?

In this study, we used a unique PWS adoption dataset obtained from Weather Underground, which is one of the largest crowdsourced PWS platforms. This dataset consists of the location of more than 100,000 PWS in the contiguous U.S. PWS rainfall representation in United States Geological Survey (USGS) level-12 Hydrological Unit Codes (U.S. Geological Survey, 2021) (HUC-12) watersheds in 12 selected metropolitan areas were analyzed to quantify to what extent current PWS adoption can contribute to rainfall observations. We further applied logistic regression models using socio-economic and flood-related data to identify the factors that influence the spatial bias in PWS adoption. Moreover, we measured and analyzed the marginal effects of resulting models to quantify the PWS adoption disparities across neighborhoods.

2. Materials and methods

2.1. PWS adoption data acquisition

PWS adoption dataset used in this study was obtained from Weather Underground database using their previous version of Application Programming Interface (API). The analysis dataset contains metadata including the ID and the geographic location (latitude and longitude) of more than 100,000 PWSs. The geographical location of these PWSs were then mapped using ArcGIS to explore the PWS adoption spatial pattern in the contiguous U.S. Note that at the end of 2019, this version of the API has been retired. The new version of the API (introduced in 2020) requires a Weather Underground key which can be obtained through connecting a PWS to the platform. Furthermore, the number of API calls per day in order to download the data was greatly limited, and some of the metadata including PWS installation date were no longer available. Therefore, to preserve such details in the dataset, this study focused on the PWS adoption dataset gathered on April 2019.

2.2. PWS rainfall representation calculation

PWS rainfall representation was evaluated based on the PWS density in an urban watershed within a metropolitan area (MA). In this study, an urban watershed is defined as HUC-12 watersheds that intersected with United States Census Bureau (USCB) delineated urban areas within the MA boundary (U.S. Census Bureau, 2021). To compute the PWS density in an urban watershed of each MA, we used the point density tool in ArcGIS (version 10.6) to convert PWS location point data to gridded raster which represents the PWS density, followed by using the zonal statistics tool to compute the average PWS point density in a watershed (see Fig. 1 for the workflow). Like traditional rain gauges, rainfall data recorded from PWSs are point observations which can only be representative of the rainfall of a specific area due to the spatial variability of rainfall events (Cristiano et al., 2017). To have a sufficient

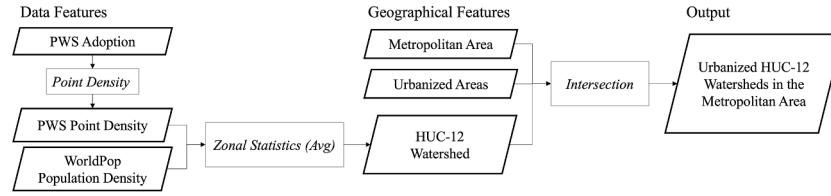


Fig. 1. GIS workflow of the PWS rainfall representation calculation. The ArcGIS tools used are the italics text in the gray box.

representation of rainfall spatial variability in terms of rain gauge density in a monitoring network, the World Meteorological Organization recommends at least one rain gauge per 10–20 km² for urban areas (World Meteorological Organization (WMO), 2008). Therefore, we used an average PWS point density of 0.1 PWS per km² (1 PWS per 10 km²) as the threshold to assess PWS rainfall representation. An underrepresented (UR) watershed is defined as having a point density lower than the threshold, while a non-UR watershed is defined as having a point density above the threshold. Using an one-way ANOVA test, we further extracted the population density estimates from the WorldPop population dataset (Tatem, 2017) to test if a significant mean difference in mean population density exists between UR and non-UR watersheds.

2.3. PWS adoption logistic regression model

We used logistic Regression to assess the association between the PWS adoption and selected socio-economic data. The built-in “glm” package in R programming language (version 3.6.1) was applied to run the regression models. Census tract (CT) level socio-economic data were obtained from the USCB’s 2017 American Community Survey 5-year estimate, including population, median household income, household density, and owner-occupied household ratio (the number of owner-occupied households divided by the total number of households). CT was selected as the common geographic unit for analysis because it is the smallest geographic boundary used by the Census Bureau. To build the logistic Regression model, we aggregated the PWS location information into total counts of PWS adopted in each CT, then classified these counts into binary groups: *PWS adoption* = 1 for CTs that have at least 1 PWS (1 + PWS) and *PWS adoption* = 0 for CTs that have no PWS adoption. The logistic Regression equation is shown as follows:

$$p(\text{PWS adoption}) = \frac{e^{\beta_0 + \sum_{i=1}^n \beta_i Q_i + \sum_{i=n+1}^m \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^n \beta_i Q_i + \sum_{i=n+1}^m \beta_i X_i}} \quad (1)$$

where p is the probability of 1 + PWS adoption in a CT, β_0 is the intercept, $\beta_1, \beta_2, \dots, \beta_{n-1}$ are the coefficients for the categorical variable $Q_i (i = 1, 2, \dots, n-1)$ and Q_i are the dummy variables for each category with the value of either 0 or 1. $\beta_n, \beta_{n+1}, \dots, \beta_m$ are the coefficients for the continuous variables X_i . The regression coefficients are estimated by maximum likelihood. In the explanatory variables, to account for household income differences across MAs, the median household income (MHI) was converted into categorical variables using quartile groups *MHI_Q1*, *MHI_Q2*, *MHI_Q3*, and *MHI_Q4*. In each analyzed MA, a CT falling within first quartile of MA MHI was assigned to *MHI_Q1*; a CT falling between first and second quartile assigned to *MHI_Q2*; a CT falling between second and third quartile was assigned to *MHI_Q3*. a CT falling above the fourth quartile was assigned to *MHI_Q4*. Population (POP), household density (HHD) and owner-occupied household ratio (OOHR) are used as continuous variables for the logistic regression model.

2.4. Flood vulnerability and PWS adoption

Two types of flood-related datasets at the CT level, (i) the total number of flood claims (FC) and (ii) percent housing units in the 100-year flood zone (FZ), were used to assess the association between

potential flood risk and PWS adoption. FC data were obtained from National Flood Insurance Program (NFIP) Redacted Claims, which is a large database containing more than two million claims transactions since the NFIP launched (Federal Emergency Management Agency, 2021). FZ data were obtained from a dataset published by NYU Furman Center. This dataset was created by combining housing and population data with FEMA floodplain maps to calculate the percent of housing units intersecting with a FEMA 100-year floodplain (NYU Furman Center, 2021). In this study, we used the FC and FZ data to represent the flood vulnerability of a CT. We assumed that a CT with a higher number of flood insurance claims is more likely to have a higher flood risk, and a CT with a higher percentage of housing units in the floodplain also implies that this CT is more vulnerable to flooding. For each analyzed MA, we classified CTs into *high/low* FC groups based on the FC value, with the *low* FC group having FCs below the median value of the CTs within the MA representing the lower flood risk CTs, and *high* FC group having FCs above the median values representing the higher flood risk CTs. Similarly, the FZ data were used to classify CTs into *in FZ* and *not in FZ* groups, with *in FZ* group representing CTs that have any percent housing units in the floodplain, and *not in FZ* group representing CTs that have zero percent in the floodplain.

3. Results

3.1. PWS adoption in the contiguous U.S.

The spatial pattern of PWS shows that adoption is concentrated in the metropolitan areas in the east and west coast of the contiguous U.S. (Fig. 2a). Unlike the agency-operated rainfall network where rainfall stations are usually uniformly distributed, PWS is spatially biased toward populous areas. At the metropolitan area (MA) level, PWS adoption is also highly correlated with population (Fig. 2b), with a correlation coefficient of 0.88 for all MAs in the contiguous U.S. To assess the PWS representation in the MAs in the contiguous U.S., the top 12 MAs by PWS adoption (Fig. 2b) were selected in this study. The analyzed MAs are distributed across the contiguous U.S., with the number of PWS adoption ranging from 1,300 to 2,569 units per MA and PWS density ranging from 0.03 to 0.27 PWS per km².

3.2. PWS rainfall representation in HUC-12 watersheds

In this study, watersheds with PWS point density < 0.1 are considered underrepresented (UR); otherwise, they are considered well-represented (non-UR). The results of the PWS representation calculation showed that the current PWS rainfall representation exhibits three characteristics (Table 1). First, at the MA level, PWS rainfall representation in HUC-12 watersheds varied across the analyzed MAs. The average PWS point density ranges from 0.14 in Atlanta to 0.59 in San Francisco. Most of the analyzed MAs (besides Chicago) have a maximum PWS point density above 0.40, indicating that PWS has the potential to contribute to the coverage of rainfall observations for at least every 2.5 km² in those watersheds. Among the analyzed MAs, San Francisco has the highest maximum PWS point density of 1.34 PWS per km², which could provide a considerable rainfall representation that is even greater than the current resolution of radar-derived rainfall (e.g., The Next Generation Weather Radar (NEXRAD) typically has 1 km by 1 km resolution).

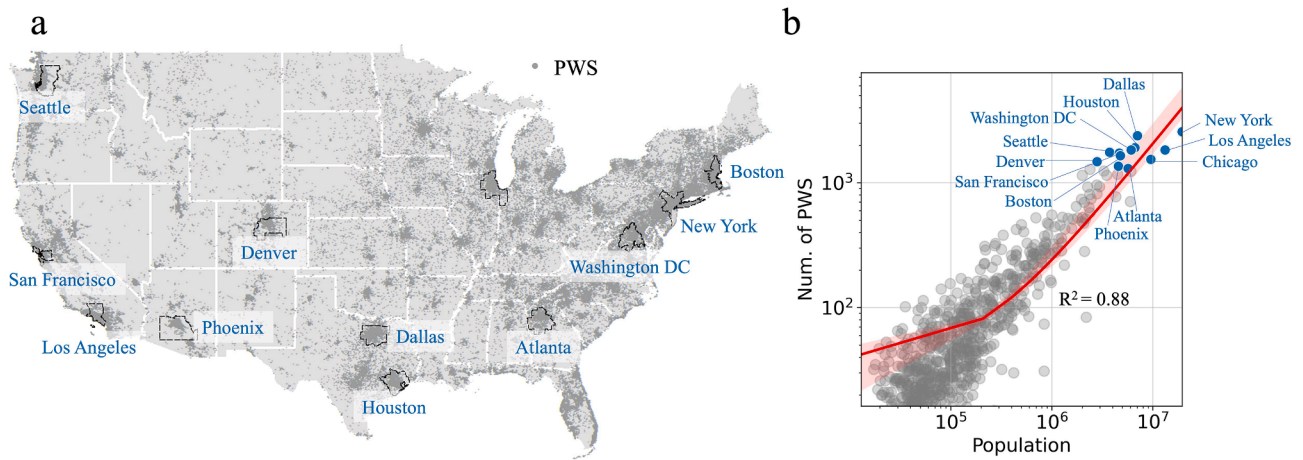


Fig. 2. PWS adoption pattern of the contiguous U.S. as of April 2019. a Each gray dot represents a PWS. The 12 metropolitan areas (MAs) with the largest number of PWS adoption were selected for analysis. b PWS adoption highly correlates with the population in MAs in the contiguous U.S.

Table 1

PWS representation in HUC-12 watersheds for the 12 analyzed metropolitan areas (MAs). The one-way ANOVA test showed that the mean population density difference in underrepresented (UR) and non-UR watersheds is not significant.

MA	N. of watersheds in urbanized area	PWS point density (/km ²)			Num. of UR	UR %	Population density (/km ²)		
		Mean	Max	Std			Avg. UR	Avg. Non-UR	F-test p-value
New York	111	0.18	0.43	0.07	15	14%	1039	1928	0.189
Dallas	53	0.29	0.64	0.16	6	11%	1130	1285	0.440
Houston	30	0.25	0.48	0.15	9	30%	1076	1459	0.100
Los Angeles	28	0.26	0.46	0.11	1	4%	2011	2467	–
Washington DC	37	0.25	0.62	0.15	6	16%	751	1418	0.032
Seattle	20	0.40	0.73	0.15	0	0%	–	1195	–
San Francisco	17	0.59	1.34	0.29	0	0%	–	2163	–
Boston	38	0.22	0.58	0.09	1	3%	600	789	–
Chicago	84	0.15	0.37	0.07	19	23%	972	1235	0.340
Denver	15	0.42	0.67	0.16	0	0%	–	1322	–
Phoenix	18	0.28	0.44	0.11	1	6%	601	1404	0.155
Atlanta	88	0.14	0.40	0.10	35	40%	626	870	0.002

Second, though the average PWS point density for analyzed 12 MAs is well-represented at the MA level, large disparities occurred at the HUC-12 watershed level when uneven representation begins to appear. For example, in Seattle and Denver, PWS adoptions are more uniformly distributed where no UR watersheds were observed. However, the percentage of UR watersheds is larger in MAs such as Atlanta (40%) and Houston (30%). Third, the mapping of the PWS representation shows that the PWS representation has certain spatial distribution patterns which merit further analysis. As shown in Fig. 2, UR watersheds are generally clustered in a specific region of a MA. For example, in Houston, UR watersheds are concentrated on the east portion of the MA where PWS adoption in these watersheds is relatively low. In Chicago and Atlanta, large portions of UR watersheds are concentrated in the southern part of the MA.

We hypothesized that population density is explanatory for the PWS adoption spatial disparities. To test this hypothesis, population density estimates for each HUC-12 watershed were computed using the WorldPop dataset (Tatem, 2017). As shown in Table 1, the average population density in non-UR watersheds was higher than in UR watersheds. However, the result of the one-way ANOVA test showed that the mean difference of population density in UR and non-UR watersheds in most MAs was not statistically significant. Although there is a notable PWS adoption difference in UR and non-UR watersheds, population density did not fully explain the spatial bias in PWS adoption. As can be seen in the scatterplots in Fig. 3, most UR watersheds (shown in red dots) have similar levels of population density compared to non-UR watersheds (shown in blue dots) while lacking PWS representation. Therefore,

we implemented further analysis at a finer geographic scale (census tract) to assess other factors affecting the spatial pattern of PWS adoption.

3.3. PWS adoption logistic regression model

Logistic regression models were built for the 12 analyzed MAs (12 regression models) to assess the factors affecting the spatial pattern of PWS adoption. Table 2 shows the summary statistics of the response variable (PWS adoption) and the selected socio-economic explanatory variables at the census tract (CT) level. Multicollinearity diagnostics performed using the variance inflation factor (VIF) indicated that the selected input variables do not pose a concern of collinearity since all the VIFs are below 3.0 (Midi et al., 2013). Using the PWS point density threshold of 0.1 per km² in the previous section, we assumed that the presence of 1 PWS could provide the minimum rainfall representation for a CT, since the area of a CT in the analyzed MAs is mostly below 10 km². Therefore, the CTs were classified into binary groups of “1 + PWS adoption” and “no PWS” as the response variables in the Logistic Regression models. The summary statistics (Table 2) showed that the percentage of CTs with 1 + PWS adoption ranged from 21% (New York) to 77% (Seattle), and is generally lower in MAs with a larger number of CTs, such as New York, Los Angeles, and Chicago. Notably, the comparison of CTs in “no PWS” and “1 + PWS” groups (Table 2) across MAs shows that the median household income (MHI) and owner-occupied household ratio (OOHR) are significantly higher in the CTs with 1 + PWS, while household density (HHD) is mostly higher in CTs with no

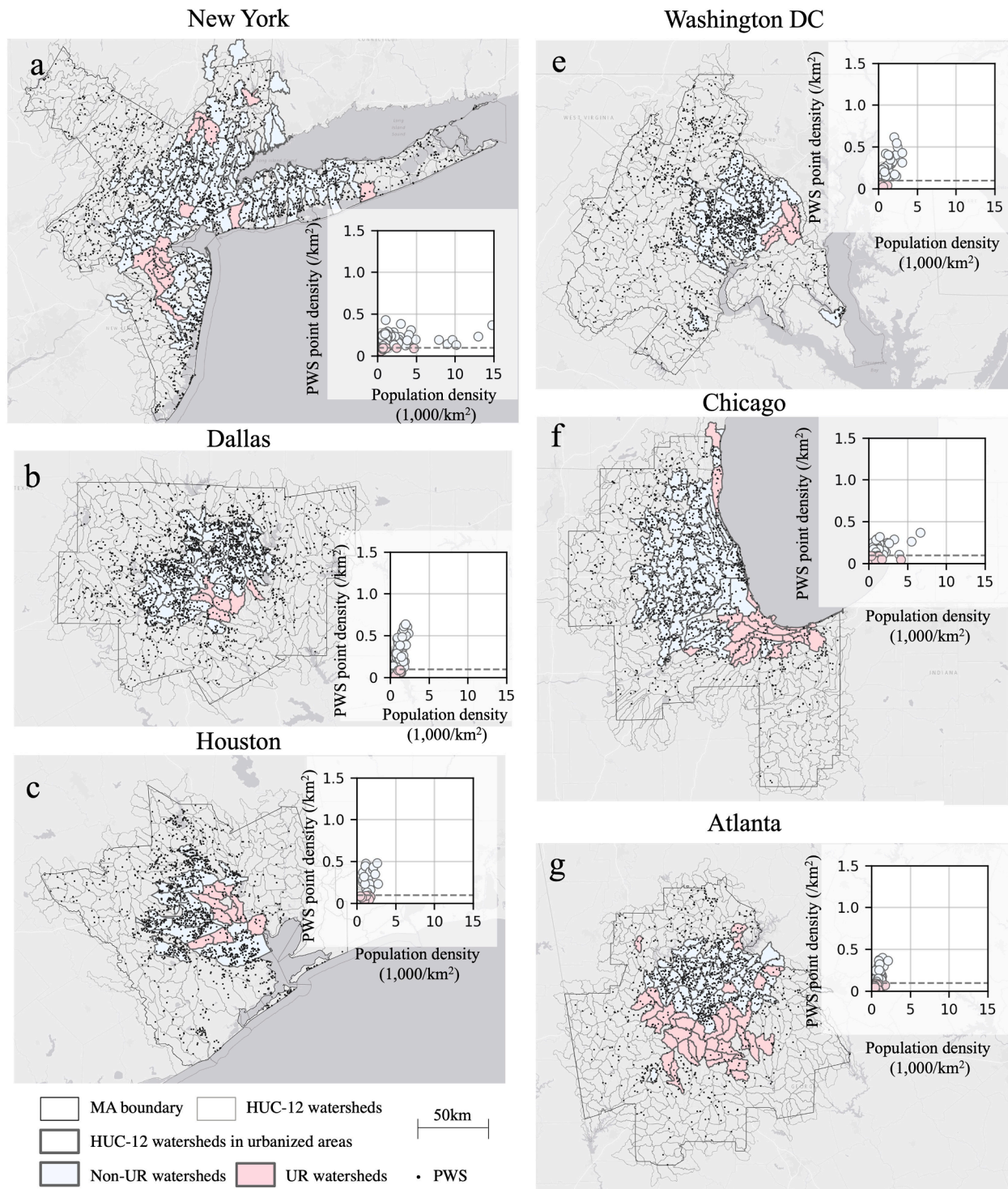


Fig. 3. The PWS adoption pattern in HUC-12 watersheds for the MAs with higher percentage of UR watersheds. The scatterplots of PWS point density and population density show that underrepresentation of PWS adoption can occur in watersheds with similar level of population density as non-UR watersheds.

PWS. The mean population in the two groups was similar, since CTs, by definition, are delineated based on the population.

The exponentiated coefficients (odds ratio) of the logistic regression models are shown in Table 3. Among the input variables to the models, MHI, POP, and OOH have positive effects on PWS adoption (odds ratio greater than 1), while HHD has negative effects (odds ratio less than 1) which indicates negative effects on PWS adoption. The MHI variables were converted into categorical variables using quartile groups. Based on the coefficients of the MHI for the 12 analyzed MAs, the odds ratio of

the fourth MHI quartile group (MHI_Q4) was the highest, followed by the third (MHI_Q3) and second (MHI_Q2) MHI quartile group. Notably, in MAs such as Dallas and Houston, the odds ratio of the third and fourth MHI quartile was significantly greater than first MHI quartile group (MHI_Q1), which indicated that PWS adoptions are much more likely to occur in wealthier neighborhoods. The coefficients of population variables were significant for every analyzed MA in predicting PWS adoption, suggesting that an increase of 1,000 in population could lead to an increase of 9.6% (Houston) – 48.5% (Boston) of the odds ratio that a CT

Table 2
Summary statistics for census-level PWS adoption and the selected explanatory variables.^a

MA	N (census tracts)	N of 1 + PWS	Mean of CTs with no PWS				Mean of CTs with 1 + PWS			
			MHI (\$)	POP	HHD (/km ²)	OOHR	MHI (\$)	POP	HHD (/km ²)	OOHR
New York	3586	758 (21%)	66,521	4,073	5,529	0.41	102,044	4,968	2,178	0.69
Dallas	917	467 (51%)	53,938	4,469	923	0.46	83,949	5,649	612	0.64
Houston	743	353 (48%)	50,488	5,123	890	0.47	83,013	6,499	672	0.63
Los Angeles	2141	516 (24%)	55,801	4,313	1,969	0.40	86,607	4,854	1,232	0.59
Washington DC	924	430 (47%)	86,810	4,014	1,783	0.51	125,384	4,794	1,016	0.71
Seattle	517	396 (77%)	66,789	4,663	1,272	0.49	86,380	5,387	967	0.62
San Francisco	511	262 (51%)	73,320	4,153	3,833	0.38	109,486	4,857	1,937	0.59
Boston	580	291 (50%)	73,701	4,080	2,742	0.43	98,589	5,340	1,134	0.66
Chicago	1797	548 (30%)	57,369	3,781	1,878	0.54	86,831	4,921	957	0.74
Denver	408	242 (59%)	56,757	4,174	1,142	0.50	82,503	4,750	849	0.68
Phoenix	721	366 (51%)	49,714	4,272	840	0.51	67,391	4,697	665	0.65
Atlanta	660	338 (51%)	51,092	4,814	558	0.46	84,032	6,550	419	0.69

^a MHI: median household income (\$); POP: total population; HHD: household density (/km²); HHOR: owner-occupied household ratio.

Table 3
PWS adoption logistic regression model exponentiated coefficients (odds ratio).^{a,b}

MA	Intercept (MHI_Q1)	MHI_Q2	MHI_Q3	MHI_Q4	POP	HHD	OOHR
New York	0.021 **	1.890 **	3.156 **	6.847 **	1.311 **	0.897 **	1.011 **
Dallas	0.136 **	2.568 **	5.274 **	14.306 **	1.279 **	0.495 **	1.001
Houston	0.271 **	2.446 **	6.413 **	20.051 **	1.096 **	0.559 **	0.994
Los Angeles	0.044 **	1.375	3.902 **	10.944 **	1.175 **	0.886	1.003
Washington DC	0.053 **	3.400 **	5.840 **	10.982 **	1.336 **	0.866 **	1.005
Seattle	0.189 **	1.849 *	3.266 **	3.726 **	1.353 **	0.967	1.012
San Francisco	0.145 **	2.042 *	4.116 **	7.691 **	1.212 **	0.864 **	1.009
Boston	0.037 **	2.069 *	2.698 **	3.468 **	1.485 **	0.876 *	1.017 *
Chicago	0.021 **	3.671 **	8.275 **	12.789 **	1.252 **	0.823 **	1.009 *
Denver	0.114 **	1.439	2.279 *	5.387 **	1.231 **	0.920	1.018 *
Phoenix	0.139 **	1.596	2.516 **	3.951 **	1.162 **	0.792	1.014 **
Atlanta	0.042 **	2.478 **	5.879 **	14.749 **	1.253 **	0.965	1.011

^a * indicates 5% significance (p-value < 0.05), and ** indicates 1% significance (p-value < 0.01).

^b MHI: median household income (\$); POP: total population; HHD: household density (/km²); HHOR: owner-occupied household ratio.

has 1 + PWS adoption. The coefficients of OOHR showed that PWS adoption is more likely to occur in CTs with a larger percentage of owner-occupied households, suggesting that an increase of 1% in OOHR in a CT could lead to an increase of 3% – 11% of the 1 + PWS adoption odds ratio. However, PWS adoption is less likely in densely populated CTs. For an increase of 1000 households per km² in HHD in a CT, the odds ratio of PWS adoption could be decreased by 3.5% to 50.5% across analyzed MAs.

3.4. The effects of median household income on PWS adoption

Marginal effects of MHI were calculated to provide an intuitive comparison of MHI effects on PWS adoption across analyzed MAs. A marginal effect (ME) is defined as the change in the response variable associated with a change in one explanatory variable while holding other variables at a specific value. In the PWS adoption logistic regression models, the MEs of MHI demonstrate the discrete change in predicted probability of PWS adoption from the reference category (MHI_Q1) to other categories (MHI_Q2, MHI_Q3 and MHI_Q4), keeping POP, HHD and OOHR at their mean values. The MEs of MHI for the analyzed MAs are shown in Fig. 4a. Based on the results of the ME, common patterns were observed for every analyzed MA. The predicted probability of PWS adoption is always the lowest in the reference category (MHI_Q1), followed by the second (MHI_Q2), third (MHI_Q3). The predicted probability of PWS adoption of the MHI_Q4 is consistently the highest. This indicates that in a hypothetical CT with identical POP, HHD, and OOHR, the probability of PWS adoption is greater in upper MHI quartile categories. For example, for Houston, the ME for the MHI_Q1 category is 19%, while in MHI_Q2, MHI_Q3 and MHI_Q4 the MEs are 34%, 58% and 81%, respectively.

Despite the common pattern that predicted PWS adoption is generally higher in upper MHI quartile groups, the level of PWS adoption disparities due to MHI varied significantly across analyzed MAs. As shown in Fig. 3a, Seattle generally has a higher probability (greater than 66%) of PWS adoption, regardless of the MHI quartile group. However, the adoption probability varied largely in MAs such as Houston (MHI_Q1 = 19%, MHI_Q4 = 81%) and Chicago (MHI_Q1 = 7% and MHI_Q4 = 45%). The ratio of the average MEs of the upper MHI quartile group (MHI_Q2, MHI_Q3, and MHI_Q4) to the MHI_Q1 was computed to quantify the level of disparities. As shown in Fig. 4b, in MAs such as Seattle and Denver, the level of disparities between MHI quartile group are lower (ratio to Q1 were 1.3 and 1.5, respectively), while in MAs such as Houston and Chicago, the level of disparities is much higher (ratio to Q1 were 3.3 and 5.4, respectively). The PWS adoption pattern versus MHI quartiles of these example MAs are shown in Fig. 5.

3.5. The effects of potential flood risk on PWS adoption

In this study, we further theorized that flood risk may influence PWS adoption. Therefore, two types of flood related dataset at the CT level were used to assess the association between potential flood risk and PWS adoption (i) total number of flood claim (FC), and (ii) percent housing units in the 100-year flood zone (FZ). A comparison of FZ and FC data showed that CTs with a lower number of flood claims (*low FC group*) are associated with a lower percentage of CTs that are in the *in FZ group*, while CTs with a higher number of flood claims (*high FC group*) are associated with a higher percentage of CTs that are in the *in FZ group*. As can be seen in Fig. 6, this relationship is consistent for the analyzed MAs, which indicates that CTs that have any percent of occupied housing units in the FEMA 100-year flood plain are more likely to have a greater

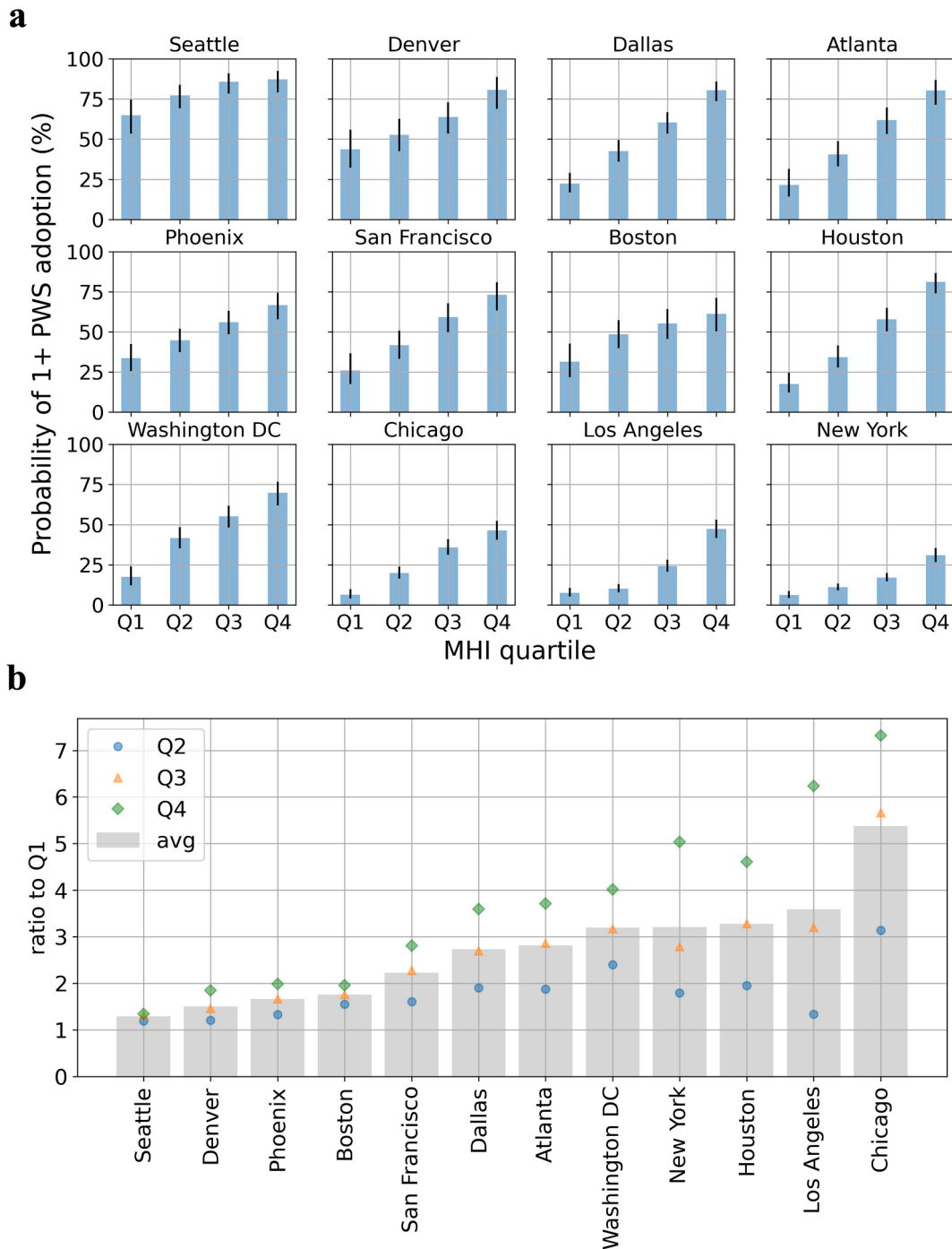


Fig. 4. Marginal effect of median household income (MHI) quartile groups in analyzed MAs. a Predicted PWS adoption probability (the error bars represent 95% confidence interval) of the analyzed MAs (sorted by average probability). b PWS adoption disparities regarding MHI (sorted by average ratio of Q2, Q3 and Q4 to Q1 from low to high).

number of flood claims. For example, in New York, only 4.6% of the CTs in the *low FC* group are in the *In FZ* group, while 59.3% of the CTs in the *high FC* group are also in the *In FZ* group.

Next, we look at the relationship between PWS adoption and the two flood-related variables to investigate whether the vulnerability to flood affects the likelihood of PWS adoption. Logistic regression models were built for the 12 analyzed MAs using socio-economic explanatory variables along with FC and FZ data to assess the potential flood risk affecting the spatial pattern of PWS adoption. Note that these models were built separately from the models in previous section, since the FZ data have larger percentage of CTs with missing data than in analyzed

MAs (100%, 27%, 18% and 17% missing in Seattle, San Francisco, Washington DC, and Chicago, respectively). In addition, since the correlation exists between FC and FZ variable, two separated regression models (one adding FC data and another one adding FZ data) were built. The coefficients for the models adding FC and FZ data are shown in Table 4 and Table 5, respectively. Compared to the reference group of *low FC*, *high FC* groups for most of the analyzed MAs have odds ratio greater than 1.0, indicating positive effects on the PWS adoption. Similarly, compared to the reference group of *Not in FZ* group, the odds ratios for *In FZ* group are greater than 1.0 which also suggesting positive effect on PWS adoption.

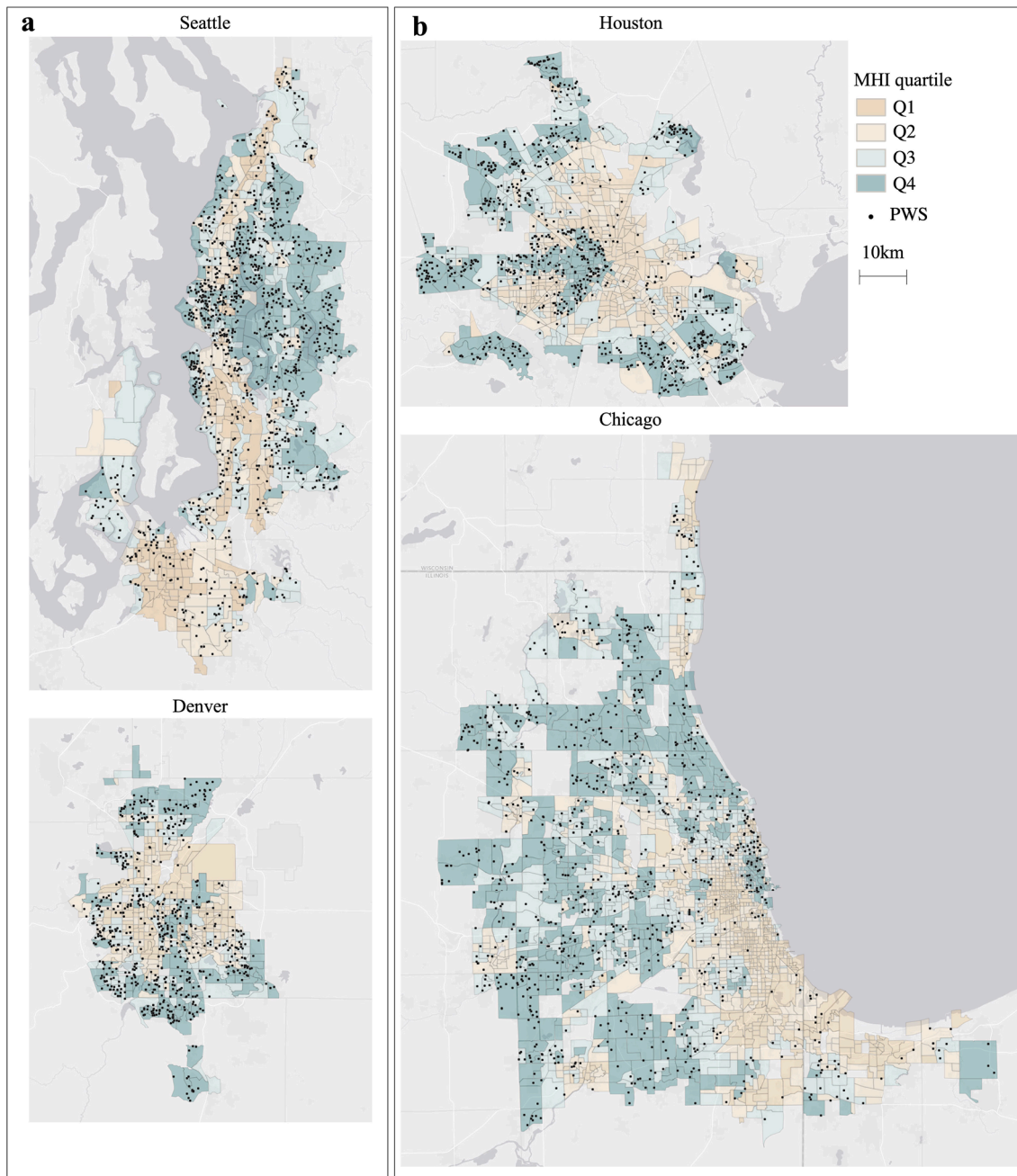


Fig. 5. The spatial pattern of PWS adoption shows clusters in census tracts with higher median household income (MHI). a In the example of Seattle and Denver, the PWS adoption disparities across MHI quartile are less significant, but b in the example of Houston and Chicago, the PWS adoption are more biased toward wealthier census tracts.

We further used Marginal Effects (ME) to demonstrate the effects of flood vulnerability on PWS adoption. Using the ME adjusted for the second quartile MHI group as an example, as shown in Fig. 7, a consistent relationship was found that the MEs of the predicted probability of PWS adoption in *high FC* groups were higher than the MEs in the *low FC* groups (Fig. 7a), which means that CTs that have higher number of flood claims will be more likely to have PWS adoption. Similarly, the MEs of the FZ variable also suggested that CTs that intersected the 100-year floodplain have higher probability of PWS adoption than those not in the floodplain (Fig. 7b). These findings suggest that, with the assumption that the number of flood claims and the percentage of households in the 100-year floodplain represent the flood vulnerability of a CT, current PWS adoption pattern is spatially biased toward flood-

prone regions.

4. Discussion

In this study, we analyzed a large PWS adoption dataset to explore the rainfall representation of crowdsourced PWS in metropolitan areas (MAs) in the contiguous U.S. Consistent with previous literature, PWS adoptions are generally concentrated in populous MA (Muller et al., 2015). However, our analysis showed that PWS adoption exhibited significant spatial biases, which may result in overrepresentation and underrepresentation of crowdsourced rainfall observations across HUC-12 watersheds within a MA.. In the 12 analyzed MAs, the results of logistic regression models revealed that current PWS rainfall

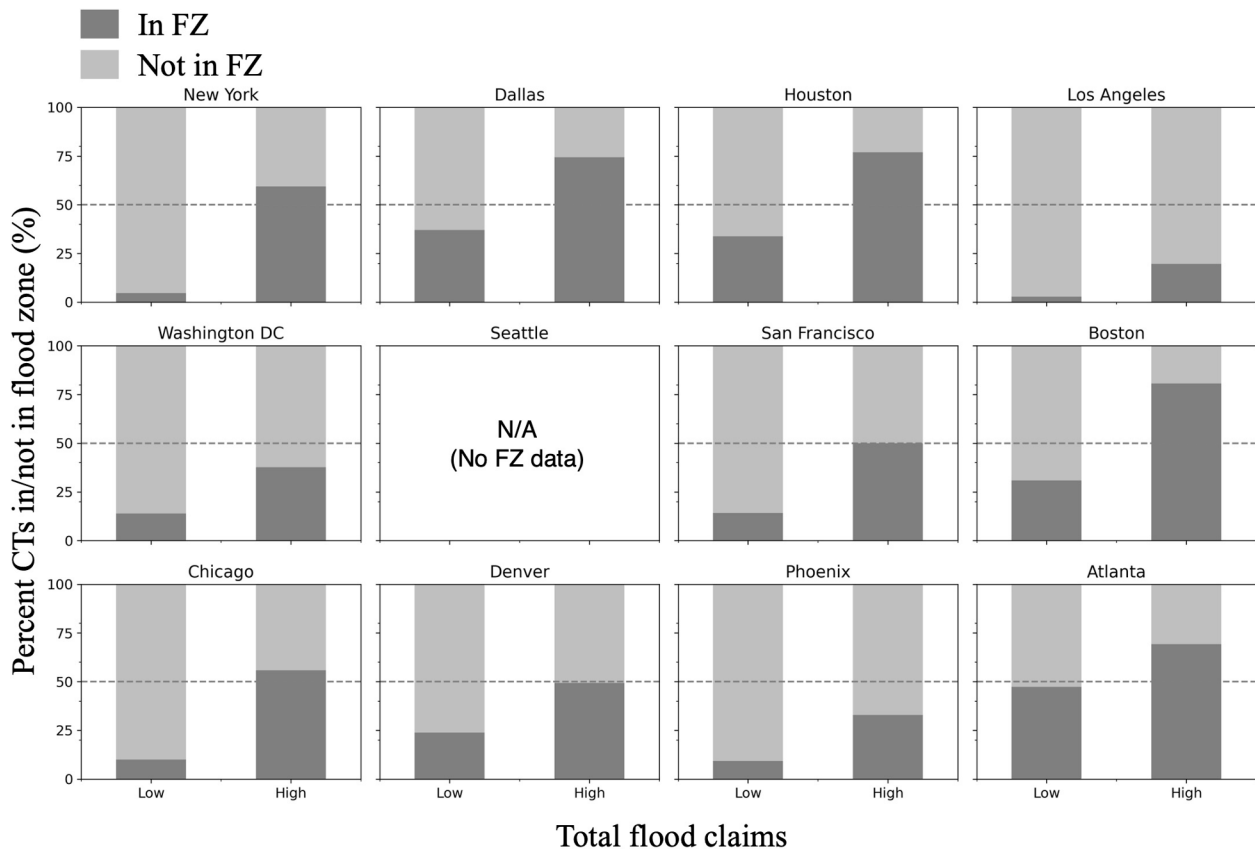


Fig. 6. The relationship between any percent of occupied housing units in/not in 100-year flood zone (FZ) and total NFIP flood claims from 1970 to 2020 (FC). Higher percentage of census tracts with zero percent FZ (in black) are in low FC group (below median FC), while census tracts with greater than zero percent FZ are more likely to be in higher FC group (above median FC).

Table 4
PWS adoption logistic regression model coefficients (odds ratio) adding FC data.^a

MA	Intercept	MHI_Q2	MHI_Q3	MHI_Q4	POP	HHD	OOHR	FZ
New York	0.02 **	1.84 **	2.94 **	6.04 **	1.27 **	0.91 **	1.01 **	1.57 **
Dallas	0.12 **	2.54 **	5.35 **	15.55 **	1.28 **	0.52 **	0.999	1.61 **
Houston	0.23 **	2.37 **	6.20 **	19.85 **	1.09 **	0.59 **	0.993	1.46 *
Los Angeles	0.04 **	1.35	3.70 **	9.53 **	1.17 **	0.89	1.003	1.45 **
Washington DC	0.05 **	3.43 **	5.79 **	10.73 **	1.33 **	0.87 *	1.004	1.25
Seattle	0.19 **	1.83	3.24 **	3.67 **	1.35 **	0.97	1.01	1.12
San Francisco	0.10 **	2.26 **	4.44 **	8.06 **	1.18 **	0.90 *	1.01	2.40 **
Boston	0.04 **	2.11 *	2.71 **	3.50 **	1.47 **	0.88 *	1.02 *	1.20
Chicago	0.02 **	3.64 **	8.16 **	12.67 **	1.24 **	0.84 **	1.01 *	1.27
Denver	0.11 **	1.47	2.32 *	5.37 **	1.22 **	0.93	1.02 *	1.29
Phoenix	0.13 **	1.61	2.52 **	3.96 **	1.16 **	0.80	1.01 **	1.19
Atlanta	0.04 **	2.47 **	5.89 **	14.51 **	1.25 **	0.97	1.01	1.20

^a * indicates 5% significance (p-value < 0.05), and ** indicates 1% significance (p-value < 0.01).

representation is biased toward wealthier neighborhoods and higher flood risk regions due to adoption disparities.

Potential causes of these biases could be because wealthier families have more resources and leisure to participate in PWS adoption, since purchasing and maintaining PWSs are often regarded as a hobby rather than a necessity in the household (Gharesifard and Wehn, 2016). On the other hand, continuous urban development might be driving the population growth in the floodplains, which potentially explains the higher possibility of PWS adoption in floodplains due to increased population (Qiang, 2019). In addition, areas with higher number of flood claims are usually associated with homeownership and higher value homes (Kousky and Michel-Kerjan, 2017), which is the proportion of demographics that are more likely to adopt PWS. However, these are

speculations based on the correlation and regression model results. Future studies are needed to test and better understand causality for these correlations.

Policies could be made to direct the distribution of resources of rainfall data collection efforts in PWS-underrepresented regions. Unlike the non-uniform PWS spatial adoption, traditional rainfall monitoring networks are often designed as uniformly deployed rainfall stations across the watersheds. For example, Harris County Flood Control District (HCFCD) maintains a large rainfall network consisted of ~ 175 rain gauges uniformly distributed across Harris County. As shown in Fig. 8, the point density of PWS and HCFCD comparison shows that in HUC-12 watersheds, the point density of HCFCD maintained network ranges from 0.01 to 0.07, whereas the point density of PWS widely ranges from

Table 5
PWS adoption logistic regression model coefficients (odds ratio) adding FZ variable.^a

MA	Intercept	MHI_Q2	MHI_Q3	MHI_Q4	POP	HHD	OOHR	FZ
New York	0.02 **	1.76 **	2.84 **	5.52 **	1.28 **	0.92 **	1.01 **	1.65 **
Dallas	0.10 **	2.61 **	5.66 **	15.84 **	1.27 **	0.55 **	0.999	1.79 **
Houston	0.24 **	2.40 **	6.13 **	19.71 **	1.10 **	0.59 **	0.994	1.12
Los Angeles	0.05 **	1.35	3.86 **	10.69 **	1.16 **	0.88	1.003	1.32
Washington DC	0.07 **	2.64 **	3.86 **	6.04 **	1.34 **	0.83 **	1.01	1.06
Seattle	–	–	–	–	–	–	–	–
San Francisco	0.10 **	1.76	4.41 **	6.30 **	1.16 *	1.01	1.01	1.43
Boston	0.04 **	2.18 *	2.82 **	4.04 **	1.44 **	0.90	1.01	1.80 **
Chicago	0.02 **	3.75 **	7.89 **	13.60 **	1.22 **	0.86 **	1.01	1.66 **
Denver	0.13 **	1.43	2.29 *	5.52 **	1.21 **	0.91	1.02 *	1.07
Phoenix	0.14 **	1.58	2.50 **	4.02 **	1.15 **	0.80	1.01 **	1.32
Atlanta	0.04 **	2.49 **	5.90 **	14.63 **	1.26 **	0.94	1.01	0.86

^a * indicates 5% significance (p-value < 0.05), and ** indicates 1% significance (p-value < 0.01).

0.01 to 0.48. This observation suggests that to make the best use of the limited public resources, city managers and engineers should consider the spatial pattern of PWS adoption when designing future rainfall monitoring networks. Rather than evenly deploying rain gauges, relocating resources to set up more rain gauges in PWS-underrepresented regions would support the current non-uniform PWS adoption. Moreover, local agencies and organizations could leverage their resources to conduct science, technology, engineering, and mathematics (STEM) related programs and workshop that incentivize the communities in PWS-underrepresented regions to participate in data-driven decision-making processes, and thus increasing their representation in crowdsourced rainfall data (Mondschein et al., 2019).

Incentivizing PWS adoptions to increase participation in crowdsourced data collection could strengthen the awareness of stakeholders for their local environment in the PWS-underrepresented regions. Past examples include the use of volunteered stream monitoring to increase the community's awareness for protecting environmental resources (Overdevest et al., 2004). The application of participative geographic information approaches could strengthen work relations among local actors and authorities to prevent river flooding (Usón et al., 2016). Crowdsourced rainfall data collection in the Community Collaborative Rain, Hail and Snow Network (CoCoRaHS) were shown to have educational benefits that improve the climate literacy of the participants (Reges et al., 2016). Additionally, the increased participation in PWS adoption could improve the usefulness of the crowdsourced rainfall data. Unlike traditional rainfall networks that are maintained and collected by experts with rigorous quality control procedures, the utility of crowdsourced PWS rainfall network is often compromised by its data quality because of limited quality control processes and lack of trust for data contributed from non-experts (Muller et al., 2015). As PWS quality control and trustworthiness assessments methods often require a sufficient number of neighboring PWSs for optimal performance (Chen et al., 2021; de Vos et al., 2019), increased number of PWS adoption in the PWS-underrepresented areas can therefore help ensure the quality and trustworthiness of the crowdsourced data, and thus improve the usefulness of the crowdsourced rainfall data. Future work could also focus on the design of crowdsourcing incentivizing programs for PWS adoption by coupling the regression model results with agent-based modeling to better identify the behavior of crowdsourcing participants and therefore increase adoption (Yang et al., 2019).

Our research merits further exploration in terms of methodology and underlying datasets, which limited the potential of our analytical framework in understanding the spatial biases of crowdsourced environmental data. First, to explore the PWS adoption spatial patterns, we only considered socio-economic factors of population, including median household income, household density and owner-occupied household ratio. Incorporating other demographic characteristics such as the proportion of retired people/students (which potentially form a sizable proportion of the PWS owners) or marginalized/vulnerable groups

(which may likely to have lower uptake) may aid the understanding of the spatial pattern of PWS adoption (Paul et al., 2021). Second, this study focused on analyzing the factors that affect PWS rainfall representation in a census tract. Therefore, binary results of PWS adoption were used for building regression models to assess the socio-economic and flood vulnerability differences in census tracts with adoption or no adoption. Future work could expand the results by building regression models to predict the number of PWSs or per capita PWS to understand the factors that influence the PWS growth (De Groote et al., 2016). Finally, this study focused on the PWS adoption at the census tract level, which may not be adequate to fully capture the factors that affect PWS adoption, since aggregation of data can inflate estimates of association between variables within a CT. Future work could be focused on surveying individual PWS owners at household level to permit understandings for finer level of PWS adoption (Ghahesifard and Wehn, 2016).

Another unexplored question in this study is the difference of adoption rates between the analyzed MAs. For example, Seattle generally has higher probability (>66%) of PWS adoption, regardless of the MHI quartile group, while in Chicago the probability is lower than 45% for every MHI quartile group. This can be caused by some other factors that were not explored in this study. For example, while PWS adoption often occur at the individual household level, large number of PWSs could be adopted by corporations that are used for the benefits for their business. PWS adoption could also be adopted by non-profit organizations (NGOs) that initiate weather monitoring programs. Furthermore, the difference in cost of living across MAs might be explanatory of PWS adopt number difference in MAs. Since PWS are devices that could be easily purchased online with the similar costs, a PWS of \$200 USD, for example, may be affordable in MAs with higher cost of living, while in MAs with lower cost of living this price could be relatively costly.

While most of the past literature on crowdsourced rainfall has focused on the usefulness, quality, and trustworthiness of the observed data, they tended to focus on a single study area (a watershed or a city) of interest (Bardossy et al., 2021; Chen et al., 2021; de Vos et al., 2017, 2019; Mandement and Caumont, 2020). This study, on the other hand, opens a door for spatial pattern analysis of broader crowdsourced data. Using PWS adoption as a case study, our study highlights a phenomenon that generates cross-disciplinary research opportunities between engineers, designers, social sciences, city managers, planners, and NGO groups. Moving forward, alongside the rapid growth of the number of crowdsourced data to support flood assessment efforts, the methodology could be applied to other types of crowdsourced data to assess the potential spatial biases in crowdsourced data collection, enabling crowdsourced networks to be better positioned for decision-makers in their flood resilience efforts.

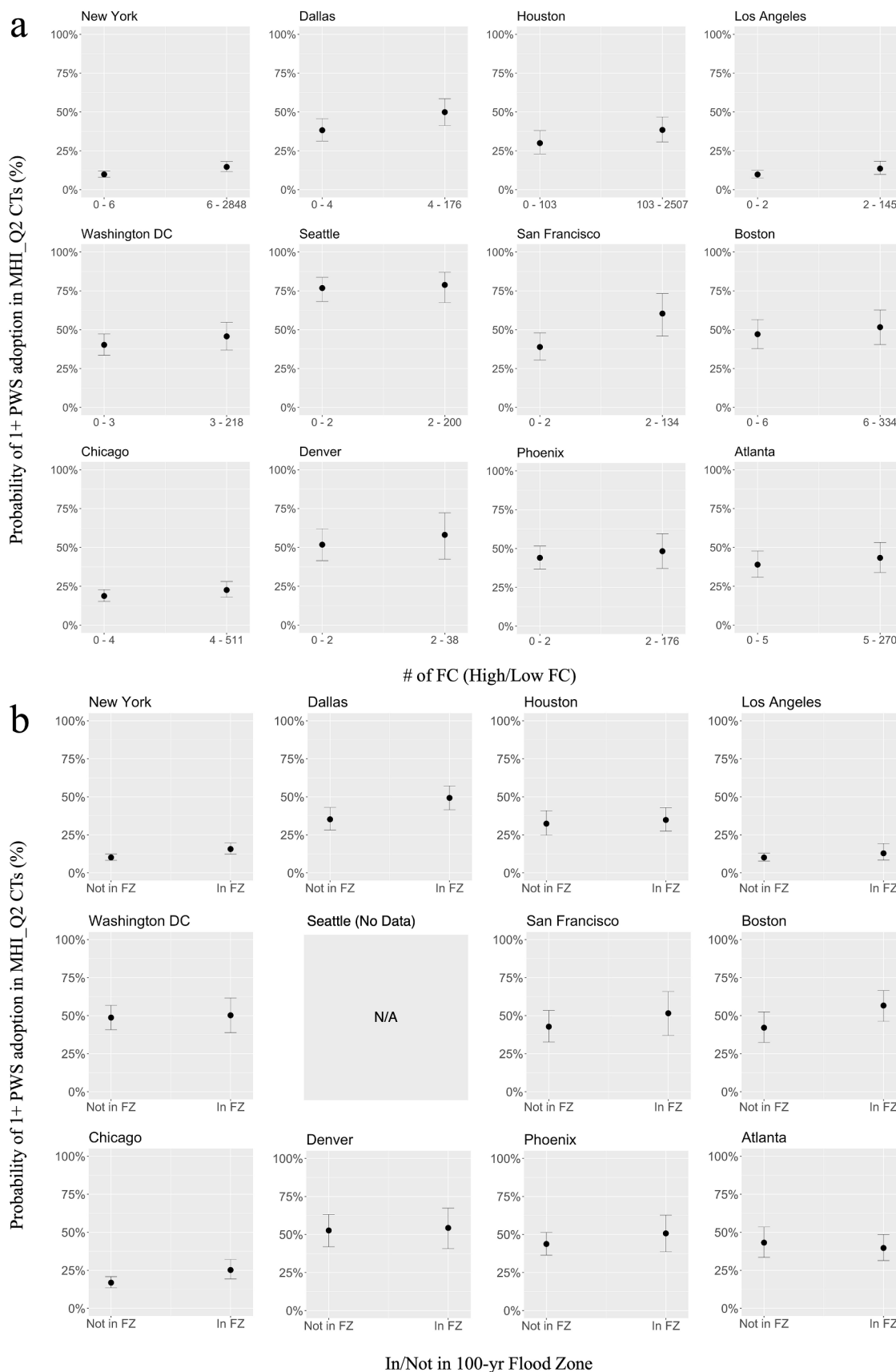


Fig. 7. Marginal effect of the predicted probability of 1 + PWS adoption in MHI_Q2 census tracts of the number of total flood claims (FC) and in/not in flood zone (FZ) for analyzed MAs. a The predicted PWS adoption probability is generally higher in census tracts in the higher FC group. b The probability is also higher in census tracts intersected with the FZ.

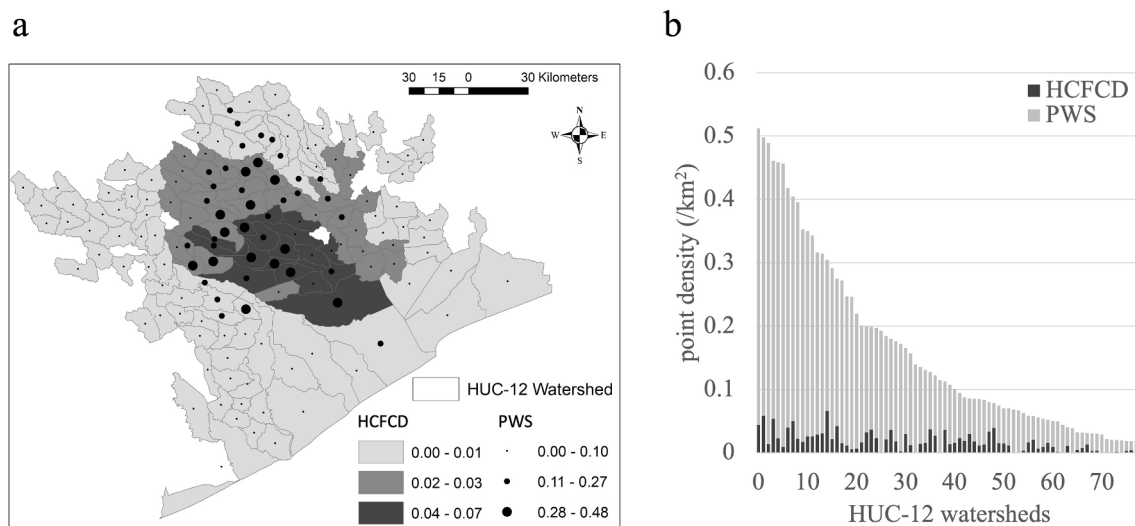


Fig. 8. Comparison of point density of HCFCF and PWS rainfall network in Houston MA. HCFCF rainfall stations are evenly distributed across HUC-12 watersheds, while PWS is spatially biased toward certain watersheds.

5. Conclusion

In this study, we first examined the rainfall spatial representation of PWSs at HUC-12 watersheds in twelve metropolitan areas in the U.S. The results show disparities across the analyzed metropolitan areas, with the percentage of PWS-underrepresented watersheds ranging from 0 to 40%. Furthermore, by modeling PWS adoption using socio-economic and flood-related data at census-tract level, the results suggest that the current PWS adoption pattern exhibits spatial biases toward wealthier neighborhoods and flood-prone regions. The findings provide insights to inform how policies could be made to distribute resources to improve the rainfall data collection efforts in PWS-underrepresented regions. As crowdsourced data are increasingly used for decision-making by policymakers, efforts to close the gap in current non-uniform PWS spatial adoption will allow crowdsourced rainfall data to be better positioned to support decision-makers in their flood resilience efforts.

Data availability

All data used in this study are publicly available. Personal Weather Station adoption data was downloaded from the Weather Underground database using the Weather Underground API (Weather Underground, 2019). HUC-12 watershed boundary was downloaded from USGS National Water Information System (U.S. Geological Survey, 2021). Socio-economic data was downloaded from the United States Census Bureau (U.S. Census Bureau, 2021). Population raster data was downloaded from Worldpop website (Tatem, 2017). NFIP redacted flood claim data is available from OpenFEMA website (Federal Emergency Management Agency, 2021), and the flood zone data was accessed from NYU Furman Center FloodzoneData.us website (NYU Furman Center, 2021).

Author contributions

All authors contributed to conceptualization of the study. A.B.C conducted all the analyses and wrote the initial draft. Z.Z. assisted with data collection and visualization. J.L.G and T.D.C aided in the study design and interpretation of the results. All authors contributed to revising and finalizing the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research is supported by the National Science Foundation under Grant No. CBET-1735587. We gratefully acknowledge Weather Underground, United States Geological Survey, United States Census Bureau, Worldpop, the Federal Emergency Management Agency, and the NYU Furman Center for access to their data.

References

- Almoradie, A., Cortes, V.J., Jonoski, A., 2015. Web-based stakeholder collaboration in flood risk management. *J. Flood Risk Manage.* 8, 19–38. <https://doi.org/10.1111/jfr3.12076>.
- Ando, A.W., Freitas, L.P.C., 2011. Consumer demand for green stormwater management technology in an urban setting: The case of Chicago rain barrels. *Water Resour. Res.* 47 <https://doi.org/10.1029/2011WR011070>.
- Assumpção, T.H., Popescu, I., Jonoski, A., Solomatine, D.P., 2018. Citizen observations contributing to flood modelling: opportunities and challenges. *Hydrol. Earth Syst. Sci.* 22, 1473–1489. <https://doi.org/10.5194/hess-22-1473-2018>.
- Bardossy, A., Seidel, J., El Hachem, A., 2021. The use of personal weather station observations to improve precipitation estimation and interpolation. *Hydrol. Earth Syst. Sci.* 25, 583–601. <https://doi.org/10.5194/hess-25-583-2021>.
- Berne, A., Delrieu, G., Creutin, J.-D., Obled, C., 2004. Temporal and spatial resolution of rainfall measurements required for urban hydrology. *J. Hydrol.* 299, 166–179. <https://doi.org/10.1016/j.jhydrol.2004.08.002>.
- Bertilsson, L., Wiklund, K., de Moura Tebaldi, I., Rezende, O.M., Veról, A.P., Miguez, M. G., 2019. Urban flood resilience – A multi-criteria index to integrate flood resilience into urban planning. *J. Hydrol.* 573, 970–982. <https://doi.org/10.1016/j.jhydrol.2018.06.052>.
- Buytaert, W., Dewulf, A., De Bièvre, B., Clark, J., Hannah, D.M., 2016. Citizen Science for Water Resources Management: Toward Polycentric Monitoring and Governance? *J. Water Resour. Plan. Manag.* 142, 01816002. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000641](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000641).
- Chen, A.B., Behl, M., Goodall, J.L., 2021. Assessing the Trustworthiness of Crowdsourced Rainfall Networks: A Reputation System Approach. *Water Resour. Res.* 57 <https://doi.org/10.1029/2021WR029721>.
- Chen, T.D., Wang, Y., Kockelman, K.M., 2015. Where are the electric vehicles? A spatial model for vehicle-choice count data. *J. Transp. Geogr.* 43, 181–188. <https://doi.org/10.1016/j.jtrangeo.2015.02.005>.
- Cheng, L., AghaKouchak, A., 2015. Nonstationary Precipitation Intensity-Duration-Frequency Curves for Infrastructure Design in a Changing Climate. *Sci. Rep.* 4 (1).
- Craig, L., Craig, N., Calgaro, E., Dominey-Howes, D., Johnson, K., 2019. People with disabilities: Becoming agents of change in Disaster Risk Reduction. *Emerg. Voices Nat. Hazards Res.* 327–356 <https://doi.org/10.1016/B978-0-12-815821-0.00020-5>.
- Cristiano, E., Veldhuis, M.C.T., Van De Giesen, N., 2017. Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas - A review. *Hydrol. Earth Syst. Sci.* 21, 3859–3878. <https://doi.org/10.5194/Hess-21-3859-2017>.

- De Groote, O., Pepermans, G., Verboven, F., 2016. Heterogeneity in the adoption of photovoltaic systems in Flanders. *Energy Econ.* 59, 45–57. <https://doi.org/10.1016/J.ENECO.2016.07.008>.
- de Vos, L., Leijnse, H., Overeem, A., Uijlenhoet, R., 2017. The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam. *Hydrol. Earth Syst. Sci.* 21, 765–777. <https://doi.org/10.5194/hess-21-765-2017>.
- de Vos, L.W., Leijnse, H., Overeem, A., Uijlenhoet, R., 2019. Quality Control for Crowdsourced Personal Weather Stations to Enable Operational Rainfall Monitoring. *Geophys. Res. Lett.* 46, 8820–8829. <https://doi.org/10.1029/2019GL083731>.
- Fan, C., Esparza, M., Dargin, J., Wu, F., Oztekin, B., Mostafavi, A., 2020. Spatial biases in crowdsourced data: Social media content attention concentrates on populous areas in disasters. *Comput. Environ. Urban Syst.* 83, 101514 <https://doi.org/10.1016/j.compenvurbysys.2020.101514>.
- Federal Emergency Management Agency, 2021. OpenFEMA Dataset: FIMA NFIP Redacted Claims | FEMA.gov [WWW Document]. URL <https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims> (accessed 6.4.21).
- Fowler, H.J., Wasko, C., Prein, A.F., 2021. Intensification of short-duration rainfall extremes and implications for flood risk: current state of the art and future directions. *Philos. Trans. A. Math. Phys. Eng. Sci.* 379, 20190541. <https://doi.org/10.1098/rsta.2019.0541>.
- Gebremedhin, E.T., Basco-Carrera, L., Jonoski, A., Iliffe, M., Winsemius, H., 2020. Crowdsourcing and interactive modelling for urban flood management. *J. Flood Risk Manag.* 13, e12602 <https://doi.org/10.1111/jfr3.12602>.
- Gharesifard, M., Wehn, U., 2016. To share or not to share: Drivers and barriers for sharing data via online amateur weather networks. *J. Hydrol.* 535, 181–190. <https://doi.org/10.1016/j.jhydrol.2016.01.036>.
- Hu, R., Fang, F., Pain, C.C., Navon, I.M., 2019. Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. 10.1016/j.jhydrol.2019.05.087.
- Kiaghadi, A., Rifai, H.S., Dawson, C.N., 2021. The presence of Superfund sites as a determinant of life expectancy in the United States. *Nat. Commun.* 12, 1–12. <https://doi.org/10.1038/s41467-021-22249-2>.
- Kousky, C., Michel-Kerjan, E., 2017. Examining Flood Insurance Claims in the United States: Six Key Findings. *J. Risk Insur.* 84, 819–850. <https://doi.org/10.1111/JORI.12106>.
- Mandement, M., Caumont, O., 2020. Contribution of personal weather stations to the observation of deep-convection features near the ground. *Nat. Hazards Earth Syst. Sci.* 20, 299–322. <https://doi.org/10.5194/nhess-20-299-2020>.
- Midi, H., Sarkar, S.K., Rana, S., 2013. Collinearity diagnostics of binary logistic regression model. *J. Interdiscip. Math.* 13, 253–267. <https://doi.org/10.1080/09720502.2010.10700699>.
- Millar, E.E., Hazell, E.C., Melles, S.J., 2019. The ‘cottage effect’ in citizen science? Spatial bias in aquatic monitoring programs. *Int. J. Geogr. Inf. Sci.* 33, 1612–1632. <https://doi.org/10.1080/13658816.2018.1423686>.
- Mondschein, A., Zhang, Z., Khafif, M. El, 2019. Community-Centered Urban Sensing: Smart, Engaged Planning and Design in a Dysfunctional Urban Context. *Int. J. E-Planning Res.*
- Mosavi, A., Ozturk, P., Chau, K.W., 2018. Flood prediction using machine learning models: Literature review. *Water (Switzerland)*. <https://doi.org/10.3390/w10111536>.
- Muller, C.L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., Leigh, R.R., 2015. Crowdsourcing for climate and atmospheric sciences: Current status and future potential. *Int. J. Climatol.* 35 (11), 3185–3203.
- NYU Furman Center, 2021. FloodzoneData.us [WWW Document]. URL [FloodzoneData.us](https://www.floodzonedata.us) (accessed 6.4.21).
- Overdevest, C., Huyck Orr, C., Stepenuck, K., 2004. Volunteer Stream Monitoring and Local Participation in Natural Resource Issues, *Source. Human Ecology Rev.*
- Paul, J.D., Bee, E., Budimir, M., 2021. Mobile phone technologies for disaster risk reduction. *Clim. Risk Manag.* 32, 100296 <https://doi.org/10.1016/J.CRM.2021.100296>.
- Paul, J.D., Buytaert, W., Allen, S., Ballesteros-Cánovas, J.A., Bhusal, J., Cieslik, K., Clark, J., Dugar, S., Hannah, D.M., Stoffel, M., Dewulf, A., Dhital, M.R., Liu, W., Nayaval, J.L., Neupane, B., Schiller, A., Smith, P.J., Supper, R., 2018. Citizen science for hydrological risk reduction and resilience building. *Wiley Interdiscip. Rev. Water* 5, e1262. <https://doi.org/10.1002/wat2.1262>.
- Praharaj, S., Chen, T.D., Zahura, F.T., Behl, M., Goodall, J.L., 2021. Estimating impacts of recurring flooding on roadway networks: a Norfolk. *Nat. Hazards* 107 (3), 2363–2387.
- Qiang, Y., 2019. Disparities of population exposed to flood hazards in the United States. *J. Environ. Manage.* 232, 295–304. <https://doi.org/10.1016/J.JENVMAN.2018.11.039>.
- Reges, H.W., Doesken, N., Turner, J., Newman, N., Bergantino, A., Schwalbe, Z., 2016. CoCoRaHS: The evolution and accomplishments of a volunteer rain gauge network. *Bull. Am. Meteorol. Soc.* 97, 1831–1846. <https://doi.org/10.1175/BAMS-D-14-00213.1>.
- Rosenzweig, B.R., McPhillips, L., Chang, H., Cheng, C., Welty, C., Matsler, M., Iwaniec, D., Davidson, C.I., 2018. Pluvial flood risk and opportunities for resilience. *WIREs Water* 5 (6). <https://doi.org/10.1002/wat2.1302>.
- Roy, A., Nelson, T.A., Fotheringham, A.S., Winters, M., 2019. Correcting Bias in Crowdsourced Data to Map Bicycle Ridership of All Bicyclists. *Urban Sci.* 3, 62. <https://doi.org/10.3390/urbansci3020062>.
- Sadler, J.M., Goodall, J.L., Morsy, M.M., Spencer, K., 2018. Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. *J. Hydrol.* 559, 43–55. <https://doi.org/10.1016/J.JHYDROL.2018.01.044>.
- Shen, Y., Morsy, M.M., Huxley, C., Tahvildari, N., Goodall, J.L., 2019. Flood risk assessment and increased resilience for coastal urban watersheds under the combined impact of storm tide and heavy rainfall. *J. Hydrol.* 579, 124159 <https://doi.org/10.1016/j.jhydrol.2019.124159>.
- Sy, B., Frischknecht, C., Dao, H., Consuegra, D., Giuliani, G., 2019. Flood hazard assessment and the role of citizen science. *J. Flood Risk Manag.* 12, e12519 <https://doi.org/10.1111/jfr3.12519>.
- Tatem, A.J., 2017. WorldPop, open data for spatial demography. *Sci. Data* 4, 1–4. <https://doi.org/10.1038/sdata.2017.4>.
- The National Academies of Science, Engineering, and M., 2019. Framing the Challenge of Urban Flooding in the United States, Framing the Challenge of Urban Flooding in the United States. National Academies Press. 10.17226/25381.
- Torralla, A., Efron, A.A., 2011. Unbiased look at dataset bias. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 1521–1528. 10.1109/CVPR.2011.5995347.
- Towe, R., Dean, G., Edwards, L., Nundloll, V., Blair, G., Lamb, R., Hankin, B., Manson, S., 2020. Rethinking data-driven decision support in flood risk management for a big data age. *J. Flood Risk Manag.* 13, e12652 <https://doi.org/10.1111/jfr3.12652>.
- Tyler, J., Sadiq, A.-A., Douglas, -, Noonan, S., Noonan, D.S., 2019. A review of the community flood risk management literature in the USA: lessons for improving community resilience to floods 96, 1223–1248. 10.1007/s11069-019-03606-3.
- U.S. Census Bureau, 2021. Metropolitan and Micropolitan [WWW Document]. URL <https://www.census.gov/programs-surveys/metro-micro.html> (accessed 4.20.21).
- U.S. Geological Survey, 2021. Hydrologic Unit Maps [WWW Document]. URL <https://water.usgs.gov/GIS/huc.html> (accessed 4.20.21).
- Usón, T.J., Klonner, C., Höfle, B., 2016. Using participatory geographic approaches for urban flood risk in Santiago de Chile: Insights from a governance analysis. *Environ. Sci. Policy* 66, 62–72. <https://doi.org/10.1016/j.envsci.2016.08.002>.
- van Oldenborgh, G.J., van der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., Hausteijn, K., Li, S., Vecchi, G., Cullen, H., 2017. Attribution of extreme rainfall from Hurricane Harvey, August 2017. *Environ. Res. Lett.* 12 (12), 124009.
- Voinov, A., Kolagani, N., McCall, M.K., Glynn, P.D., Kragt, M.E., Ostermann, F.O., Pierce, S.A., Ramu, P., 2016. Modelling with stakeholders - Next generation. *Environ. Model. Softw.* 77, 196–220.
- Wang, R.Q., Mao, H., Wang, Y., Rae, C., Shaw, W., 2018. Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data. *Comput. Geosci.* 111, 139–147. <https://doi.org/10.1016/j.cageo.2017.11.008>.
- Underground, W., 2019. Personal Weather Station Network [WWW Document]. accessed 8.28.19. <https://www.wunderground.com/weatherstation/overview.asp>.
- White, I., Kingston, R., Barker, A., 2010. Participatory geographic information systems and public engagement within flood risk management. *J. Flood Risk Manag.* 3, 337–346. <https://doi.org/10.1111/j.1753-318X.2010.01083.x>.
- World Meteorological Organization (WMO), 2008. Guide to Hydrological Practices, Volume I, WMO. WMO, Geneva.
- Yang, P., Ng, T.L., Cai, X., 2019. Reward-Based Participant Management for Crowdsourcing Rainfall Monitoring: An Agent-Based Model Simulation. *Water Resour. Res.* 55, 8122–8141. <https://doi.org/10.1029/2018WR024447>.
- Zahura, F.T., Goodall, J.L., Sadler, J.M., Shen, Y., Morsy, M.M., Behl, M., 2020. Training Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model: Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal Community. *Water Resour. Res.* 56 <https://doi.org/10.1029/2019WR027038>.