



Research papers

Exploring the relationship between flood insurance claims, crowdsourced rainfall, and tide levels for coastal urban communities: Case study for the mid-Atlantic United States

Alexander B. Chen^{1,*}, Jonathan L. Goodall, Julianne D. Quinn

Department of Civil and Environmental Engineering, University of Virginia, 151 Engineer's Way, Charlottesville, VA 22904, USA

ABSTRACT

The Federal Emergency Management Agency (FEMA) recently released a large dataset (~2.5 million records) of flood insurance claim transactions from the National Flood Insurance Program (NFIP) that are geolocated by census tract. In this study, we present a methodology for using these NFIP flood claims along with agency rainfall and tide level data, combined with high-resolution rainfall data collected from growing networks of crowdsourced Personal Weather Stations (PWSs), to investigate the characteristics of storm events that result in flood insurance claims. Using 25 cities/counties in five metropolitan areas in the Mid-Atlantic region as case studies for testing the methodology, we find that the majority of flood claim events in this region were associated with rainfall below a 1-year return period and a tide level below the NOAA minor flood risk tide level. When storm events exceeded these rainfall and tide thresholds, the probability of a flood claim occurring could reach 100% for rainfall only events and 33% for tidal only events, depending on the region. While compound events of both high rainfall and high tide were rare, when seen in Virginia Beach, the probability of insurance claims being made during these events far exceeded the probability that would be expected if rainfall and tide were independent, showing how compound events in this region exacerbate flood impacts. Additionally, analysis of 110 heavy rainfall events showed that crowdsourced PWSs, can better capture rainfall extremes associated with insurance claims due to their higher spatial density compared to federal agency rainfall networks. This suggests that the PWS networks, once quality controlled, can offer greater insights into which rainfall events are likely to contribute to insurance claims. These region-specific findings and the general methodology presented in this study can benefit multiple stakeholders including researchers, risk management officials, insurers as well as the general public in understanding the impacts of rainfall, tidal, and compound flooding, along with the value of crowdsourced data in assessing flood risk.

1. Introduction

Flooding is a leading cause of social and economic losses in the United States (NOAA National Centers for Environmental Information (NCEI), 2022). In coastal cities, floods can be triggered by heavy rainfall and high tide due to low-lying topography and impervious surfaces (Hallegatte et al., 2013). However, due to the effects of climate change and sea level rise, flooding has become more frequent and severe (Ezer and Atkinson, 2014; Fowler et al., 2021). For example, recent studies have shown that there has been an increase in both the frequency and intensity of precipitation trends in the United States, which will likely worsen the impacts from flooding (Davenport et al., 2021; Mosavi et al., 2018; Netusil and Kousky, 2021). Moreover, sea level rise has also caused increased flood risks even for low-return period storm events (Moftakhari et al., 2015). Increased urbanization is another factor that increases potential flood hazard and exposure due to impervious surfaces (Chang et al., 2021).

Understanding the characteristics of rainfall and tide that lead to flooding is a critical step toward effectively managing and building resilience against flood risk (Nofal and van de Lindt, 2020; Shen et al., 2019). This is especially important in areas that lack flooding observational data to calibrate complex flood prediction and drainage models (Gaitan et al., 2016). However, due to insufficient spatial, temporal, and geographic information regarding historical floods, past studies of flood characteristics have focused on single case studies or a limited database of flood events (Saharia et al., 2017). Recently, flood insurance claims have emerged as a potential data source of flood impacts (Gradedi et al., 2019). For example, outside the United States, flood insurance data has been used to investigate the relationship between extreme rainfall event and flood losses (Leal et al., 2019; Sørensen and Mobini, 2017; Spekkers et al., 2013; Torgersen et al., 2015). The spatial relationships between flood insurance claims and landcover have been explored to understand flood risk. Studies have also shown that flood insurance claim data can be used to inform the calibration of flood models (Wang and Sebastian,

* Corresponding author.

E-mail address: abc8fq@virginia.edu (A.B. Chen).

¹ Now with the California Nevada River Forecast Center, NOAA National Weather Service, 3310 El Camino Avenue, Sacramento, CA 95821, USA.

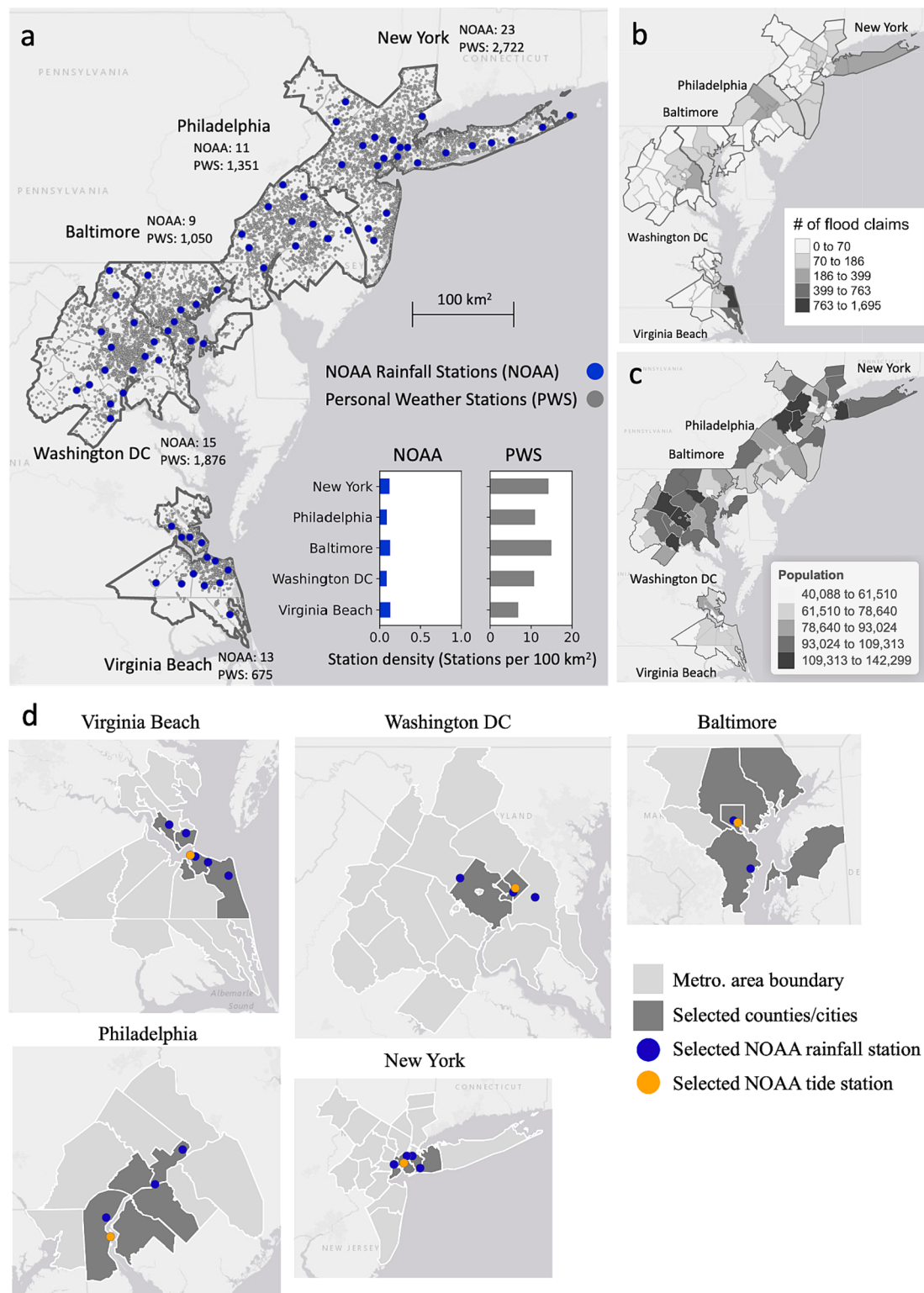


Fig. 1. (a) Comparison of NOAA rainfall network and crowdsourced PWS rainfall network for the five metropolitan areas. (b) The number of claims by county. (c) Population by county, (d) Selected NOAA rainfall and tide stations and crowdsourced PWS rainfall network in the analyzed five metropolitan areas.

2021; Zischg et al., 2018).

In the United States, the Federal Emergency Management Agency's (FEMA) National Flood Insurance Program (NFIP) is the largest provider of flood insurance policies. In the past, the NFIP flood insurance claim dataset was not publicly available. Therefore, there are only a few studies using NFIP flood insurance claims that either focus on policy aspects (i.e., supply, demand and effectiveness of the NFIP program)

(Kousky, 2018; Kousky and Michel-Kerjan, 2017) or assessing flood risk for certain study areas in case studies (Blessing et al., 2017; Czajkowski et al., 2017; Mobley et al., 2021). In June 2019, in an effort to promote FEMA's open data vision (OpenFEMA) (Federal Emergency Management Agency (FEMA), 2022), FEMA made more than two million NFIP flood claim transactions (1978 to present) available to the public in an easily-accessible and machine-readable format (Dombrowski et al.,

Table 1

NFIP flood claim information for the selected metropolitan areas (2016 – 2020). The number (percentage) of days with claims were calculated from the raw flood claim data.

Metropolitan Area	Population('000)	Number offlood claims	Total Payout (\$K)	Number (percentage) of days with claims
Virginia Beach (VB)	1,731	4,045	91,829	337 (18%)
Washington DC (DC)	6,176	1,591	21,284	382 (21%)
Baltimore (BAL)	2,797	686	22,200	198 (11%)
Philadelphia (PHL)	6,079	2,494	69,052	245 (13%)
New York (NY)	19,294	2,848	42,786	668 (37%)

2021; Federal Emergency Management Agency (FEMA), 2021). Since the release, more studies have emerged to leverage this dataset for flood studies. For example, prediction models for flood insurance claims and flood damages have been explored to construct the relationship between flooding and socioeconomic, geographic, demographic and environmental factors (Collins et al., 2022; Mobley et al., 2021; Vishnu et al., 2021; Wing et al., 2020; Yang et al., 2022). However, there is a lack of studies that analyze how rainfall and tide characteristics in coastal cities related to flood claims using this newly released NFIP flood claim dataset.

It is possible to gain insight into how rainfall and tide characteristics correspond to flood losses in coastal cities through the use of the NFIP flood claims dataset. However, in the case of rainfall, high-resolution and dense rainfall observation networks will be needed to capture weather events that result in localized flooding (Maier et al., 2020). While there are agency (government-operated) in situ rainfall monitoring networks and radar-based methods for measuring rainfall, these approaches alone do not produce the resolution of rainfall data needed to fully capture the storm event characteristics that lead to localized urban flooding (Emmanuel et al., 2012; Smith et al., 2007). The recent growth of crowdsourcing rainfall through the use of Personal Weather Stations (PWSs) has been shown to have the potential to fill in observation gaps and provide high-resolution and reliable data to supplement urban flood monitoring (Bardossy et al., 2021; Chen et al., 2022, 2021; de Vos et al., 2019, 2017; Mandement and Caumont, 2020). Although crowdsourced rainfall data have been utilized in several flooding-related studies, they have not been applied to understand flood losses.

The objective of this study, therefore, is to address these research gaps by creating a method to assess how rainfall and tide characteristics of storm events relate to the NFIP flood claims using exploratory data analysis, and second exploring if crowdsourced rainfall data collected by growing networks of Personal Weather Stations (PWSs) better captures rainfall that leads to flood losses. We use regions in the mid-Atlantic United States as test cases for applying the method and exploring the PWS data. The contributions of this research are both the methodology for analyzing the drivers of flood insurance claims in coastal regions that could be applied to other regions and time periods using this national-scale dataset, as well as region-specific findings of which hydrologic conditions lead to flood insurance losses in the Mid-Atlantic.

2. Materials and methods

2.1. Data

2.1.1. Flood claim data

The flood insurance claims data used in this study were collected from the OpenFEMA website (<https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v1>), released by the Federal Emergency Management Agency's (FEMA) National Flood Insurance Program (NFIP), the largest provider of flood insurance in the U.S. (Federal Emergency Management Agency (FEMA), 2021). The NFIP provides flood insurance to property owners, renters and businesses. The policy holders file claims when their properties are flooded. The policy and claims information are then documented in the NFIP system of record. This dataset provides claim-level fields for each claim transaction

including more than 40 attributes, such as the date of loss (date on which water first entered the insured building), location (geocoded to census tract), insurance payout (dollar amount paid on building and contents claim), and other claim and building property related characteristics.

2.1.2. Agency rainfall and tide data

The agency rainfall and tide data used in this study were collected from the National Oceanic and Atmospheric Administration (NOAA)'s National Centers for Environmental Information (NCEI). The available rainfall and tide stations, as well as their observed data, were queried and downloaded using the *rnoaa* library in the R programming language (R package version 1.3.4) (Chamberlain, 2021), which utilizes NCDC's Application Programming Interface (API) that provides access to NCDC's database (<https://www.ncdc.noaa.gov/cdo-web/webservices/v2>), as well as the NOAA Tide & Currents CO-OPS API (<https://api.tidesandcurrents.noaa.gov/api/prod/>). For the rainfall data, the NOAA Local Climatological Data stations (referred as NOAA rainfall stations hereafter) were queried automatically using NCDC's API. For the tide data, the available tide stations and data stored in the Tide & Current Product database were queried and downloaded using the CO-OPS API (<https://api.tidesandcurrents.noaa.gov/api/prod/>).

2.1.3. Crowdsourced rainfall data

The crowdsourced rainfall network used in this study consists of PWSs available through the Weather Underground, one of the largest platforms for PWS owners to share their data. The rainfall observations from the PWSs were accessed through the API provided by the Weather Underground. The PWS rainfall observation sampling interval varies from station to station. Most of the sampling intervals are between 5 and 10 min per observation.

2.2. Study area

Five metropolitan areas (Virginia Beach, Washington DC, Baltimore, Philadelphia, and New York) (Fig. 1) in the Mid-Atlantic region were selected as case study regions given their susceptibility to flooding due to their population density, low-lying topographic characteristics, rising sea level and increasing rainfall intensities (Ezer and Atkinson, 2014; Goddard et al., 2015; Wahl et al., 2015).

For these regions, agency (NOAA) rainfall and tide data and crowdsourced (PWS) rainfall data were downloaded using the methods mentioned in the previous section. Fig. 1a shows the available NOAA rainfall stations and PWSs within the five selected metropolitan areas based on the API queries. The rainfall station density of NOAA rainfall stations ranged from 0.1 to 0.2 stations per 100 km², while the density of PWSs ranged from 5 to 15 stations per 100 km². Based on the Weather Underground database archive, PWSs have been growing exponentially in recent years (Chen et al., 2022). Therefore, to assess the potential of PWS rainfall observations to understand flood claim events, this study focused on analyzing the flood claim data from 2016 to 2020 for all five metropolitan areas. Fig. 1b and 1c show the number of NFIP flood claims and the population for the selected metropolitan areas at county/city level from 2016 to 2020. As shown in Table 1, the number of flood claims ranges from 686 to 4045 per metropolitan region. The payout ranges from 2.22 million to up to 9.18 million USD.

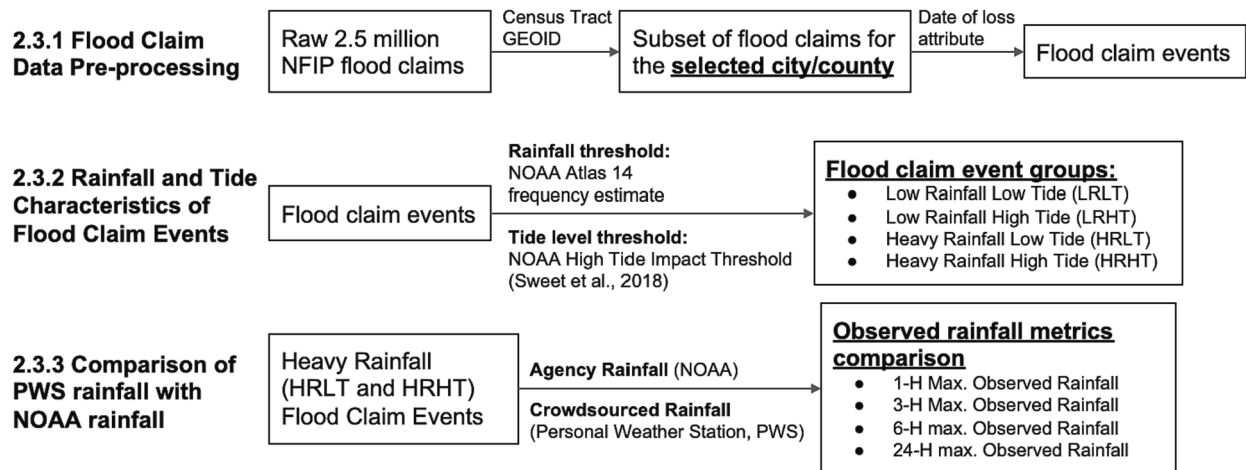


Fig. 2. Workflow of data preprocessing and analysis methods used in this study.

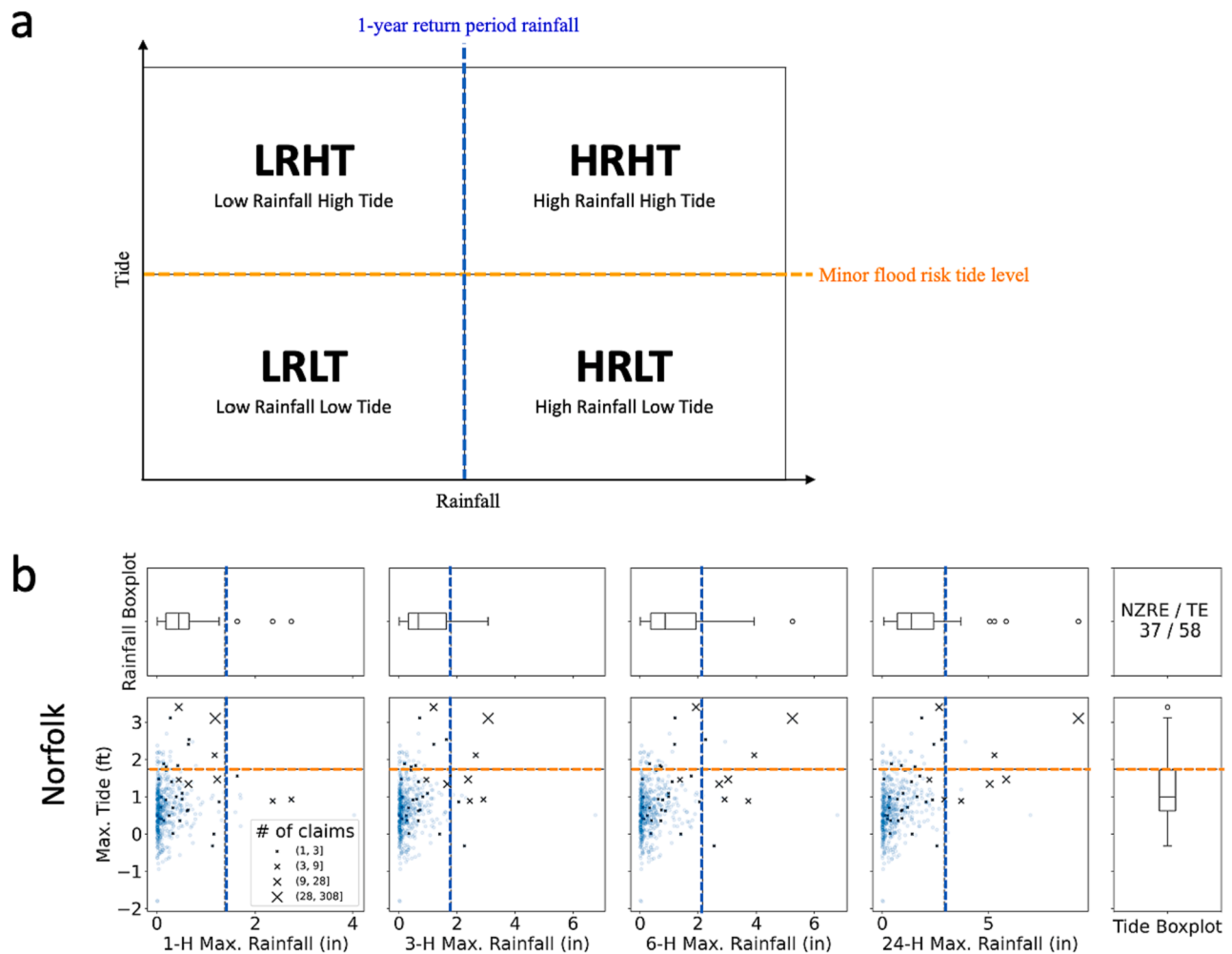


Fig. 3. (a) Flood claim events are categorized into groups based on the rainfall and tide thresholds. (b) Analysis of rainfall and tide characteristics for flood claim events using Norfolk as an example. Blue dots represent non-flood-claim events while black “X”s represent flood claim events with non-zero rainfall. The size of the symbol “X” corresponds to the number of claims. The vertical and horizontal dashed lines represent the rainfall and tide thresholds defined in Fig. 3a, respectively. NZRE and TE represent non-zero rainfall events and total events, respectively. The comparison plots for the four other counties in the Virginia Beach metropolitan area are shown in Fig. A1 in the Appendix.

To account for the effect of tide level on flood losses, we first selected one representative tide station for each metropolitan area that has the most consistent period of record (Fig. 1d). Next, assuming areas that are

closer to the representative tide station have higher impacts from tide water, we chose five counties/cities for each selected metropolitan area that (1) are the closest to the tide station and (2) touch the waterbody

that the representative tide station observes. For the representative rainfall station, we chose the NOAA rainfall station that is closest to the centroid of each selected county/city (Fig. 1d).

2.3. Methods

The methodology presented in this study can be summarized as a workflow of data pre-processing and additional steps used to investigate the relationship between rainfall, tide level and flood losses (Fig. 2). The details of each step are presented in the following subsections.

2.3.1. Flood claim data pre-processing

The raw flood claim data (claim level) was converted into flood claim events (event level) to associate claims with the observed rainfall and tide level using the following steps. For each city/county, we first subset the flood claims that occurred inside the census tracts of the selected city/county. We then sorted the subset data using the date of loss field. Flood claims that occurred only on a single date with no claims on the preceding and succeeding day were regarded as separate flood claim events. On the other hand, flood claims that spanned across multiple consecutive days were aggregated into a single flood claim event. In this case, the start and the end date of such an event was considered the first and the last date of these consecutive days, respectively. This is usually the case for typhoons or hurricanes which have prolonged flood impacts. In addition, we also expect flood risk from typhoons or hurricanes to be encompassed within the rainfall and tide level data. The tide level includes storm surge that can be caused by such events, and the rainfall associated with these events should also be captured by the rainfall gauges.

2.3.2. Analysis of rainfall and tide characteristics of flood claim events

In this study, we used the maximum 1-H, 3-H, 6-H, and 24-H accumulated rainfall and maximum tide level observed during the flood claim event period as the rainfall and tide metrics, respectively. Since different rain gauges may classify these events differently, for consistency the rainfall and tide metrics of each event were calculated from the representative NOAA rainfall and tide station for each city/county. NOAA rainfall stations were used instead of PWS rain stations because of their greater quality control, but data from these networks were later compared under different events to see if the PWSs could capture greater spatial variability in rainfall that might explain surprising flood claim distributions.

To assess the rainfall characteristics of flood claim events, we compared the rainfall metrics of the flood claim events with the 1-year return period of maximum 1-H, 3-H, 6-H and 24-H accumulated rainfall derived from NOAA Atlas 14 precipitation frequency estimates (https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_map_cont.html). These rainfall thresholds for the analyzed counties/cities were extracted using the latitude and longitude of the representative NOAA rainfall station for each city/county. Similarly, to quantify the tide characteristics of flood claim events, the tide metrics calculated for each flood claim event were compared with the minor flood threshold tide level defined by NOAA (Sweet et al., 2018). This threshold is associated with tide levels when minor flooding is likely, and a coastal flood warning of serious risks to life and property is issued by NOAA (Sweet et al., 2018). Generally, this minor flood threshold tide level ranges between a 1-year to 2-year return period event. The rainfall and tide level thresholds (Mean Higher High Water datum) for the analyzed cities/counties are shown in Table A1 in the appendix.

Using the rainfall and tide thresholds from the previous section, each flood claim event was categorized into LRLT (low rainfall low tide), HRLT (heavy rainfall low tide), LRHT (low rainfall high tide), and HRHT (heavy rainfall high tide) groups, based on the observed rainfall and tide metrics. These divisions are shown in Fig. 3a.

2.3.3. Comparison of PWS rainfall with NOAA rainfall

2.3.3.1. Heavy rainfall flood claim events selection. To assess the value of PWS data, we focus on flood claim events that are associated with heavy rainfall. Heavy rainfall events in this study are defined as events falling into the high rainfall low tide (HRLT) or high rainfall high tide (HRHT) quadrants defined in Fig. 3a, based on NOAA observed rainfall. These heavy rainfall events have rainfall exceeding the 1-year return period and resulted in flood losses, which we referred to as heavy rainfall flood claim events. These events were selected for the analysis in the next section.

2.3.3.2. Comparison of NOAA and PWS rainfall in heavy rainfall flood claim events. To assess the ability of PWSs to capture the spatial heterogeneity of observed rainfall for the heavy rainfall flood claim events, the rainfall metrics computed from agency and crowdsourced rainfall stations were compared. For each flood claim in the heavy rainfall claim events, rainfall observations from the neighboring PWSs (defined as PWSs within a specific distance from the centroid of the census tract associated with flood claim) were extracted. The distance is determined based on the availability of PWSs that were actively reporting rainfall observations and the level of consensus between those PWSs. Typically, distances of 4 km or less should be selected (Chen et al., 2021). To ensure trustworthy and quality controlled PWS rainfall data, the Reputation System for Crowdsourced Rainfall Network (RSCRN) (Chen et al., 2021) approach was applied to the PWS data. Using neighboring PWSs as a cluster for running the RSCRN, the trust score of each PWS was assigned based on its agreement or disagreement with the neighboring PWSs in a cluster. The PWSs that received trust scores above 5.0 (out of 10.0) were considered trustworthy and included in the comparison analysis with NOAA gauges. Using the trustworthy PWSs, the median of rainfall metrics (the maximum 1-H, 3-H, 6-H, and 24-H accumulated rainfall) of the neighboring PWSs were computed and compared with the average rainfall metrics computed from NOAA rainfall stations.

3. Results and discussions

3.1. Rainfall and tide characteristics of flood claim events

We present the rainfall and tide characteristics of 58 flood claim events in Norfolk, VA in the Virginia Beach metropolitan area as an example (Fig. 3b). Among the 58 flood claim events (total flood claim events, TE), there were 37 events associated with non-zero rainfall (non-zero rainfall flood claim events, NZRE) and 19 flood claim events associated with zero rainfall. To understand what event characteristics lead to flood claim events on non-zero rainfall days, scatterplots of the rainfall and tide metrics in Norfolk are shown in Fig. 3b for all non-zero rainfall events that occurred during 2016 to 2020. These events include both non-flood-claim events (defined as rainfall events during which no flood claims were reported) and flood claim events (defined as rainfall events during which flood claims were recorded). In this figure, we see that the 37 non-zero flood claim events (black "X"s in Fig. 3b), are generally associated with higher maximum 1-H, 3-H, 6-H, 24-H rainfall and higher maximum tide level than the non-flood claim events (blue dots in Fig. 3b), as shown by the size of the symbols. Fig. 3b also shows the range of rainfall for each flood claim event duration in the boxplots above the scatterplots, and of the tide level in the boxplot on the right. From these we see that for the flood events that occurred in Norfolk, the majority of 1-H maximum rainfall ranged from 0.3 to 1.5 in., with three events (1.6, 2.4, 2.7 in.) above the boxplot's upper whisker. The majority of tide levels were between 0.7 and 2.0 feet, with one event associated with a significantly higher tide (3.4 feet).

The rainfall (1-year rainfall return period) and tide (minor flood risk) threshold values are shown by the vertical and horizontal dotted lines in Fig. 3b, respectively. These lines indicate that most flood claims events

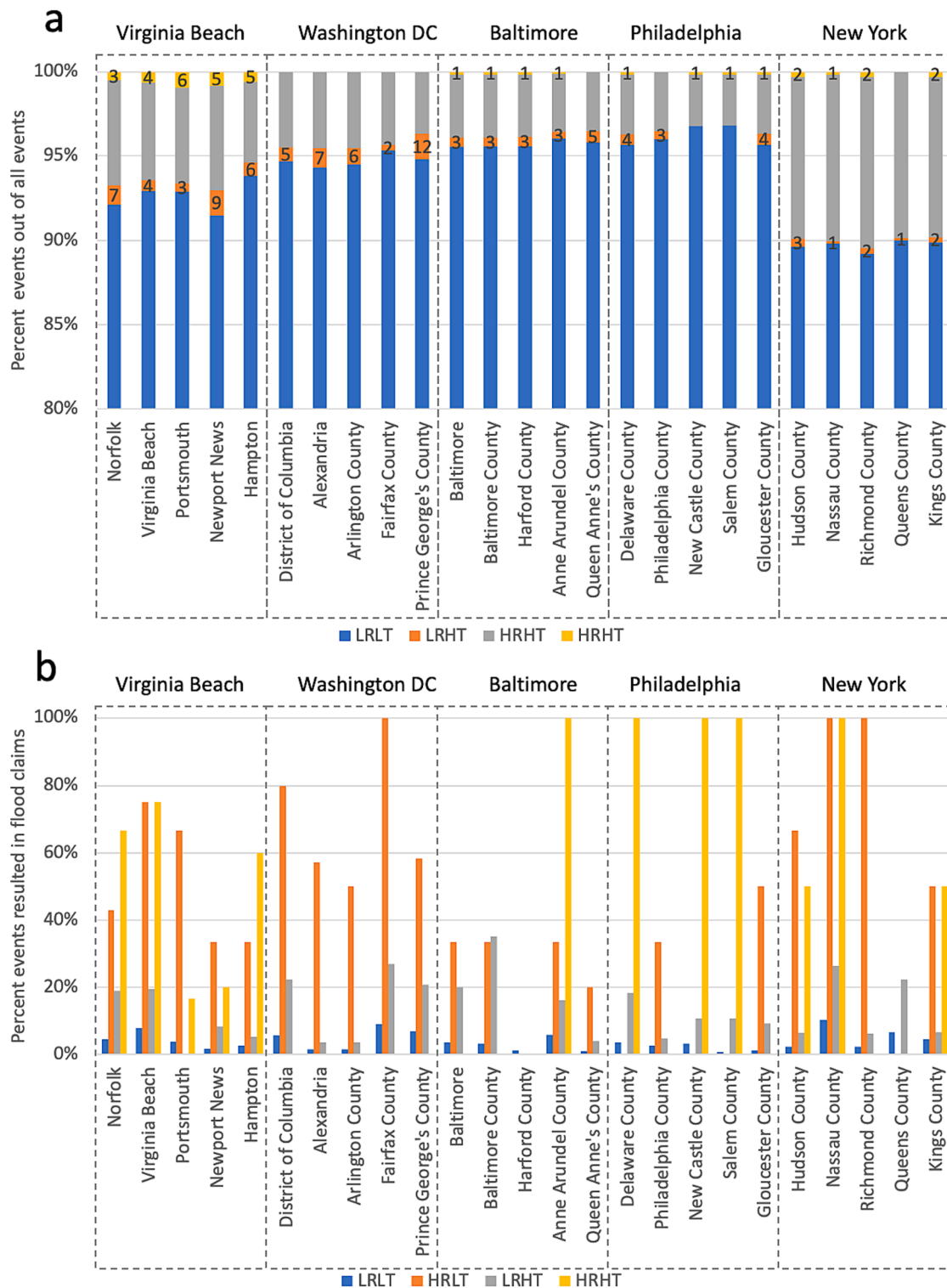


Fig. 4. (a) Percentage of events out of total events in each group. The numbers in the bar represent the number of flood claim events in LRHT and HRHT groups. (b) percentage of flood claim events out of all flood claim events in each group using 24-H rainfall duration threshold.

are in the low rainfall low tide (LRLT) or low rainfall high tide (LRHT) group as defined in Fig. 3a, which is associated with lower than 1-year return period rainfall, regardless of the duration, since these events happen more often than extreme events. The boxplot of the tide level shows that the tide threshold for Norfolk was close to the third quantile tide levels of flood claim events, which indicates that 25% of the flood claim events are in the low rainfall high tide (LRLT) or high rainfall high tide (HRHT) group with tide levels exceeding this threshold. Similar

patterns of flood claim events characteristic were observed in the other four counties in the Virginia Beach metropolitan area (Fig. A1 in the appendix). These results imply that, although most of the flood insurance policies are based on the 100-year flood plain (Kousky and Michel-Kerjan, 2017), flood losses can still occur when there is both low rainfall and a low tide level.

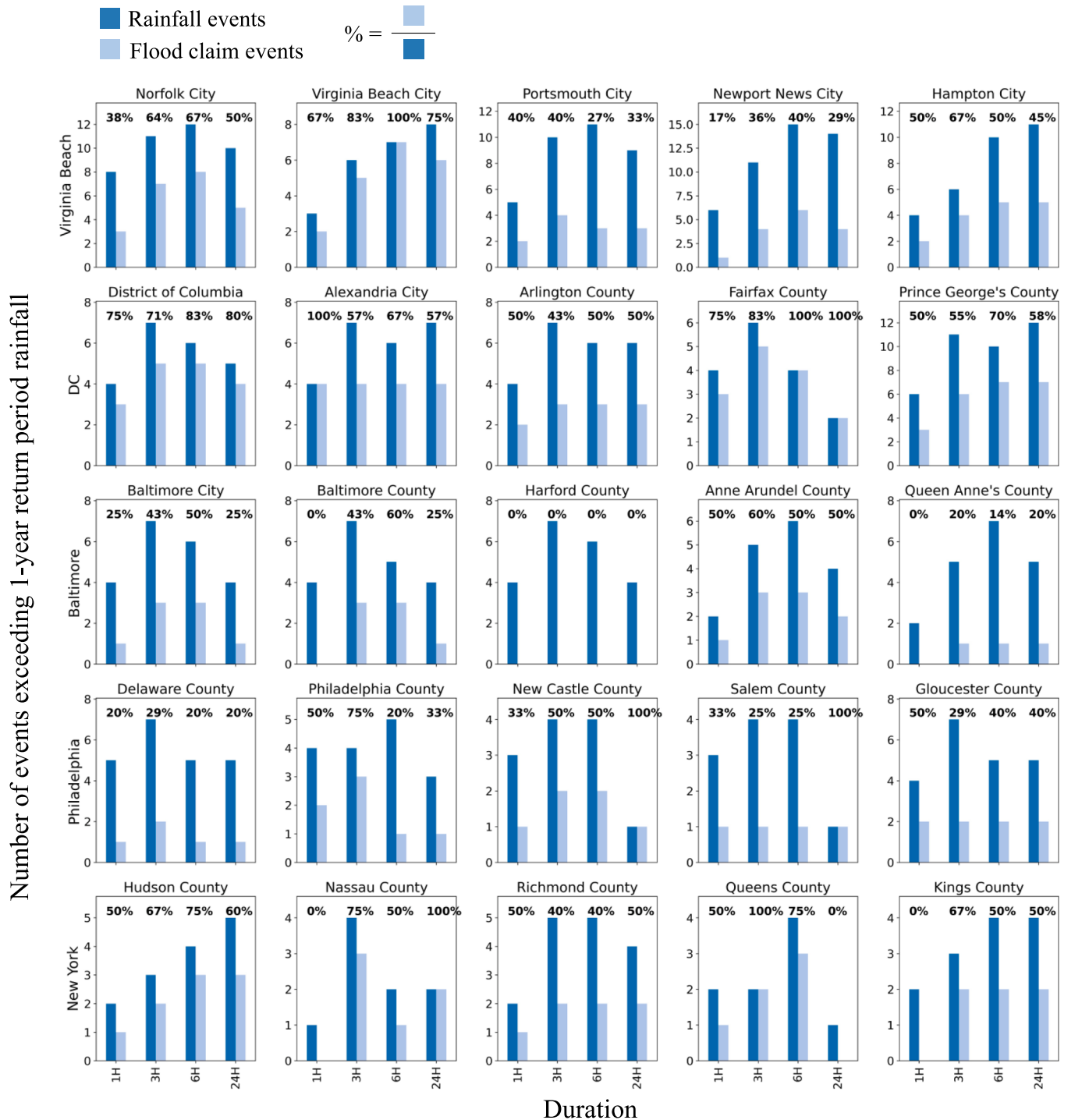


Fig. 5. Number and percentage of events exceeding the 1-year return period rainfall that resulted in flood claims. The darker blue bar represents the number of rainfall events exceeding the 1-year rainfall level. The lighter blue bar represents the number of those events that resulted in flood claim events. The percentage above the bars represents the percentage of those rainfall events that resulted in flood claim events.

3.2. Categorizing flood claim events using rainfall and tide metrics

To obtain the rainfall and tide characteristics of the analyzed metropolitan areas during the study period, we first analyzed the relationship between the rainfall threshold and the observed rainfall, as well as the tide level threshold and the observed tide level in all events (both flood claim and non-flood claim) that occurred during 2016–2020. Fig. 4a and Table A2 show our findings using 24-H rainfall as an example. Here we see that the majority of events were in the low rainfall

low tide (LRLT) (89.2 to 95.3%) and low rainfall high tide (LRHT) groups (3.0 to 10.1%). However, we found that less than 1.5% of the events were in high rainfall low tide (HRLT) group, and less than 0.8% of the events were in the high rainfall high tide (HRHT) group. As shown in Table A3 in the Appendix, across metropolitan areas, the number of events that exceed 1-year rainfall and minor flood risk tide level range from 19 to 63 and 1 to 12, respectively, which indicates minor flood risk tide events occurred 4.8 to 15.8 times and 1-year rainfall events occurred 0.25 to 3 times per year on average over the study period.

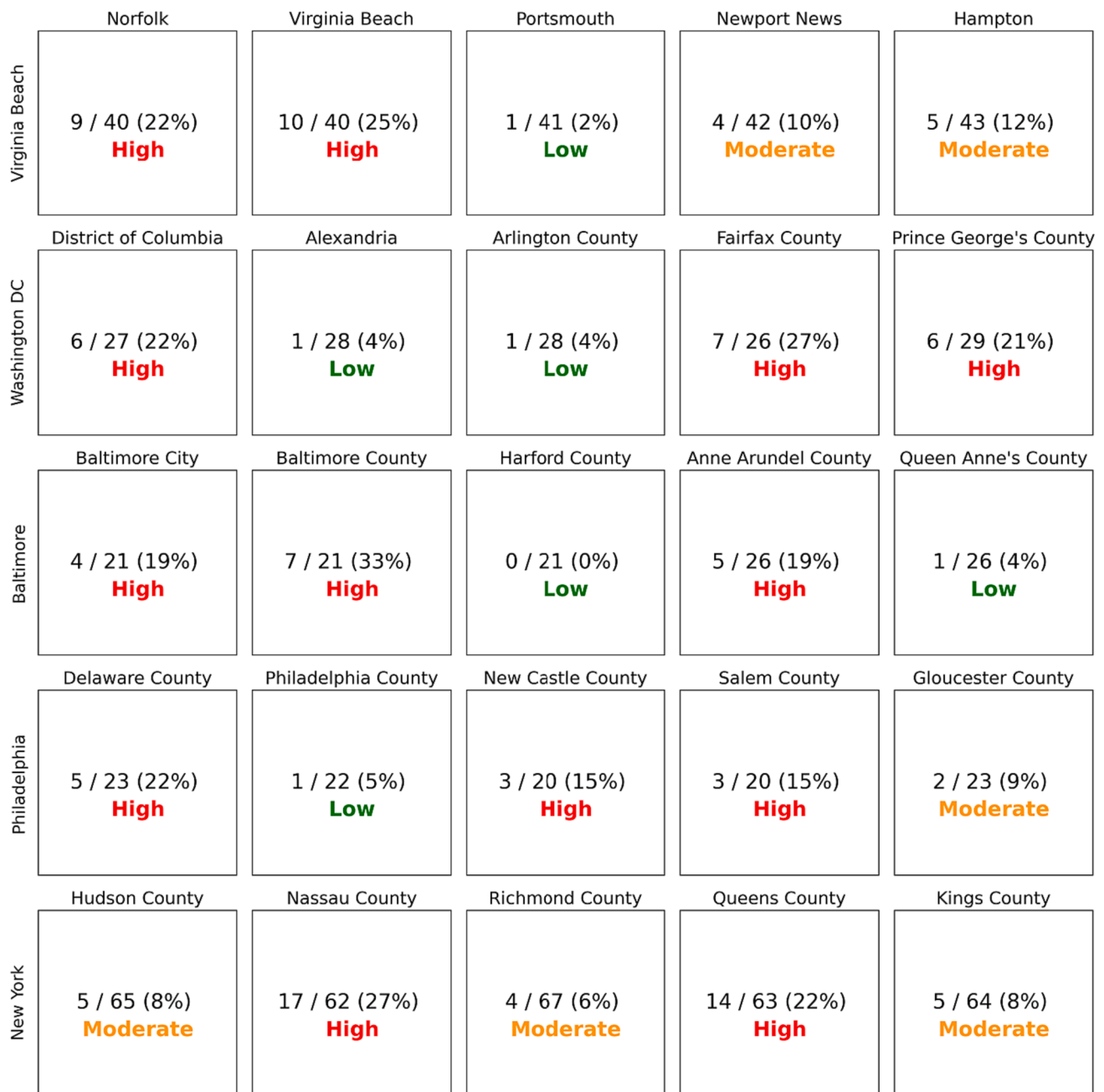


Fig. 6. Number and percentage of events exceeding the minor flood risk tide level resulting in flood claims. Based on the percentage, the analyzed counties/cities were categorized into high (greater than 15%), moderate (5–15%), and low (less than 5%) tidal flood risk groups.

Based on the 24-H rainfall threshold, counties/cities in VB and DC both have a greater number and percentage of high rainfall events compared to BAL, PHL, and NY, implying a higher risk for rainfall-driven floods. Similarly, based on the tide threshold, counties/cities in VB, DC, and NY have a higher percentage of high tide events compared to BAL and PHL, implying a higher risk of tidal floods. These findings were consistent regardless of the durations of rainfall events, as shown in Fig. A2 and A3 in the Appendix.

3.3. Quantifying the characteristics of flood claim events

Following the analysis from the previous section, we calculated the number of flood claim events in each rainfall and tide group, using 24-H rainfall duration as an example (Fig. 4b). Using the Virginia Beach metropolitan area as an example, we found that the percentage of events

resulting in flood claims are much higher in the high rainfall low tide (HRLT) (50%) and high rainfall high tide (HRHT) (48%) groups than in the low rainfall low tide (LRLT) group (4%) and the low rainfall high tide (LRHT) group (10%). This shows the expected results that, although heavy rainfall events are rare, they are much more likely to result in flood insurance claims. Based on the rainfall threshold, VB and DC have the highest number of events in the HRLT and HRHT groups, but the percentages of events in high rainfall groups (HRLT and HRHT) range from 16.7 to 75%. This also implies that not all the events exceeding the 1-year return period maximum 24-H rainfall resulted in flood loss. Interestingly, the combination of high rainfall and tide (HRHT) occurred in multiple events in VB, while no HRHT events occurred in DC, even with a similar number of HRLT events. Based on the tide threshold, the events in the LRHT group that resulted in flood losses varied. For example, in Norfolk and Virginia Beach, the percentages of events in the

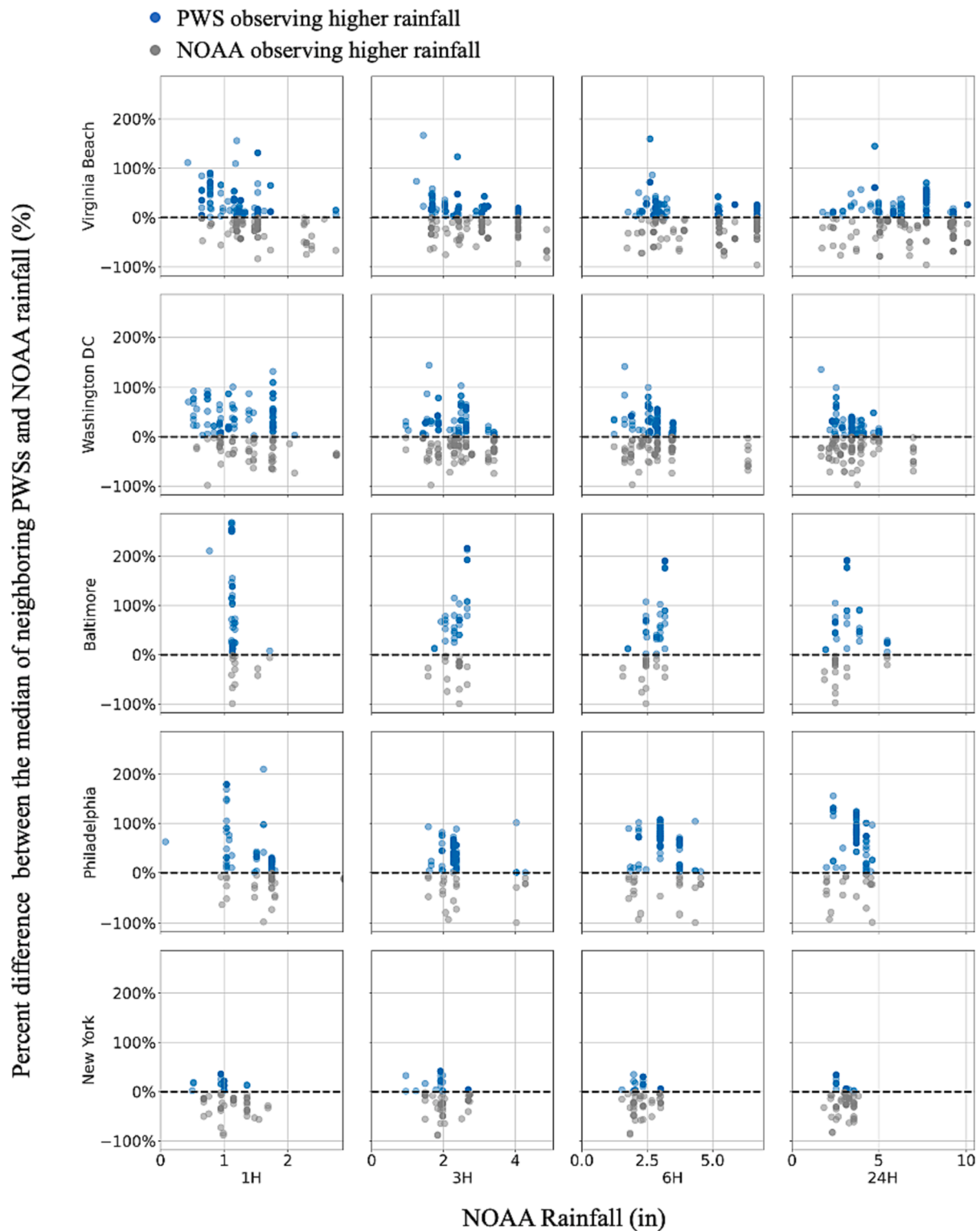


Fig. 7. The comparison of NOAA and PWS observed rainfall metrics for heavy rainfall flood claim events. The x-axis represents the average NOAA observed rainfall near the census tract (y-axis), while the y-axis represents a function of x-axis for the percent difference between (1) the median observed rainfall of neighboring trustworthy PWSs (within 5 km from the census tract the flood claim is associated with) and (2) the average NOAA observed rainfall.

LRHT group that resulted in flood losses were 19%, compared to 8.1, 5.3 and even 0% in Newport News, Hampton, and Portsmouth, respectively. This indicates that most flood claims in the analyzed counties/cities are associated with less than a 1-year return period rainfall.

The data in Fig. 4b also allowed us to investigate how rainfall and tide levels interact. If these variables are independent, the probability

that a flood event leads to flood claims when high rainfall and high tide occur together is equal to (1) the product of the probability that a flood event leads to flood claims when rainfall is high and (2) the probability that a flood event leads to flood claims when the tide is high. If the observed probability that a flood event leads to flood claims is higher than this, the effects of these variables interact to increase the risk of

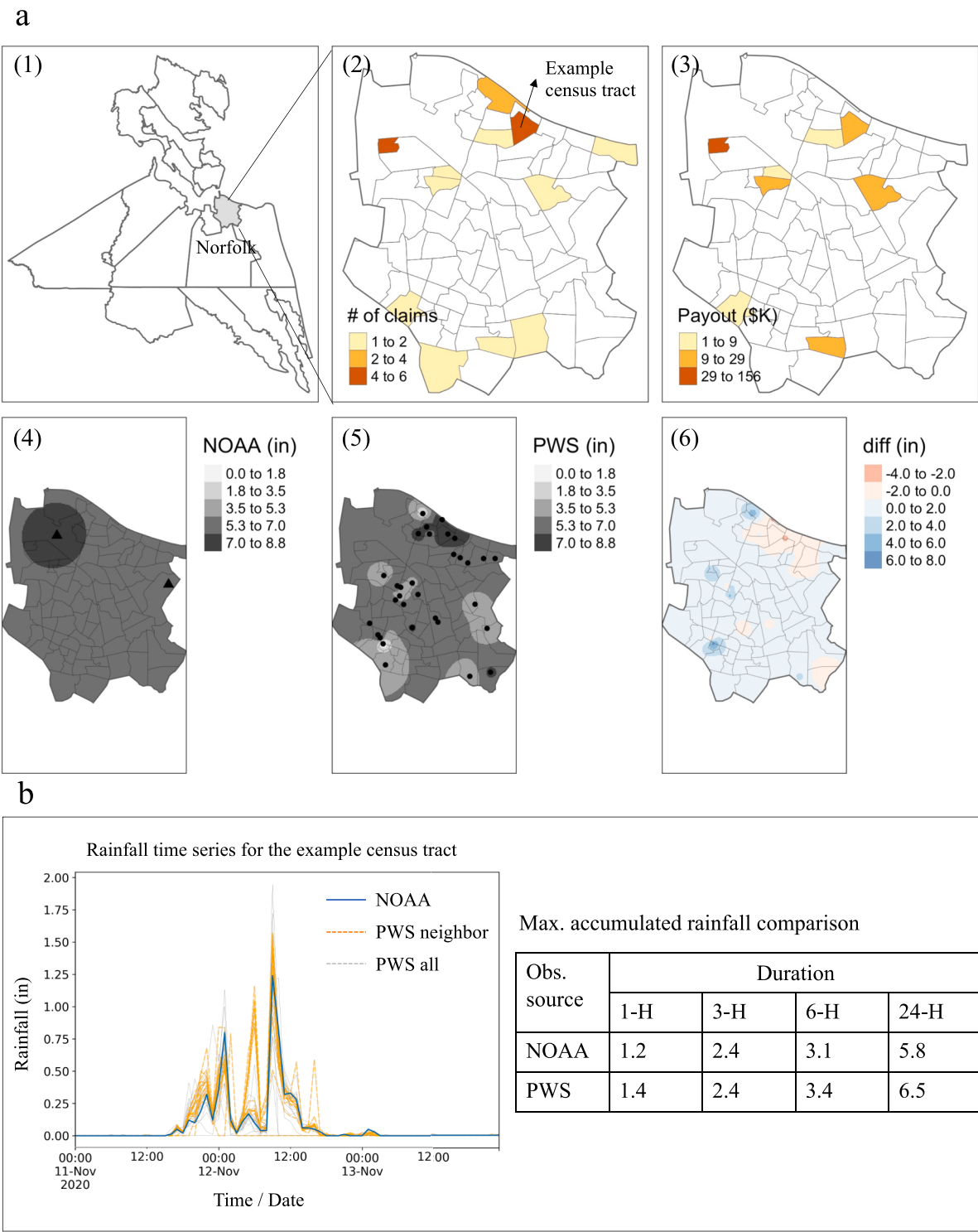


Fig. 8. First example flood claim event of PWS capturing heterogeneous storms coinciding with the location of flood claims.

flooding. As shown in Table A4 in the Appendix, in the Virginia Beach Region, the probability of a flood event leading to a claim when both the rainfall and tide is high far exceeds what would be expected if the two variables were independent, indicating a strong interaction. In the other regions, there were only 1 or 2 events in which both high rainfall and tide levels occurred together, consequently, there is not enough information to determine if these variables interact there as well.

The rainfall threshold analysis was further expanded in Fig. 5. The events that exceed the rainfall threshold (HRLT and HRHT groups) were analyzed for each duration. In VB and DC, there are more high rainfall

events, and the percentage of high rainfall events that led to flood losses varied across different durations of maximum rainfall. For example, in Fairfax County, the percentage of heavy rainfall events exceeding the rainfall threshold of 6-H and 24-H that resulted in flood losses were 100%, while one event in both the groups of exceeding 1-H and 3-H rainfall did not result in flood loss. On the other hand, in Alexandria City, 100% of the events exceeding the 1-H rainfall threshold resulted in flood losses, while only 50–70% of the events above the other duration thresholds resulted in flood losses. In summary, by comparing across metropolitan areas, the analysis shows that VB and DC have relatively

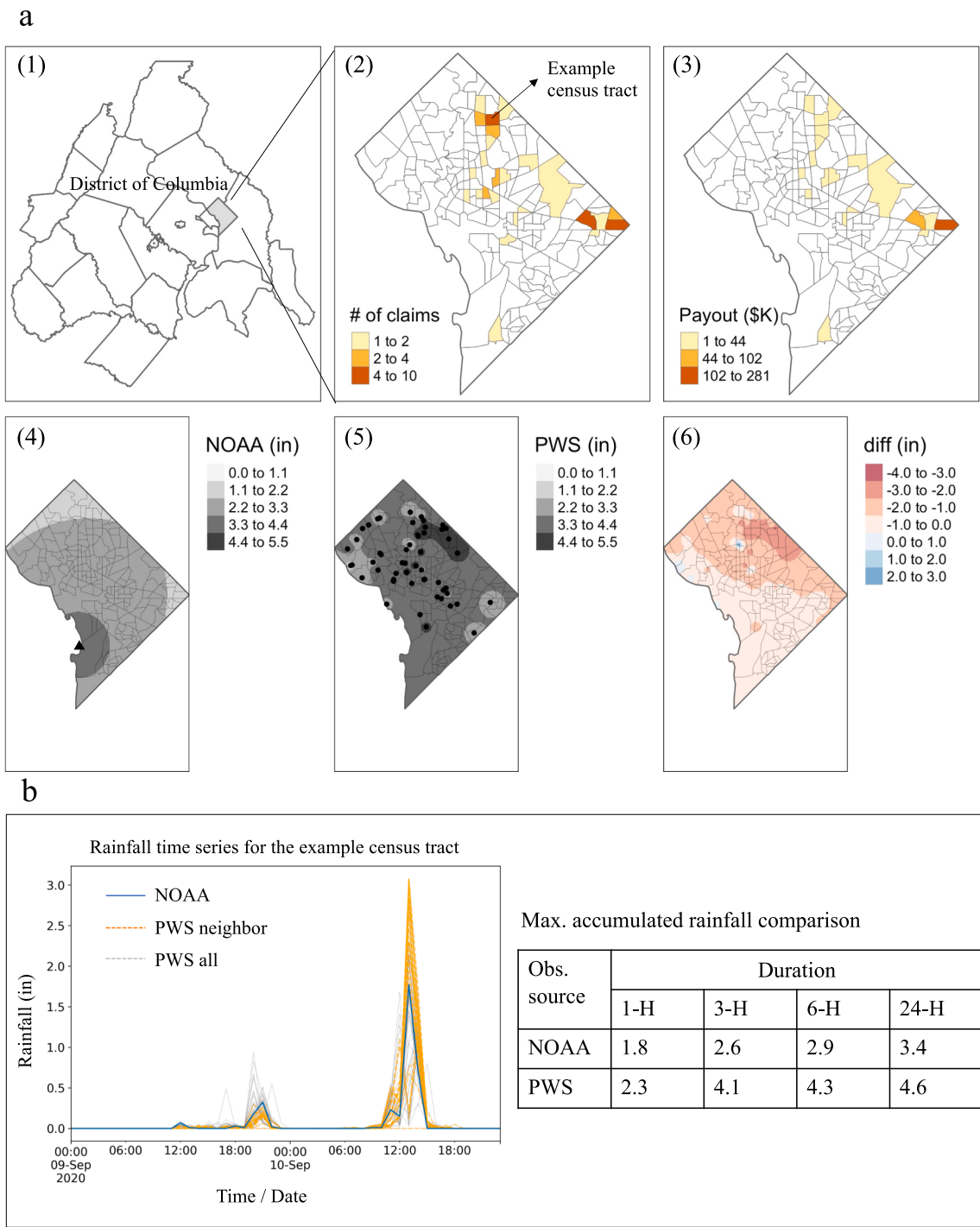


Fig. 9. Second example flood claim event of PWS capturing heterogeneous storms coinciding with the location of flood claims.

more high rainfall events and a higher percentage of high rainfall events leading to flood losses.

In Fig. 6, the number and percentage of high tide events (events in LRHT and HRHT) that resulted in flood losses were calculated. In the twenty-five analyzed counties/cities, the flood loss percentage can be categorized into high, moderate, and low tidal flood risk groups based on the percentages of high tide events resulting in flood losses. Cities/counties in the high tidal flood risk group have percentages greater than 15%, with Baltimore County in BAL (33%), Nassau County in NY (27%), and Fairfax County (27%) having the three highest percentages. The moderate flood risk groups have percentages ranging from 5 to 15%. Example cities/counties include Newport News City, Hampton City in

VB, Gloucester County in PHL, Hudson County and Kings County in NY. In the low flood risk groups, the percentages are lower than 5%. Surprisingly, the results showed that some coastal cities such as Portsmouth City in VA, Alexandria City and Arlington County in DC have very low percentages of high tide events resulting in flood losses, which may imply that flood losses in these areas are more rainfall-driven than tidal-driven.

3.4. Comparison of NOAA and PWS rainfall in heavy rainfall flood claim events

A total of 110 flood claim events associated with high rainfall (HRLT

and HRHT groups) were used to compare the rainfall metrics observed from agency (NOAA) and crowdsourced (PWS) rainfall stations. There were 34, 32, 12, 14 and 18 flood claim events with flood claims that occurred in 299, 228, 61, 118, and 88 census tracts in VB, DC, BAL, PHL, and NY, respectively. In these flood claim events, the maximum accumulated rainfall varied across events. The 1-H, 3-H, 6-H and 24-H ranged from 0.08 to 3.18, 0.95 to 4.88, 1.23 to 6.69 and 1.68 to 10.07 in. respectively, based on the NOAA observed rainfall.

The comparison of NOAA and PWS observed rainfall for each heavy rainfall flood claim event is shown in Fig. 7. The x-axis represents the average observed rainfall from NOAA rainfall gauges within 10 km of the census tract (y-axis), while the y-axis represents percent difference between (1) the median observed rainfall of neighboring trustworthy PWSs (within 5 km from the census tract the flood claim is associated with) and (2) the average NOAA observed rainfall. Note that we used median for PWSs for excluding the effect of untrustworthy PWSs and average for NOAA rainfall stations because there are usually only 2–3 NOAA rainfall stations nearby. In Fig. 7, each dot represents the rainfall metrics of census tracts associated with claims within a flood event, with the blue dots representing events in which the PWSs observed greater median rainfall than the average NOAA gauge, while the gray dots represent the reverse. The results show that there can be significant differences in rainfall measured by the PWS gauges and the NOAA gauges, suggesting there is high spatial variability in rainfall during the flood claim events. In BAL and PHL, the majority of flood claim events have PWSs observing greater rainfall (91, 67, 57 and 57% for 1-H, 3-H, 6-H, and 24-H rainfall for BAL). In VB and DC, the number of PWSs observing greater or lesser rainfall than NOAA is similar. However, the percent differences in events in which the PWSs observe greater rainfall are higher than in the events in which they observe less rainfall (e.g., 31% compared to –22% for 1-H rainfall in VB). Overall, we see that PWSs tend to observe higher rainfall than the NOAA gauges for less extreme and shorter duration events, suggesting they may be capturing more localized flooding, whereas the gauges are in closer agreement for more extreme events that generally span a larger area.

Figs. 8 and 9 compare Inverse Distance Weighted (IDW) interpolated maps of total accumulated rainfall observed from NOAA rainfall stations and PWSs, as well as the rainfall hourly time series for an example census tract and the derived rainfall metrics for two example flood claim events (one on 11/11/2020 in Norfolk (Fig. 8) and the other on 9/9/2020 in the District of Columbia (Fig. 9)). As shown in Fig. 8a (1) – (3), in the 11/11/2020 event, there was a total of 25 claims distributed across the city with \$254,364 in insurance payout. The IDW interpolated maps for NOAA and PWS in Fig. 8a (4) – (6) show that a significant difference

(~2 in) in total accumulated rainfall occurred around the example census tract area, with the PWS maps showing greater rainfall. Comparing the hourly time series of the NOAA rainfall stations with the PWS neighboring stations for the example census tract in Fig. 8b, we see that PWSs captured a significantly larger rainfall peak (~1 in) on 11/12 01:00.

In the second example (9/9/2020 event), shown in Fig. 9a (1) – (3), there were a total of 60 claims with \$516,178 in insurance payout. Similarly, the census tracts with the greatest number of flood claims coincide with areas showing significant differences in interpolated rainfall from NOAA gauges vs. PWS gauges, with the PWS gauges showing greater rainfall (Fig. 9a (4) – (6)). This suggests the coverage from the NOAA rainfall stations was insufficient to capture this area of heavy rainfall that resulted in claims. As shown in Fig. 9b, the rainfall peak on 9/10 13:00 observed from PWSs (~3 in) was more than 1 in. greater than NOAA (~1.7 in). Both flood claim events suggest that the spatial density of PWSs provides the ability to capture localized storms that resulted in flood claims, which greatly assists understanding the cause of the flood losses.

3.5. Limitations and future research

This study utilizes the recently released and publicly available NFIP flood claim data to analyze the rainfall and tide characteristics during storm events that lead to flood losses. Given the novelty of the dataset, an exploratory analysis was presented that resulted in a straightforward method for understanding the rainfall and tide characteristics of flood claim events captured in the NFIP flood insurance claim data. The results from this study provide a first step toward understanding the information and potential insights from this newly released dataset, and our hope is that it paves the road for more sophisticated statistical analysis in future studies. The method used in this study can be easily generalized and applied in other regions or time periods. However, there are limitations within the dataset used and method presented in this study that are worth pointing out for potential future studies.

First, in this dataset, there are more than 40 fields providing key information such as locational, structural, occupancy and claim variables (Dombrowski et al., 2021). However, some important features of the insured property (e.g., property value and financial damages) and, the flood hazard characteristics (e.g., flood depth and flood category) are not included or have been heavily redacted (Vishnu et al., 2021). Most importantly, the finest geographical level of where the flood claims occurred was the census tract level, as the coordinates of the insured properties were redacted. Additionally, since the private flood insurance

Table A1
Flood claim, rainfall, tide, and threshold values used in this study.

MA	County/City	NFIP Flood Claim(2016 – 2020)			NOAA Rainfall and Tide			Threshold Values for Analysis				
		# of claims	# of claim events	Payout (\$ M)	Rainfall StationID / Name	Tide Station ID / Name		Rainfall (in.)				Tide (ft.)
VB	Norfolk City	472	58	6.1	72,308,013,737	Norfolk International Airport	8,638,610Sewell’s Point, VA	1.40	1.79	2.17	2.93	1.74
	Virginia Beach City	1695	86	55.8	72,307,513,769	Oceana NAS		1.42	1.82	2.23	3.00	
	Portsmouth City	269	37	5.7	72,308,513,750	Norfolk NAS		1.41	1.80	2.17	2.93	
	Newport News City	94	21	3.9	72,308,693,741	Newport News International Airport		1.40	1.80	2.16	2.93	
	Hampton City	206	41	2.7	74,598,013,702	Langley Air Force Base		1.41	1.81	2.18	2.93	
DC	District of Columbia	212	69	1.4	72,405,013,743	Washington Reagan National Airport	8,594,900Washington, DC	1.22	1.50	1.84	2.56	1.77
	Alexandria City	214	26	5.1	72,405,013,743							
	Arlington County	80	24	1.4	72,405,013,743							

(continued on next page)

Table A1 (continued)

	Fairfax County	289	89	4.7	72,403,093,738	Washington Dulles International Airport		1.16	1.47	1.81	2.50
	Prince George's County	399	99	3.8	74,594,013,705	Camp Springs Andrews AFB		1.20	1.52	1.86	2.62
BAL	Baltimore City	150	41	4.0	74,594,493,784	Maryland Science Center	8,574,680	1.18	1.52	1.89	2.67
	Baltimore County	117	45	1.9	74,594,493,784		Baltimore, MD				1.71
	Harford County	28	15	0.4	74,594,493,784						
	Anne Arundel County	118	65	0.7	72,215,899,999	Annapolis US Naval Academy		1.19	1.53	1.89	2.66
	Queen Anne's County	48	15	0.5	72,215,899,999						
PHL	Delaware County	568	40	18.8	72,408,013,739	Philadelphia International Airport	8,551,910	1.19	1.56	1.93	2.68
	Philadelphia County	498	26	13.9	72,408,594,732	Northeast Philadelphia Airport	Reedy Point, DE	1.18	1.56	1.96	2.75
	New Castle County	106	35	2.0	72,418,013,781	Wilmington New Castle CO Airport		1.19	1.55	1.91	2.69
	Salem County	37	12	0.8	72,418,013,781						
	Gloucester County	59	13	0.6	72,408,013,739	Philadelphia International Airport		1.19	1.56	1.93	2.68
NY	Hudson County	166	37	1.8	72,505,394,728	NY City Central Park,	8,518,750	1.07	1.63	2.06	2.93
	Nassau County	375	129	4.7	74,486,094,789	JFK International Airport	The Battery, NY	1.11	1.68	2.03	2.67
	Richmond County	70	31	0.3	72,502,014,734	Newark Liberty International Airport		1.13	1.54	1.98	2.73
	Queens County	161	75	1.6	72,503,014,732	LaGuardia Airport		1.07	1.65	2.02	2.69
	Kings County	109	55	0.3	72,505,394,728	NY City Central Park,		1.07	1.63	2.06	2.93

markets have been growing in recent years, the NFIP flood claims may only account for a portion of the flood losses that actually occurred during the storm events (Kousky and Michel-Kerjan, 2017). These factors are important for understanding the effect of social inequality on flood claims, as flood losses highly depend on the vulnerability of the infrastructure and social-economic status of the populations affected (Tate et al., 2021). Therefore, research into the characteristics of rainfall and tide for flood claim events could be expanded in future studies if more information about flood claims become available.

Second, we found that a relatively high percentage (~90%) of claims

occurred with observed rainfall lower than a 1-year return period. Moreover, about 30% of the claims occurred on zero-rainfall days (days on which the total accumulated precipitation is zero). About 10–20% of these non-zero rainfall flood claims were associated with high tide levels. Another potential explanation for these claims is water-related damage such as roof, wall, or pipe leakage (Spekkers et al., 2015). It may also be possible that the date of loss documented in the flood claims was incorrect. In these cases, the rainfall and tide observations on days surrounding the occurrence of flood losses could be examined to verify the cause of the flood losses.

Table A2

Number (percentage) of events during 2016–2020 in each group among non-zero rainfall days using 24-H rainfall duration threshold.

MA	County/City	Number of non-zero rainfall days	Number (percentage) in each group out of all days/events			
			LRLT	LRHT	HRLT	HRHT
VB	Norfolk	598	551 (92.1%)	37 (6.2%)	7 (1.2%)	3 (0.5%)
	Virginia Beach	624	580 (92.9%)	36 (5.8%)	4 (0.6%)	4 (0.6%)
	Portsmouth	622	578 (92.9%)	35 (5.6%)	3 (0.5%)	6 (1.0%)
	Newport News	600	549 (91.5%)	37 (6.2%)	9 (1.5%)	5 (0.8%)
	Hampton	795	746 (93.8%)	38 (4.8%)	6 (0.8%)	5 (0.6%)
DC	District of Columbia	602	570 (94.7%)	27 (4.5%)	5 (0.8%)	0 (0.0%)
	Alexandria	615	580 (94.3%)	28 (4.6%)	7 (1.1%)	0 (0.0%)
	Arlington County	617	583 (94.5%)	28 (4.5%)	6 (1.0%)	0 (0.0%)
	Fairfax County	598	570 (95.3%)	26 (4.3%)	2 (0.3%)	0 (0.0%)
	Prince George's County	787	746 (94.8%)	29 (3.7%)	12 (1.5%)	0 (0.0%)
BAL	Baltimore	538	514 (95.5%)	20 (3.7%)	3 (0.6%)	1 (0.2%)
	Baltimore County	540	516 (95.6%)	20 (3.7%)	3 (0.6%)	1 (0.2%)
	Harford County	542	518 (95.6%)	20 (3.7%)	3 (0.6%)	1 (0.2%)
	Anne Arundel County	732	703 (96.0%)	25 (3.4%)	3 (0.4%)	1 (0.1%)
	Queen Anne's County	740	709 (95.8%)	26 (3.5%)	5 (0.7%)	0 (0.0%)
PHL	Delaware County	620	593 (95.6%)	22 (3.5%)	4 (0.6%)	1 (0.2%)
	Philadelphia County	624	599 (96.0%)	22 (3.5%)	3 (0.5%)	0 (0.0%)
	New Castle County	622	602 (96.8%)	19 (3.1%)	0 (0.0%)	1 (0.2%)
	Salem County	626	606 (96.8%)	19 (3.0%)	0 (0.0%)	1 (0.2%)
	Gloucester County	623	596 (95.7%)	22 (3.5%)	4 (0.6%)	1 (0.2%)
NY	Hudson County	657	589 (89.6%)	63 (9.6%)	3 (0.5%)	2 (0.3%)
	Nassau County	618	555 (89.8%)	61 (9.9%)	1 (0.2%)	1 (0.2%)
	Richmond County	641	572 (89.2%)	65 (10.1%)	2 (0.3%)	2 (0.3%)
	Queens County	641	577 (90.0%)	63 (9.8%)	1 (0.2%)	0 (0.0%)
	Kings County	653	587 (89.9%)	62 (9.5%)	2 (0.3%)	2 (0.3%)

LRLT: low rainfall low tide; LRHT: low rainfall high tide; HRLT: high rainfall low tide; HRHT: high rainfall high tide.

Another limitation of our analysis is that we only utilized the rainfall data observed from NOAA rainfall stations to categorize the flood claim events. We did this because it is a trustworthy and national consistent dataset. However, other sources of rainfall data could be considered in future studies. For example, regional rainfall networks could also be available through local municipality agencies such as the Hampton

Roads Sanitary District in the Hampton Roads, VA region, who maintain their own rainfall monitoring network (Sadler et al., 2018). This study also investigated the potential for crowdsourced rainfall data collected through PWSs to supplement the rainfall observations from NOAA rainfall stations. Future studies could also incorporate remote sensing rainfall (e.g., radar-derived rainfall estimates from Next Generation

Table A3

Number (percentage) of flood claim events out of all events for each group using 24-H rainfall duration threshold.

MA	County/City	Number (percentage) of flood claim days/events out of the number of days/events in each group			
		LRLT	LRHT	HRLT	HRHT
VB	Norfolk	25/551 (4.5%)	7/37 (18.9%)	3/7 (42.9%)	2/3 (66.7%)
	Virginia Beach	45/580 (7.8%)	7/36 (19.4%)	3/4 (75.0%)	3/4 (75.0%)
	Portsmouth	21/578 (3.6%)	0/35 (0.0%)	2/3 (66.7%)	1/6 (16.7%)
	Newport News	9/549 (1.6%)	3/37 (8.1%)	3/9 (33.3%)	1/5 (20.0%)
	Hampton	19/746 (2.5%)	2/38 (5.3%)	2/6 (33.3%)	3/5 (60.0%)
DC	District of Columbia	32/570 (5.6%)	6/27 (22.2%)	4/5 (80.0%)	0/0 (N/A)
	Alexandria	9/580 (1.6%)	1/28 (3.6%)	4/7 (57.1%)	0/0 (N/A)
	Arlington County	9/583 (1.5%)	1/28 (3.6%)	3/6 (50.0%)	0/0 (N/A)
	Fairfax County	51/570 (8.9%)	7/26 (26.9%)	2/2 (100.0%)	0/0 (N/A)
	Prince George's County	52/746 (7.0%)	6/29 (20.7%)	7/12 (58.3%)	0/0 (N/A)
BAL	Baltimore	18/514 (3.5%)	4/20 (20.0%)	1/3 (33.3%)	0/1 (0.0%)
	Baltimore County	16/516 (3.1%)	7/20 (35.0%)	1/3 (33.3%)	0/1 (0.0%)
	Harford County	6/518 (1.2%)	0/20 (0.0%)	0/3 (0.0%)	0/1 (0.0%)
	Anne Arundel County	41/703 (5.8%)	4/25 (16.0%)	1/3 (33.3%)	1/1 (100.0%)
	Queen Anne's County	7/709 (1.0%)	1/26 (3.8%)	1/5 (20.0%)	0/0 (N/A)
PHL	Delaware County	21/593 (3.5%)	4/22 (18.2%)	0/4 (0.0%)	1/1 (100.0%)
	Philadelphia County	15/599 (2.5%)	1/22 (4.5%)	1/3 (33.3%)	0/0 (N/A)
	New Castle County	19/602 (3.2%)	2/19 (10.5%)	0/0 (0.0%)	1/1 (100.0%)
	Salem County	4/606 (0.7%)	2/19 (10.5%)	0/0 (0.0%)	1/1 (100.0%)
	Gloucester County	7/596 (1.2%)	2/22 (9.1%)	2/4 (50.0%)	0/1 (0.0%)
NY	Hudson County	13/589 (2.2%)	4/63 (6.3%)	2/3 (66.7%)	1/2 (50.0%)
	Nassau County	57/555 (10.3%)	16/61 (26.2%)	1/1 (100.0%)	1/1 (100.0%)
	Richmond County	13/572 (2.3%)	4/65 (6.2%)	2/2 (100.0%)	0/2 (0.0%)
	Queens County	38/577 (6.6%)	14/63 (22.2%)	0/1 (0.0%)	0/0 (N/A)
	Kings County	26/587 (4.4%)	4/62 (6.5%)	1/2 (50.0%)	1/2 (50.0%)

LRLT: low rainfall low tide; LRHT: low rainfall high tide; HRLT: high rainfall low tide; HRHT: high rainfall high tide.

Table A4

The interaction probability between tide and rainfall. Column 1–4 represent the probability a rainfall event results in a claim given (1) the rainfall event exceeds the 1-year level, (2) the tide level is above NOAA's minor tide level, or (3) both and compares to the probability in case 3 to (4) for what the probability would be if the variables' effects were independent.

MA	County/City	(1)P(claim HR)	(2)P(claim HT)	(3)P(claim HR and HT) if independent	(4) ObservedP(claim HR and HT) [# HR and HT events]
	Norfolk	0.50	0.23	0.11	0.67 [3]
	Virginia Beach	0.75	0.25	0.19	0.75 [4]
	Portsmouth	0.33	0.02	0.01	0.17 [6]
	Newport News	0.29	0.10	0.03	0.2 [5]
	Hampton	0.45	0.12	0.05	0.6 [5]
	District of Columbia	0.80	0.22	0.18	NA [0]
	Alexandria	0.57	0.04	0.02	NA [0]
	Arlington County	0.50	0.04	0.02	NA [0]
	Fairfax County	1.00	0.27	0.27	NA [0]
	Prince George's County	0.58	0.21	0.12	NA [0]
	Baltimore	0.25	0.19	0.05	0 [1]
	Baltimore County	0.25	0.33	0.08	0 [1]
	Harford County	0.00	0.00	0.00	0 [1]
	Anne Arundel County	0.50	0.19	0.10	1 [1]
	Queen Anne's County	0.20	0.04	0.01	NA [0]
	Delaware County	0.20	0.22	0.04	1 [1]
	Philadelphia County	0.33	0.05	0.02	NA [0]
	New Castle County	1.00	0.15	0.15	1 [1]
	Salem County	1.00	0.15	0.15	1 [1]
	Gloucester County	0.40	0.09	0.03	0 [1]
NY	Hudson County	0.60	0.08	0.05	0.5 [2]
	Nassau County	1.00	0.27	0.27	1 [2]
	Richmond County	0.50	0.06	0.03	0 [2]
	Queens County	0.00	0.22	0.00	NA [0]
	Kings County	0.50	0.08	0.04	0.5 [2]

MA: Metropolitan area.

Red indicates $P > \text{if independent}$.

Blue indicates $P < \text{if independent}$.

Black indicates no observed events.

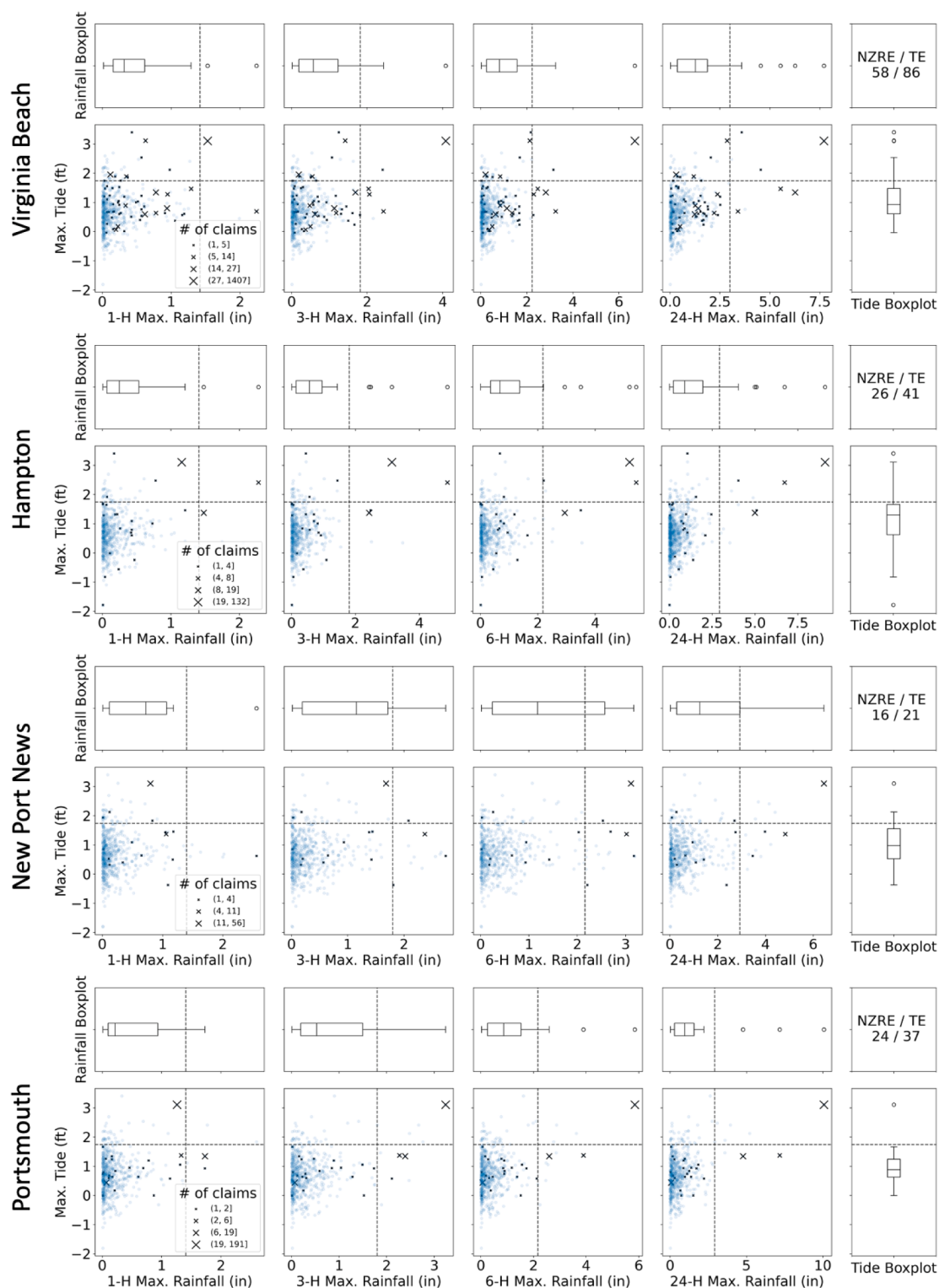


Fig. A1. Analysis of rainfall and tide characteristics for flood claim events for other cities/counties in the Virginia Beach metropolitan area.

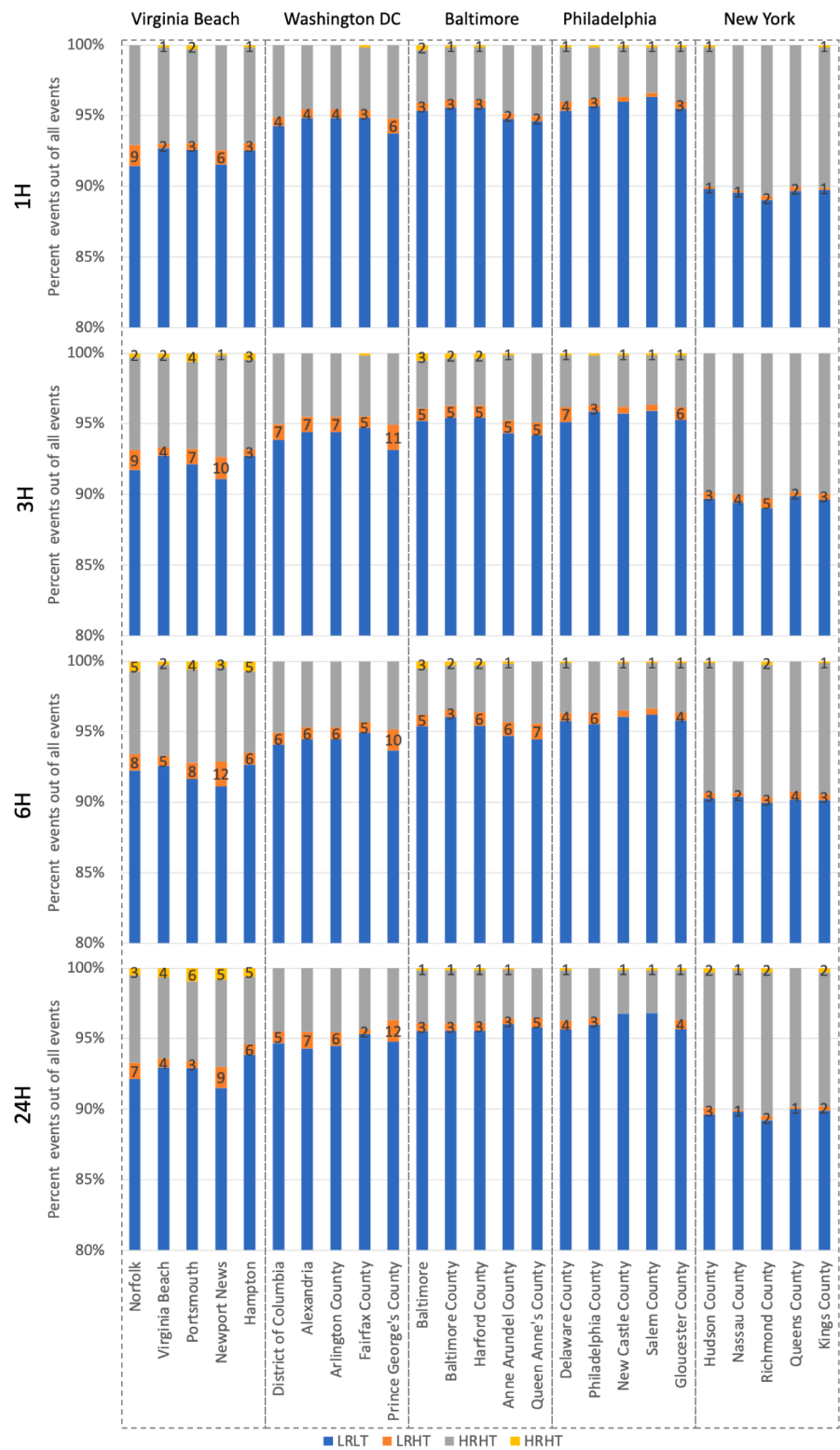


Fig. A2. Percentage of events out of total events in each group for all cities/counties in this study.

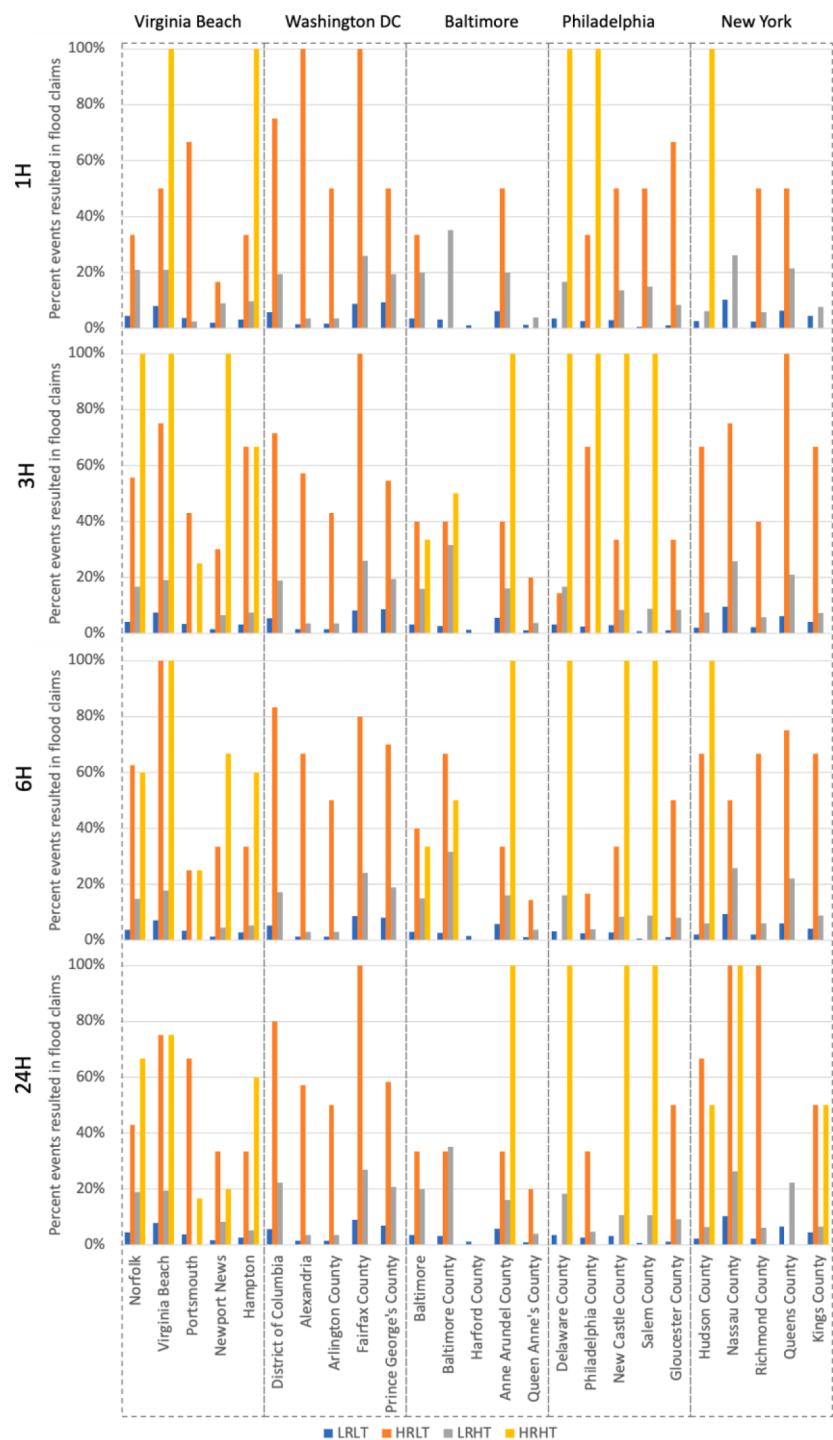


Fig. A3. Percentage of flood claim events out of all flood claim events in each group using 24-h rainfall duration thresholds for all cities/counties in this study.

Weather Radar (NEXRAD) (Cunha et al., 2015)) into the analysis to better compare and contrast the rainfall spatial heterogeneity and its relationship to flood claims. Extending the rainfall analysis geographical extent and time period could also be completed for a more comprehensive understanding of the trends and patterns in flood claims.

Lastly, as mentioned earlier, this study utilized only exploratory analysis for the comparison of NOAA and PWS rainfall during heavy rainfall flood claim events. Future studies could focus on building a model of the number (Poisson regression), presence (logistic regression), or amount (linear regression) of claims as a function of rainfall in the census tract to determine if using PWSs for these models yields more accurate predictions than NOAA gauges. This would provide additional confirmation of the value of PWSs in contributing to our understanding of localized flooding resulting in flood losses.

4. Conclusion

This study presents a methodology for using the NFIP flood insurance claim data to investigate rainfall and tide characteristics of storm events that result in flood losses. The methodology was applied to twenty-five counties/cities in five metropolitan areas (Virginia Beach, Washington DC, Baltimore, Philadelphia, and New York) in the mid-Atlantic region in the U.S. The results showed that the majority of flood claim events in this region are associated with rainfall lower than the 1-year return period and tide level lower than a minor flood risk tide level. However, in storm events exceeding the rainfall and tide threshold, the probability of the occurrence of flood claims could average more than 50% and 20%, respectively. In Virginia Beach, this probability far exceeds what would be expected if these two variables were independent, suggesting they interact to exacerbate flood impacts, while there were not enough compound events to see if this was the case in the other metropolitan areas.

In the analyzed metropolitan areas, storm events associated with longer duration (6H and 24H) rainfall exceeding 1-year return period were found to generally have higher probability of resulting in flood losses. Storm events associated with shorter duration (1H and 3H) exceeding the 1-year return period were also found to increase the chance of flood losses in cities/counties in Virginia Beach and Washington DC metropolitan areas. With respect to tide levels, their effect on flood losses was found to vary across the analyzed cities/counties, with 53% of the analyzed cities/counties having higher than 20%, and 24% having lower than 5% of the high tide events resulting in flood losses, respectively.

Comparison of observed rainfall from NOAA stations and crowdsourced personal weather stations (PWSs) showed that PWSs observe an average of 30% more rainfall than agency rainfall stations at the location of the flood claims across analyzed heavy rainfall events. This suggests the high spatial density of PWSs may be capturing localized rainfall missed by the lower spatial density NOAA gauges, better representing the rainfall spatial heterogeneity of storms that contribute to localized flood losses.

The methodology presented in this paper can be used in future work in other regions or time periods to further advance understanding of the rainfall and tide characteristics of flood claim events captured in the NFIP flood insurance claim data. The results for the Mid-Atlantic region provide a start toward better understanding the role of rainfall-driven and high-tide flood losses based on the flood insurance claims. These findings could benefit multiple stakeholders in the region, including researchers, risk management officials, insurers, and the general public in assessing and building resilience against increased flood risks in coastal cities.

CRedit authorship contribution statement

Alexander B. Chen: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Jonathan L. Goodall:**

Supervision, Conceptualization, Methodology, Writing – review & editing. **Julianne D. Quinn:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – review & editing.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgement

This research is supported by the United States National Science Foundation under Grant No. CBET-1735587. We gratefully acknowledge Federal Emergency Management Agency, National Oceanic and Atmospheric Administration and Weather Underground for access to their data.

Appendix

Table A1.
Table A2.
Table A3.
Table A4.
Fig. A1.
Fig. A2.
Fig. A3.

References

- Federal Emergency Management Agency (FEMA), 2022. OpenFEMA [WWW Document]. URL <https://www.fema.gov/about/reports-and-data/openfema>.
- 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>.
- Blessing, R., Sebastian, A., Asce, S.M., Brody, S.D., 2017. Flood Risk Delineation in the United States: How Much Loss Are We Capturing? *Nat. Hazards Rev.* 18, 04017002. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000242](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000242).
- Chamberlain, S., 2021. rnoaa: “NOAA” Weather Data from R [WWW Document]. URL <https://cran.r-project.org/package=rnoaa>.
- Chang, H., Pallathadka, A., Sauer, J., Grimm, N.B., Zimmerman, R., Cheng, C., Iwaniec, D.M., Kim, Y., Lloyd, R., McPhearson, T., Rosenzweig, B., Troxler, T., Welty, C., Brenner, R., Herreros-Cantis, P., 2021. Assessment of urban flood vulnerability using the social-ecological-technological systems framework in six US cities. *Sustain. Cities Soc.* 68, 102786 <https://doi.org/10.1016/J.SCS.2021.102786>.
- 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, A.B., Goodall, J.L., Chen, T.D., Zhang, Z., 2022. Flood resilience through crowdsourced rainfall data collection: Growing engagement faces non-uniform spatial adoption. *J. Hydrol.* 609, 127724 <https://doi.org/10.1016/J.JHYDROL.2022.127724>.
- Collins, E.L., Sanchez, G.M., Terando, A., Stillwell, C.C., Mitasova, H., Sebastian, A., Meentemeyer, R.K., 2022. Predicting flood damage probability across the conterminous United States. *Environ. Res. Lett.* 17, 034006 <https://doi.org/10.1088/1748-9326/AC4F0F>.
- Cunha, L.K., Smith, J.A., Krajewski, W.F., Baeck, M.L., Seo, B.-C., 2015. NEXRAD NWS polarimetric precipitation product evaluation for IFloods. *J. Hydrometeorol.* 16, 1676–1699.
- Czajkowski, J., Villarini, G., Montgomery, M., Michel-Kerjan, E., Goska, R., 2017. Assessing Current and Future Freshwater Flood Risk from North Atlantic Tropical Cyclones via Insurance Claims. *Sci. Reports* 2017 71 7, 1–10. [10.1038/srep41609](https://doi.org/10.1038/srep41609).
- Davenport, F.V., Burke, M., Diffenbaugh, N.S., 2021. Contribution of historical precipitation change to US flood damages. *Proc. Natl. Acad. Sci. U. S. A.* 118, 2017524118. https://doi.org/10.1073/PNAS.2017524118/SUPPL_FILE/PNAS.2017524118.SAPP.PDF.
- de Vos, L.W., 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.
- 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>.

- Dombrowski, T., Ratnadiwakara, D., Slawson Jr, V.C., 2021. The FIMA NFIP's Redacted Policies and Redacted Claims Datasets. *J. Real Estate Lit.* 28, 190–212.
- Emmanuel, I., Andrieu, H., Lebouis, E., Flahaut, B., 2012. Temporal and spatial variability of rainfall at the urban hydrological scale. *J. Hydrol.* 430–431, 162–172. <https://doi.org/10.1016/j.jhydrol.2012.02.013>.
- Ezer, T., Atkinson, L.P., 2014. Accelerated flooding along the U.S. East Coast: On the impact of sea-level rise, tides, storms, the Gulf Stream, and the North Atlantic Oscillations. *Earth's Futur.* 2, 362–382. <https://doi.org/10.1002/2014EF000252>.
- Federal Emergency Management Agency (FEMA), 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>.
- Gaitan, S., van de Giesen, N.C., ten Veldhuis, J.A.E., 2016. Can urban pluvial flooding be predicted by open spatial data and weather data? *Environ. Model. Softw.* 85, 156–171. <https://doi.org/10.1016/j.envsoft.2016.08.007>.
- Goddard, P.B., Yin, J., Griffies, S.M., Zhang, S., 2015. An extreme event of sea-level rise along the Northeast coast of North America in 2009–2010. *Nat. Commun.* 6, 6346. <https://doi.org/10.1038/ncomms7346>.
- Gradeci, K., Labonnote, N., Sivertsen, E., Time, B., 2019. The use of insurance data in the analysis of Surface Water Flood events – A systematic review. *J. Hydrol.* 568, 194–206. <https://doi.org/10.1016/J.JHYDROL.2018.10.060>.
- Hallegatte, S., Green, C., Nicholls, R.J., Corfee-Morlot, J., 2013. Future flood losses in major coastal cities. *Nat. Clim. Chang.* 2013 39 3, 802–806. 10.1038/nclimate1979.
- Kousky, C., 2018. Financing Flood Losses: A Discussion of the National Flood Insurance Program. *Risk Manag. Insur. Rev.* 21, 11–32. <https://doi.org/10.1111/RMIR.12090>.
- 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>.
- Leal, M., Boavida-Portugal, I., Fragoso, M., Ramos, C., 2019. How much does an extreme rainfall event cost? Material damage and relationships between insurance, rainfall, land cover and urban flooding. <https://doi.org/10.1080/02626667.2019.1595625> 64, 673–689. 10(1080/02626667), pp. 1595625, 2019.
- Maier, R., Krebs, G., Pichler, M., Muschalla, D., Gruber, G., 2020. Spatial Rainfall Variability in Urban Environments—High-Density Precipitation Measurements on a City-Scale. *Water* 2020, Vol. 12, Page 1157 12, 1157. 10.3390/W12041157.
- 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>.
- Mobley, W., Sebastian, A., Blessing, R., Highfield, W.E., Stearns, L., Brody, S.D., 2021. Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: A pilot study in southeast Texas. *Nat. Hazards Earth Syst. Sci.* 21, 807–822. <https://doi.org/10.5194/NHESS-21-807-2021>.
- Moftakhari, H.R., AghaKouchak, A., Sanders, B.F., Feldman, D.L., Sweet, W., Matthew, R. A., Luke, A., 2015. Increased nuisance flooding along the coasts of the United States due to sea level rise: Past and future. *Geophys. Res. Lett.* 42, 9846–9852. <https://doi.org/10.1002/2015GL066072>.
- Mosavi, A., Ozturk, P., Chau, K.W., 2018. Flood prediction using machine learning models: Literature review. *Water (Switzerland)*. <https://doi.org/10.3390/w10111536>.
- NOAA National Centers for Environmental Information (NCEI), 2022. U.S. Billion-Dollar Weather and Climate Disasters [WWW Document]. 10.25921/stkw-7w73.
- Netusil, N.R., Kousky, C., 2021. The Coming Storm: How U.S. Cities Are Managing Stormwater from Increasingly Extreme Rainfall Events.
- Nofal, O.M., van de Lindt, J.W., 2020. Understanding flood risk in the context of community resilience modeling for the built environment: research needs and trends. <https://doi.org/10.1080/23789689.2020.1722546>. 10(1080/23789689), pp. 1722546, 2020.
- 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>.
- Saharia, M., Kirstetter, P.E., Vergara, H., Gourley, J.J., Hong, Y., 2017. Characterization of floods in the United States. *J. Hydrol.* 548, 524–535. <https://doi.org/10.1016/J.JHYDROL.2017.03.010>.
- 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>.
- Smith, J.A., Baeck, M.L., Meierdiercks, K.L., Miller, A.J., Krajewski, W.F., 2007. Radar rainfall estimation for flash flood forecasting in small urban watersheds. *Adv. Water Resour.* 30, 2087–2097. <https://doi.org/10.1016/j.advwatres.2006.09.007>.
- Sørensen, J., Mobini, S., 2017. Pluvial, urban flood mechanisms and characteristics – Assessment based on insurance claims. *J. Hydrol.* 555, 51–67. <https://doi.org/10.1016/J.JHYDROL.2017.09.039>.
- Spekkers, M.H., Kok, M., Clemens, F.H.L.R., Ten Veldhuis, J.A.E., 2013. A statistical analysis of insurance damage claims related to rainfall extremes. *Hydrol. Earth Syst. Sci.* 17, 913–922. <https://doi.org/10.5194/HESS-17-913-2013>.
- Spekkers, M.H., Clemens, F.H.L.R., Ten Veldhuis, J.A.E., 2015. On the occurrence of rainstorm damage based on home insurance and weather data. *Nat. Hazards Earth Syst. Sci.* 15, 261–272. <https://doi.org/10.5194/NHESS-15-261-2015>.
- Sweet, W. (William V.), Dusek, G. (Gregory P.), Obeysekera, J.T.B., Marra, J.J., 2018. Patterns and projections of high tide flooding along the U.S. coastline using a common impact threshold. NOAA Tech. Rep. NOS CO-OPS 086. <https://doi.org/10.7289/V5/TR-NOS-COOPS-086>.
- Tate, E., Rahman, M.A., Emrich, C.T., Sampson, C.C., 2021. Flood exposure and social vulnerability in the United States. *Nat. Hazards* 106, 435–457. <https://doi.org/10.1007/S11069-020-04470-2/TABLES/6>.
- Torgersen, G., Bjerkholt, J.T., Kvaal, K., Lindholm, O.G., 2015. Correlation between extreme rainfall and insurance claims due to urban flooding—case study Fredrikstad. *Norway. J. Urban Environ. Eng.* 9, 127–138.
- Vishnu, N., Wojtkiewicz, R., Ramanathan, K., 2021. Understanding flood vulnerability of buildings in US: A perspective through claims data. FLOODrisk 2020 - 4th Eur. Conf. Flood Risk Manag. null-null. 10.3311/FLOODRISK2020.11.14.
- Wahl, T., Jain, S., Bender, J., Meyers, S.D., Luther, M.E., 2015. Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nat. Clim. Chang.* 5, 1093–1097. <https://doi.org/10.1038/nclimate2736>.
- Wang, Y., Sebastian, A., 2021. Community flood vulnerability and risk assessment: An empirical predictive modeling approach. *J. Flood Risk Manag.* 14, e12739.
- Wing, O.E.J., Pinter, N., Bates, P.D., Kousky, C., 2020. New insights into US flood vulnerability revealed from flood insurance big data. *Nat. Commun.* 11 <https://doi.org/10.1038/s41467-020-15264-2>.
- Yang, Q., Shen, X., Yang, F., Anagnostou, E.N., He, K., Mo, C., Seyyedi, H., Kettner, A.J., Zhang, Q., 2022. Predicting Flood Property Insurance Claims over CONUS, Fusing Big Earth Observation Data. *Bull. Am. Meteorol. Soc.* 103, E791–E809. <https://doi.org/10.1175/BAMS-D-21-0082.1>.
- Zischg, A.P., Mosimann, M., Bernet, D.B., Röthlisberger, V., 2018. Validation of 2D flood models with insurance claims. *J. Hydrol.* 557, 350–361. <https://doi.org/10.1016/J.JHYDROL.2017.12.042>.