

PAPER • OPEN ACCESS

Comparing multi-source urban flood indicators: satellite, simulation, and citizen-reported data

To cite this article: Jangho Lee *et al* 2025 *Environ. Res.: Water* **1** 035007

View the [article online](#) for updates and enhancements.

You may also like

- [Diagnostic analysis of waterlogging in Zhenjiang City by using PCSWMM](#)
Q Q Yang, Z D Bao, Y Fu et al.
- [From managing risk to increasing resilience: a review on the development of urban flood resilience, its assessment and the implications for decision making](#)
Viktor Rözer, Sara Mehryar and Swenja Surminski
- [The Initial Rain Water Storage Tank Volume Determination and Pollution Interception Efficiency Based on the Swmm Model and Computer Software Analysis](#)
Ling Shi, Honghai Liu, Chen Xi et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology



**249th
ECS Meeting**
May 24-28, 2026
Seattle, WA, US
*Washington State
Convention Center*

Spotlight Your Science

***Submission deadline:
December 5, 2025***

SUBMIT YOUR ABSTRACT

ENVIRONMENTAL RESEARCH WATER



PAPER

OPEN ACCESS

RECEIVED
18 July 2025

REVISED
12 September 2025

ACCEPTED FOR PUBLICATION
24 September 2025

PUBLISHED
14 October 2025

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Comparing multi-source urban flood indicators: satellite, simulation, and citizen-reported data

Jangho Lee^{1,*} , Sun Young Park^{2,8} , Abhinav Wadhwa^{3,8} , Aaron I Packman⁴ , Stephen W Nesbitt⁵ , Ashish Sharma^{3,5,6} , Max Berkelhammer¹ , Marcelo H Garcia² , Rao Kotamarthi⁶ , Deanna Hence⁵ and William M Miller⁷

¹ Department of Earth and Environmental Sciences, University of Illinois Chicago, Chicago, IL 60607, United States of America

² Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, United States of America

³ Discovery Partners Institute, University of Illinois Urbana-Champaign, Chicago, IL 60606, United States of America

⁴ Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL 60208, United States of America

⁵ Department of Climate, Meteorology & Atmospheric Sciences, University of Illinois Urbana-Champaign, Urbana, IL 61801, United States of America

⁶ Environmental Science Division, Argonne National Laboratory, Lemont, IL 60439, United States of America

⁷ Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, United States of America

⁸ These authors contributed equally to this work.

* Author to whom any correspondence should be addressed.

E-mail: jholee@uic.edu

Keywords: SAR, SWMM, 311, citizen science, urban flooding

Supplementary material for this article is available [online](#)

Abstract

Urban flooding arises from complex mechanisms, making it challenging to capture accurately with a single detection method. This study evaluates three complementary approaches to detect flooding across three Chicago neighborhoods: (i) Sentinel-1 synthetic aperture radar (SAR), offering weather-independent, high-resolution (10 m) imagery of surface inundation; (ii) the storm water management model (SWMM), simulating combined sewer overflow and drainage performance; and (iii) citizen-generated 311 service requests, capturing observed flooding impacts. By analyzing six storms ranging from severe to mild, we examine how each source uniquely contributes to identifying urban flood events. SAR imagery effectively identifies standing water but can miss brief flooding due to satellite revisit constraints. SWMM provides detailed insights into system-wide drainage behavior yet may underestimate localized street-level flooding. Meanwhile, 311 calls reflect real-world flooding impacts but are vulnerable to underreporting. Statistical overlap analysis highlights chronic flood hotspots repeatedly identified across multiple detection methods, indicating persistent infrastructure and topographic vulnerabilities. Temporal analysis further reveals that while SWMM flooding aligns closely with rainfall peaks, 311 calls typically precede or persist beyond these peaks. Our findings emphasize the value of using satellite observations, hydrological modeling, and resident-reported data in a complementary manner to better interpret patterns in flood timing, severity, and spatial distribution—providing insights that can inform targeted infrastructure improvements and contribute to urban flood resilience planning.

1. Introduction

1.1. Background and motivation

Flooding ranks among the most damaging natural disasters globally, causing substantial economic losses and significant human impacts (IFRC 2018, Hamers *et al* 2023). These events include pluvial (flash and urban), coastal, and fluvial (river) floods. Furthermore, flooding often is compounded with other hazards and consequences, such as landslides, agricultural loss, degradation of water supply, and public health issues,

resulting in compounded losses which are much more than from flooding alone (AghaKouchak *et al* 2020, Zhang and Villarini 2020, Ming *et al* 2022).

Urban flooding refers to flooding that occurs in developed areas due to overloading of local drainage systems rather than overflow of natural water bodies (Hammond *et al* 2015, Mignot *et al* 2019, Qin 2020, Balaian *et al* 2024), and it poses significant challenges for cities worldwide (Tingsanchali 2012, Huong and Pathirana 2013, National Academies of Sciences 2019, Zhang and Villarini 2020). Unlike riverine or coastal floods, urban floods are often hyper-local, sometimes flooding one street or building while the adjacent area remains dry. A major reason is the prevalence of impervious surfaces (roads, rooftops, parking lots) in cities, which prevent rainfall from infiltrating into soil. Instead, rainwater rapidly runs off, leading to large volumes of surface flow in a short time. If storm sewers and drains cannot carry this runoff away, water accumulates in streets and basements (Sörensen and Mobini 2017, Huang *et al* 2018, Wang *et al* 2022, Zhang *et al* 2023). Many urban drainage systems around the world were built decades ago and are now aging and struggling to cope with modern extremes (Marsalek *et al* 1993, Delleur 2003, Piadeh *et al* 2022). These older systems often operate under outdated capacity assumptions, making them vulnerable to extreme precipitation, increased impervious cover, and infrastructure failures (e.g. sediment-clogged or collapsed pipes) causing property damage, traffic disruptions, and public health hazards (Lane *et al* 2013, Fernandez *et al* 2015, Singh *et al* 2018, Zhong *et al* 2018, Zhu *et al* 2018).

1.2. Study area

The city of Chicago in Illinois, United States, presents a compelling case study of urban flooding due to its unique hydrological context (Markus *et al* 2012, Zhu *et al* 2016, Cousins 2017, Wing *et al* 2020). Chicago is a coastal megacity (metro area population 9.3 M) located on the shore of Lake Michigan, the world's third-largest freshwater lake in the world by its surface area. The city is served by a combined sewer system that carries both stormwater and sanitary sewage in the same pipes. During heavy rain, this combined system is prone to becoming overwhelmed. When these pipes and storage tunnels surcharge, the excess flow is released as combined sewer overflows (CSOs) into local waterways. To address this, Chicago invested in a massive Tunnel and Reservoir Plan (TARP, or 'Deep Tunnel') to hold excess stormwater (Dalton and Kenny 1980, Scalise and Fitzpatrick 2012). Even after decades of construction and expansion of this system, moderate storms still trigger CSOs and urban floods. The city's hydrology dynamics is further complicated by its relationship with Lake Michigan. Historical engineering projects reversed the flow of the Chicago River to send wastewater out of the Lake Michigan basin. However, during extreme storm events, the city sometimes must open emergency outlets to divert floodwaters into Lake Michigan, which is Chicago's source of drinking water. These backflows and overflows pose environmental and health risks, as they can contaminate both inland rivers and the lake with untreated sewage. In summary, Chicago's aging combined sewer infrastructure, extensive paved surfaces, and proximity to Lake Michigan create a perfect storm of factors that make urban flooding a frequent and pressing challenge (National Academies of Sciences 2019).

1.3. Approaches to flood detection and management

Addressing this challenge requires accurate detection and prediction of flood events. Yet, this is complicated by limited data availability, the interplay between natural and built environments, and the complex hydrodynamic processes that govern water movement in highly developed areas. In a multi-source urban flood monitoring context, an additional difficulty is that available datasets differ fundamentally in what they measure, as well as in their spatial and temporal resolutions and susceptibility to specific biases. Direct comparisons therefore require careful alignment and interpretation, with the understanding that these datasets do not necessarily capture the same aspects of a flood event.

Satellite-based remote sensing has emerged as a powerful tool for flood detection and mapping (Mason *et al* 2009, Wu *et al* 2012, Tanim *et al* 2022). It offers wide-area coverage and access to flooded regions that may be dangerous or inaccessible on the ground. Satellites equipped with synthetic aperture radar (SAR) (e.g. Sentinel-1, RADARSAT) actively emit microwave signals toward Earth and measure the return signal (backscatter) (Tsyganskaya *et al* 2018, Amitrano *et al* 2024). Radar has the advantage of being weather-independent—microwaves penetrate clouds and are not reliant on sunlight, so SAR can image floods at night or through storm clouds. Floodwater is often detectable in SAR because a calm water surface reflects the radar signal away (acting like a mirror), returning little energy to the satellite and appearing as a dark area. This makes SAR extremely useful for flood mapping, especially when optical images are unavailable due to cloud cover. In terms of spatial detail, modern SAR offers resolutions of 10 m or better, comparable to optical methods (Uddin *et al* 2019, Bauer-Marschallinger *et al* 2022, Islam and Meng 2022). However, the temporal resolution of SAR data can be low; Sentinel-1, for instance, has a 6–12 day revisit period, and it can cause difficulty in providing a complete time series of flood progression. The additional challenge with SAR is that city landscapes produce very complex radar backscatter: buildings, vehicles, and

other structures cause multiple reflections, bright spots, and shadows in the radar image. Floodwater in urban streets may be obscured by the ‘clutter’ of strong returns from buildings or even appear as bright radar returns due to corner reflections (Notti *et al* 2018, Shen *et al* 2019, Amitrano *et al* 2024).

Another approach to urban flood monitoring and prediction is through hydrological and hydraulic modeling (Mignot *et al* 2019, Bulti and Abebe 2020, Nkwunonwo *et al* 2020, Luo *et al* 2022a, 2022b). These physics-based models simulate how rainfall becomes runoff and how that runoff moves through urban drainage networks. A prominent example is the United States Environmental Protection Agency’s storm water management model (SWMM) (Jiang *et al* 2015, Bisht *et al* 2016, Rai *et al* 2017). SWMM tracks water movement from rainfall to runoff to flow in pipes and can simulate surcharging manholes and overflows onto streets. By simulating an entire storm event, the model can predict where and when flooding will occur in the urban system, given the network configuration. A key strength of SWMM is its balance between physical detail and computational efficiency. It incorporates the essential physics and infrastructure of urban drainage, so it can accurately capture phenomena like backwater effects in pipes or detention basin storage. However, the accuracy of a SWMM model heavily depends on the quality of input data (e.g. rainfall patterns and rates) and calibration. Building a detailed SWMM model for a city like Chicago requires extensive data, such as maps of the sewer network (e.g. pipe diameters, invert elevations, connectivity), land use data to estimate imperviousness and hydrologic parameters, and high-resolution rainfall data (from rain gauges or weather radar) to drive the simulations.

In recent years, crowdsourced information has emerged as a valuable source of data for urban flood monitoring, such as 311 calls (Elmore *et al* 2014, 2015, Mazzoleni *et al* 2017, Dixon *et al* 2021, Helmrich *et al* 2021, Lee *et al* 2025). In many cities, including Chicago, residents report flooding issues through 311 service requests—a non-emergency municipal hotline or mobile phone application for city services (Rainey *et al* 2021, Agonafir *et al* 2022a, 2022b, Negri *et al* 2023). These 311 flood reports are essentially citizen-generated data points indicating where and when flooding has been observed on the ground. This approach to flood monitoring leverages the observation capability of the community to capture localized flood impacts in real time. Because they are reported by citizens experiencing the event, 311 flood reports can provide very timely information at a fine spatial resolution (often an exact address or intersection). They can fill gaps between city-managed monitoring systems and research-grade measurements by providing additional, real-time observations. Furthermore, 311 data directly indicate impacts on residents and infrastructure. This human perspective is crucial for emergency response and damage assessment. However, not all floods are reported by 311 calls. The likelihood of a resident calling 311 can depend on awareness, trust in city services, and socio-economic factors. Research has found that 311 flood data may be biased by neighborhood demographics—some high-risk areas generate surprisingly few reports, possibly due to residents feeling nothing will be done, while other areas report more frequently for even minor issues (Wang *et al* 2017, White and Trump 2018, Kontokosta and Hong 2021). Thus, absence of 311 reports does not always mean absence of flooding; it might reflect under-reporting. Conversely, communities with higher levels of digital access, civic engagement, and familiarity with the 311 system may generate multiple reports for the same flooded intersection. Furthermore, a limitation of relying on 311 alone is that it provides a binary indication (flood happened or not, at a location) but not the depth or extent of flooding. It is thus hard to quantify severity from a service request without additional context (Choi *et al* 2018, White and Trump 2018, Hagen *et al* 2019).

1.4. Study objectives

Given the three distinct flood detection approaches outlined above, this study aims to compare three independent flood indicators—satellite-based SAR imagery, physics-based SWMM simulation, and citizen-reported 311 service requests—in an aligned spatial and temporal framework. Our goal is to evaluate their relative strengths, limitations, and areas of convergence, rather than to define an absolute ‘true’ flood extent, which is not available for Chicago. We focus on identifying recurrent hotspots where multiple datasets independently indicate flooding, as well as divergence patterns that reveal data-specific biases and uncertainties. In this context, ‘integrated approach’ refers to triangulating multiple independent indicators to identify consistent spatial patterns and systematic divergences, rather than combining datasets through physical data assimilation. This framing reflects our aim to reveal complementary insights from datasets that capture related but not identical aspects of flooding.

In this study, we will examine where these data sources converge or diverge in detecting flood events across focus communities of Chicago. By analyzing past flood events, we will identify cases where one data source catches flooding that others miss and where overlap occurs across datasets. The comparative assessment will highlight the strengths and weaknesses of each. More importantly, we seek to demonstrate their complementarity—how combining these approaches can lead to a more robust urban flood monitoring system. Ultimately, the insights derived from this study will inform improved flood management strategies. By leveraging the relative strengths of satellites, models, and crowdsourced data together, city officials and

emergency managers can move toward a more reliable and timely urban flood monitoring and warning system. The overarching objective is to demonstrate that an integrated approach—in this case, for the city of Chicago—offers a more complete picture of urban flooding than any single data source alone, thereby enhancing the city’s ability to prepare for, respond to, and recover from flood events.

2. Data

2.1. Study region and dates

For the analysis, we focus on the three community regions in Chicago: Humboldt Park, Woodlawn, and Chatham (figure 1). These sites are three historically flood-prone community areas in Chicago. Their combined sewer networks, constructed in the late nineteenth and early twentieth centuries, face capacity challenges when intense rainfall events exceed system design thresholds, leading to basement backups and street ponding. Humboldt Park’s dense impervious surfaces channel runoff quickly into undersized sewers, causing localized flash flooding (Lee *et al* 2024, Mohamed *et al* 2025). Woodlawn, located along the South Side lakefront, is characterized by predominantly impermeable soils and limited natural drainage capacity, which exacerbates flood vulnerability when combined with its proximity to Lake Michigan (McGuire 2018). In Chatham, which sits on low-lying land, stormwater readily pools, causing repeated damage to homes that lie below street grade (Keenan *et al* 2019, Feldman 2021). Beyond these physical conditions, the three neighborhoods also share socioeconomic challenges, including aging housing stock, higher proportions of low- to moderate-income households, and historic underinvestment in infrastructure, which collectively increase vulnerability to flood impacts (Cousins 2017, Keenan *et al* 2019, Wing *et al* 2020). Their documented history of severe inundations offers valuable data for assessing different flood detection methods, while their diverse geographic and social settings make them representative of broader citywide issues.

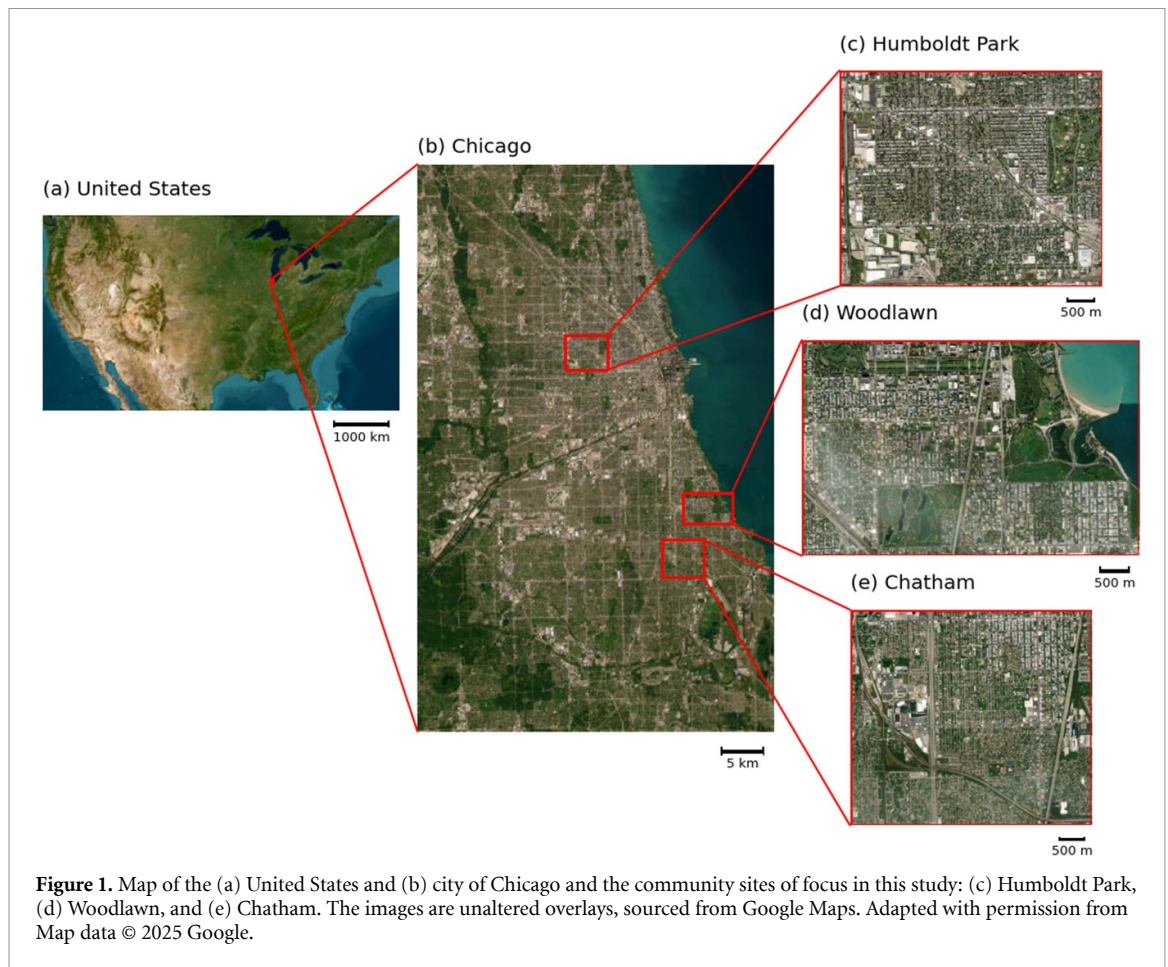
In selecting target days for analysis, we aimed to capture a range of flood severities—from severe to mild—while also accommodating the revisit interval and data availability constraints of Sentinel-1 SAR imagery. Events were chosen based on the following criteria: (i) classification as local storms by the National Weather Service (NWS), (ii) availability of SAR scenes with minimal cloud contamination, and (iii) coverage across the severity spectrum (high, moderate, and low precipitation). The six selected events are 8 May 2019 (high precipitation), 20 May 2019 (high), 1 June 2019 (high), 17 September 2023 (high), 14 April 2019 (moderate), and 9 March 2019 (low precipitation). We note that only one moderate and one low-precipitation event were included; as such, conclusions for these categories should be interpreted with caution. The first four events listed above (17 September 2023; 1 June 2019; 20 May 2019; and 8 May 2019) were classified as severe storms (exceeding the 90th percentile in daily maximum precipitation, averaged over Chicago metropolitan region, based on the historical record), 14 April represented a moderate storm (above the 75th percentile), and 9 March featured low precipitation (below the 25th percentile). This carefully curated selection ensures a robust and diverse evaluation of urban flood conditions that aligns with the NWS-defined impactful weather events.

2.2. Precipitation data

The precipitation data in this study come from the Multi-Radar/Multi-Sensor (MRMS) system. MRMS system provides high-resolution precipitation data with exceptional spatial (1 km) and temporal (2 min) resolution (Zhang *et al* 2016). By integrating radar networks, surface and upper-air observations, lightning detection systems, satellite imagery, and forecast models, MRMS offers comprehensive and robust precipitation estimates across North America. Studies confirm its reliability in capturing spatial and temporal rainfall dynamics when validated against ground-based measurements and single radar systems (Zhang *et al* 2016, Bytheway *et al* 2019, Habibi *et al* 2021, Moazami and Najafi 2021). The typical error rates of MRMS compared to *in situ* measurements are reported to be under 10 mm in RMSE or 5%–10% of precipitation rate (Zhang *et al* 2016, Rivera-Giboyeaux and Weinbeck 2024). We validated the MRMS product with available NWS weather station around Chicago region for the selected study period and it showed average of 7.3 mm of RMSE and 3% error of precipitation rate.

2.3. Remote sensing data (SAR)

SAR imagery used in this study originates from Sentinel-1 satellites (Geudtner *et al* 2014, Mullissa *et al* 2021). The preprocessing workflow begins with thermal noise removal using specialized filters designed to enhance the signal-to-noise ratio, followed by radiometric calibration to correct sensor-dependent variations and mitigate atmospheric influences (Filipponi 2019). Speckle filtering techniques, such as Lee and Gamma MAP filters, are then applied to reduce granular noise inherent to SAR data while preserving important spatial resolution and textural details (Lee *et al* 1994). Additionally, geometric correction and precise image registration are carried out to align multi-temporal SAR datasets, facilitating accurate temporal comparisons



of flood extents. Collectively, these preprocessing steps optimize the SAR imagery for robust and reliable flood detection and analysis (Mullissa *et al* 2021).

Subsequent to preprocessing, convolutional neural networks (CNNs) are employed to automatically classify flooded and non-flooded areas by learning complex patterns directly from the SAR backscatter data, thereby eliminating reliance on predefined threshold values (Fan *et al* 2019, Sharma and Saharia 2025). CNN utilizes convolutional layers with local receptive fields to extract spatial features effectively from SAR data, analyzing spatial distributions and textures of backscatter intensities. The network was trained specifically for this study on a labeled dataset derived from urban flooding events in Chicago, enabling it to identify subtle differences among water bodies, flooded regions, and other urban land-cover types through temporal analysis of SAR imagery. The CNN model dynamically adjusts decision thresholds based on learned features, thus accurately identifying flooded areas even in complex urban settings characterized by radar-specific artifacts, such as shadowing effects or double-bouncing signals. This adaptive, context-sensitive thresholding significantly enhances flood-mapping accuracy and efficiency in heterogeneous urban environments. We emphasize that all SAR data utilized were processed entirely in-house from raw Sentinel-1 Level 1 ground range detected products rather than external preprocessed datasets.

Although Sharma and Saharia (2025) are referenced to contextualize the broader CNN-based methodological framework, the CNN classifier in this research was independently developed, trained, and applied specifically to this study achieving an overall accuracy of 81%. Additional methodological details, including comprehensive CNN model parameters and training procedures, are available in the supplementary material.

Following the collection of the high-resolution (10 m) flood extent data, we preprocessed it to align with other data sources by aggregating values onto a 0.025° (approximately 250 m) latitude–longitude grid over the three selected sites. For each grid cell, we calculated the fraction of flood-inundated surface area to represent the magnitude of flooding, but excluded portions covered by buildings or permanent water bodies. Building footprints were obtained from the Chicago Data Portal (Kassen 2013), and permanent water coverage came from National Oceanic and Atmospheric Administration's (NOAA) C-CAP archive (NOAA). As a result, the preprocessed SAR product represents the fraction of exposed land surface that is flooded. Grid cells in which more than 90% of the area is occupied by buildings or permanent water were masked out

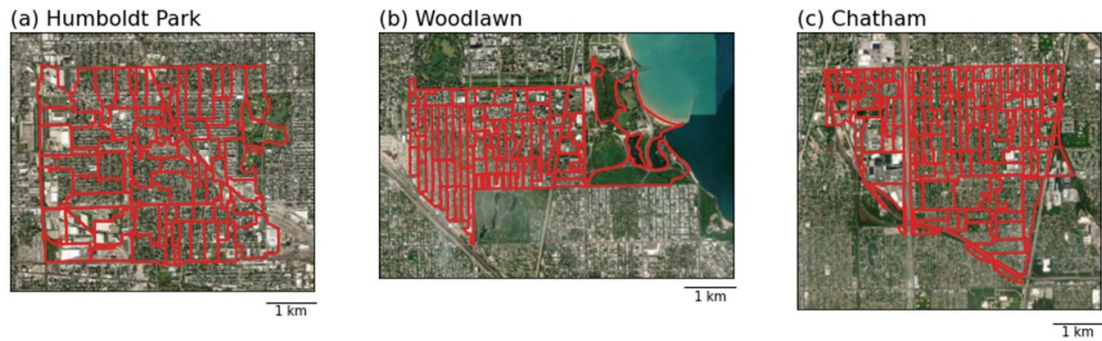


Figure 2. SWMM subcatchment boundaries (red) overlaid with satellite images across the three study neighborhoods: (a) Humboldt Park, (b) Woodlawn, and (c) Chatham. The images are unaltered overlays from Google Maps. Adapted with permission from Map data © 2025 Google.

in the final dataset. In this study, a pixel is considered flooded in SAR when the CNN-classified flood fraction exceeds the binary threshold (0.05) after excluding buildings (32% of total data) and permanent water bodies (3% of total data). This definition applies to all detections within the event period, regardless of duration, and results are aggregated at the event level.

2.4. SWMM model outputs

The SWMM model used in this study incorporates a detailed representation of the urban drainage system, simulating hydrologic and hydraulic processes across the Chicago region (Cantone and Schmidt 2011, Naniá et al 2015, Luo et al 2021). It consists of 699 subcatchments, 32 441 nodes, and 34 051 links, including key hydraulic components such as CSO junctions, drop shafts, pumps, orifices, weirs, gates, and storage units. High-resolution MRMS rainfall data from 633 rain grids drive the hydrological processes, accurately capturing the spatial variability of rainfall across the study area.

The model employs the Modified Green–Ampt infiltration method (Almedeij and Esen 2014, Mao et al 2016) to dynamically account for soil moisture conditions, providing precise rainfall partitioning between infiltration and surface runoff. Dynamic wave flow routing (Rossman and Supply 2006) simulates stormwater movement within the drainage system, capturing unsteady flow conditions, backwater effects, surcharging, and flow reversals typical of complex urban drainage systems. The EXTRAN surcharge method assesses node capacity exceedance, identifying urban flooding hotspots when infrastructure becomes overwhelmed (Pachaly et al 2020). Surface ponding in depressions is explicitly represented, accounting for temporary water accumulation.

SWMM-derived flood values are initially associated with corresponding subcatchments (figure 2), with flooding assumed to affect entire drainage areas. Excess runoff movement is tracked through the stormwater network, computing potential surcharging and surface ponding locations. The routing component evaluates conditions at over 32 000 nodes, pinpointing infrastructure limitations. Model simulations consistently exhibited low continuity errors, with runoff quantity errors below 0.05% (where the error is calculated as following equation (1) below) across events. These low errors affirm the reliability and numerical stability of hydrologic and hydraulic computations, accurately capturing the severity and distribution of modeled flooding (i.e. the number and density of nodes exceeding capacity and associated ponding volumes) under varying storm conditions, rather than literal inundation footprints.

$$\text{Continuity Error (\%)} = 100 \times \frac{\text{Inflow} - (\text{Outflow} + \text{Final Storage})}{\text{Inflow}}. \quad (1)$$

However, SWMM has inherent limitations in representing certain urban flooding scenarios, primarily capturing flooding within stormwater networks and not simulating extensive overland flow beyond drainage infrastructure. Consequently, flooding from poor surface drainage, blocked storm inlets, basement flooding, or inadequate storm drain maintenance might be underestimated.

For compatibility with remote sensing data, SWMM flood outputs are interpolated onto a 250 m grid, enabling direct comparison with SAR-derived flood extents. To aggregate SWMM results onto the 250 m grid, we applied an area-weighted averaging procedure: the flooding volume estimated for each subcatchment was distributed across the grid cells intersecting that subcatchment, proportional to the fraction of subcatchment area contained in each cell. The resulting flood value for each grid cell represents the sum of these contributions, normalized by the grid-cell area. This method ensures that the gridded field reflects the spatial distribution of modeled flooding severity without bias toward larger subcatchments. Grid

cells without overlapping catchment areas were masked to maintain data alignment. We emphasize that SWMM outputs represent a process-based scenario, not a direct observation. In the absence of neighborhood-scale *in situ* measurements, we verified plausibility through internal continuity checks and agreement in flood-prone zones with documented historical problem areas from City of Chicago reports.

For further discussion on SWMM model parameterization and validation, refer to the supplementary section S2. A ‘flood’ in SWMM output refers to any modeled surcharge or surface ponding at a node within the event window, regardless of whether the exceedance lasted for minutes or hours. All such exceedances are included equally in the spatial analysis. The SWMM output is presented in cubic feet per second (cfs) and represents the overflow rate from surcharged nodes in the drainage network. This value quantifies the amount of water that exceeds the sewer system’s capacity and is discharged onto the surface, directly indicating locations of model-predicted street-level flooding.

2.5. 311 Service requests

We obtained 311 service requests data from the Chicago Data Portal, which is a public data resource where the City of Chicago shares all of its municipal records and datasets (Kassen 2013). Each record contains the nature of the request, the date and time of submission, and latitude–longitude coordinates. Specifically for basement and street flooding, we filtered requests containing the phrases ‘water in basement’ or ‘water on street.’

For the spatial analysis, we used each request’s location for both the flooding event day and the following day. Next, we applied a Gaussian smoothing technique ($\alpha = 0.5$) on a 1 m resolution grid, where the smoothing parameter α controls the spatial spread of influence from each report location (Diggle 1985, Nychka 2000). A value of $\alpha = 0.5$ was selected to represent localized flooding effects within a limited radius while avoiding excessive smoothing that could dilute spatial precision. We then aggregated the smoothed results to 250 m grid cells, consistent with our other data sources. This two-step process ensures a more spatially continuous representation of 311 flood-related requests, thereby providing a broader view of localized flooding patterns.

2.6. Summary of resolution alignment

To enable direct, quantitative comparison across datasets with different native resolutions and temporal characteristics, we aligned both dimensions before analysis.

- Temporal alignment: SAR scenes capture standing water conditions at a single overpass time; SWMM outputs represent instantaneous surcharges at 5 min intervals; 311 service requests are logged at the time of report, which may lag flood onset by minutes to hours. For each flood event, SAR imagery was selected at the closest acquisition to peak rainfall, SWMM outputs within ± 1 h of that time, and 311 reports from the event day plus the following day were included to capture residual flooding.
- Spatial alignment: All datasets were aggregated to a 0.025° (~ 250 m) grid as mentioned in previous sections.

3. Methods

3.1. Statistical methods

In this section, we present a detailed analysis of spatial agreement between flood datasets using both binary and continuous metrics that are widely used in spatial statistics and flood related works as an evaluation tool (Rahmati *et al* 2020, Deidda *et al* 2021, Safavi *et al* 2021, Darabi *et al* 2022, Macharyulu *et al* 2022, Rydén 2022, Chandran *et al* 2024). The rationale for employing both types of metrics is to capture complementary aspects of agreement: binary metrics focus on the spatial overlap of detected flood areas, while continuous metrics assess the similarity in the magnitude and spatial distribution of flood intensities.

3.1.1. Binary metrics

Binary metrics are calculated by first thresholding each dataset to create binary flood maps. The thresholds—SAR (0.05), SWMM (0.0), and 311 (0.01)—were determined using the 10th percentile of pixel values from multiple events, combined with visual inspection of flood maps, to ensure that retained pixels represented meaningful flooding signals while minimizing noise. After thresholding, a pixel is considered ‘flooded’ if its value exceeds the threshold, and ‘not flooded’ otherwise. This approach allows for direct comparison of spatial patterns, even when the original data are on different scales or units. Here we apply three binary metrics: Matthews correlation coefficient (MCC) (Chicco and Jurman 2020), Intersection over Union (IoU) (Rezatofighi *et al* 2019), and Dice Coefficient (Shamir *et al* 2019).

MCC is a robust measure of binary classification agreement, particularly valuable in the presence of class imbalance. It is defined as:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \quad (2)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. In cases of little or no detected flooding, MCC effectively accounts for class imbalance and offers a balanced view of agreement by simultaneously considering both flooded and non-flooded pixels. However, MCC frequently becomes undefined in low-flooding scenarios due to its strict dependence on nonzero counts in all contingency-table elements. Hence, its applicability is somewhat limited to scenarios where flooding is extensive enough to produce sufficient positive detections.

IoU, or the Jaccard Index, quantifies the spatial overlap between two binary masks:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \quad (3)$$

IoU values range from 0 (no overlap) to 1 (perfect overlap). Generally, IoU values peak during major flood events due to a higher proportion of true-positive pixels, signifying strong spatial alignment between datasets. Conversely, during minor or localized flood events, IoU values typically decline sharply because overlaps are minimal. IoU is intuitive, widely understood, and directly interpretable as the proportion of overlapping flooded areas. However, IoU is particularly sensitive to the selected threshold levels and can be harsh when overlaps are limited, potentially underestimating agreement when flood areas are small or spatially fragmented.

Dice Coefficient, or the F1 score, is another metric of spatial overlap:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \quad (4)$$

Dice values also span from 0 (no overlap) to 1 (perfect overlap). Dice Coefficient, closely related to IoU, generally provides slightly more forgiving overlap assessments due to its greater weighting of true positives. While Dice shares IoU's threshold sensitivity, its slightly lenient nature is beneficial when evaluating events characterized by small-scale flooding, thereby offering a complementary perspective alongside IoU.

3.1.2. Continuous metrics

Continuous metrics assess the similarity in magnitude and spatial patterns of flooding intensities between datasets, extending beyond the binary overlap approach. Due to inherent differences in measurement scales among the SWMM, SAR, and 311 datasets—which respectively use overflow rate (cfs), flood fraction, and smoothed request density (i.e. the Gaussian-smoothed density of 311 flood reports)—all values were standardized to a common 0–1 range prior to analysis. This step was necessary to enable a direct comparison of relative magnitudes and spatial patterns for the metric calculations. The metrics employed include Pearson's correlation coefficient (Sedgwick 2012), Kendall's τ coefficient (Noether 1981), and Willmott's Index of Agreement (Willmott *et al* 2012).

Pearson's r measures the linear relationship between the continuous flood intensity values of two datasets:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (5)$$

High positive r values indicate strong linear agreement in flood intensity patterns, while negative values indicate inverse relationships. Pearson's correlation coefficient is highly informative regarding linear relationships between flood intensity distributions, particularly useful when datasets exhibit strong linear associations. However, Pearson's coefficient assumes normality and linearity of data distributions, making it sensitive to outliers and potentially misleading when flood intensities have skewed distributions or nonlinear relationships.

Kendall's τ coefficient can be calculated as follows:

$$\tau = \frac{(\text{Number of concordant pairs}) - (\text{Number of discordant pairs})}{\frac{1}{2}n(n-1)} \quad (6)$$

where a pair is concordant if the order of values is the same in both datasets, and discordant otherwise. A positive τ indicates consistent ranking order between datasets, while negative τ values suggest disagreements

in intensity rankings across pixels. Kendall's τ provides robust ordinal-level insights into dataset similarity, as it does not require normality or linearity assumptions. This robustness makes it well-suited for datasets with non-normal or non-linear intensity distributions. Nevertheless, τ captures only rank agreement rather than actual flood magnitude, thereby potentially overlooking important differences in flood intensity magnitudes.

Willmott's Index of agreement (absolute version) quantifies the degree to which one dataset matches another, accounting for both bias and variance:

$$d = 1 - \frac{\sum_{i=1}^n (x_i - y_i)}{\sum_{i=1}^n (|y_i - \bar{x}| + |x_i - \bar{x}|)} \quad (7)$$

where x_i and y_i are the values from the two datasets and $\bar{x}$ is the mean of the reference dataset. In this study, we adopted a consistent convention: when 311 was part of the comparison, 311 served as the reference dataset (e.g. SAR-311, SWMM-311); for SAR-SWMM comparisons, SWMM served as the reference. This approach reflects the role of 311 as direct observations of community impact and SWMM as the physics-based simulation benchmark. Values of Willmott's Index range from 0 (no agreement) to 1 (perfect agreement). Higher index values indicate greater similarity between the spatial distribution and magnitude of flooding intensities between datasets. Because the denominator depends on the mean of the reference dataset, the index is not symmetric, and swapping reference datasets can change the resulting value. Therefore, results should be interpreted with caution and in the context of the chosen reference. Nonetheless, because the index is applied consistently across all datasets and storm events in this study, it provides a reliable basis for the relative comparison of agreement between the different data sources. Willmott's Index of Agreement delivers a comprehensive and balanced measure of dataset similarity, combining both magnitude and variability into one metric.

Collectively, employing multiple metrics mitigates the weaknesses of any single measure, enabling a robust and comprehensive characterization of dataset agreement. The combination of binary and continuous metrics ensures that both the spatial extent and intensity patterns of flooding are thoroughly evaluated, providing valuable insights into dataset performance and reliability.

3.1.3. Consideration of metric behaviors

When evaluating spatial agreement across multiple flood datasets, it is crucial to recognize that each metric may respond uniquely under scenarios of systematic bias, such as consistent overprediction or underprediction of flooding. Systematic bias occurs when one dataset repeatedly indicates flooding, regardless of actual flood presence, potentially leading to misleading interpretations if only a single metric is considered. For instance, binary metrics such as IoU and Dice coefficients primarily measure spatial overlap, which means they can yield artificially high scores in situations where one dataset significantly overestimates flood extents. Even when the actual flood conditions are limited or localized, large false-positive areas can result in substantial overlap with true positives, thereby inflating IoU and Dice values. Thus, while these metrics are excellent for capturing the direct spatial alignment of flood events, their sensitivity to systematic overprediction biases is limited.

In contrast, MCC is specifically structured to penalize biased predictions more effectively, given its explicit incorporation of true negative counts. Because MCC accounts for both correct and incorrect classifications across flooded and non-flooded areas, it typically provides a more rigorous indication of spatial agreement under biased conditions, sharply reducing scores when systematic overprediction occurs. However, MCC itself becomes limited in low-flooding scenarios, as the coefficient may become unstable or undefined due to insufficient positive detections. Therefore, MCC, while robust in moderate to extensive flooding scenarios, should be supplemented by other metrics to fully understand dataset performance.

Continuous metrics, including Pearson's correlation coefficient, Kendall's τ , and Willmott's Index of Agreement further complement the binary metrics by explicitly evaluating agreement in flood intensity distributions. Pearson's correlation coefficient, sensitive to linear magnitude relationships, tends to decline significantly under biased scenarios where flood intensities systematically deviate from the true distribution. Meanwhile, Kendall's τ , focusing on rank-order consistency, robustly reveals situations where the relative ranking of flood intensities remains consistent or diverges across datasets, offering insight into underlying structural differences in intensity distributions. Willmott's Index, by accounting for both magnitude bias and variance, is particularly effective at detecting and quantifying intensity-level discrepancies. Substantially reduced Willmott's values signal magnitude mismatches or systematic bias in flood characterization.

Consequently, no single metric alone adequately captures all aspects of spatial and intensity agreement—particularly under systematic bias conditions. Employing multiple complementary metrics is therefore critical. Integrating binary metrics (IoU and Dice for direct spatial overlaps), MCC (for balanced binary classification evaluation), and continuous metrics (Pearson's correlation, Kendall's τ , and Willmott's Index) ensures a comprehensive and nuanced characterization of agreement.

3.2. Identification of multi-source flooding overlap

For three sources of data—SAR, SWMM and 311—collected on four high-precipitation dates in three neighborhoods, we adopt a unified framework to detect places where the indicators overlap. Each dataset was first aligned to a shared spatial grid and evaluated for flooding in every pixel at each date. We then required that at least half of these events (i.e. two out of four) show simultaneous flooding to classify an overlap. In other words, pairwise overlap (e.g. SAR-SWMM) was flagged if two datasets agreed on at least two of the four dates, while triple overlap was flagged if all three sources agreed for at least two events. This procedure yielded a series of ‘agreement maps,’ highlighting pixels that consistently appear across independent detection methods, thereby identifying potential flood hotspots and underscoring where data sources diverge.

4. Results

The following subsections describe trends and differences observed across SAR, SWMM, and 311 datasets; these are not intended as validations against a definitive ground truth, but as comparative perspectives on urban flooding.

4.1. Patterns observed in SAR data

Figures 3–5 show the 250 m-gridded MRMS precipitation, SAR, SWMM, and 311 flooding data for six target dates across the three neighborhoods. The MRMS precipitation data represent total accumulated precipitation during the storm event and were linearly interpolated to match the 250 m analysis grid. The aggregated SAR data tend to depict broad flooding coverage because the original 10 m resolution—after CNN-based removal of most building backscattering noise—retains small reflectivity signals when resampled to 250 m, which may appear as flooded cells. Nevertheless, differences in intensity between events and neighborhoods are evident. SAR systematically detects a larger flood extent compared to both SWMM and 311. This pattern arises because radar-based flood mapping is sensitive to even small changes in surface water reflectivity, meaning shallow ponding or wet surfaces can be classified as flooding, whereas SWMM requires that drainage thresholds be exceeded before flooding is recorded. The tendency for SAR to relatively overclassify flooding is particularly notable in Chatham and Woodlawn, where large contiguous areas appear flooded despite limited corroboration from SWMM or 311. Additionally, the spatial distribution of MRMS precipitation does not always correspond to SAR-detected flooding. This mismatch occurs because surface water may persist or be detected in locations with shallow ponding or low infiltration rates regardless of rainfall intensity. In Woodlawn, the high impervious cover amplifies surface runoff, but SAR-based detection still appears to overestimate flood coverage relative to SWMM simulations, likely because SAR cannot differentiate between shallow wetting and actual inundation depth. This behavior, while sometimes leading to discrepancies with modeled or reported data, also means SAR can highlight areas with surface water retention that other datasets might miss.

4.2. Insights from SWMM simulations

The SWMM model often predicts no flooding in a catchment for moderate or low precipitation events—such as those on 14 April 2019 and 9 March 2019—because it requires drainage thresholds to be exceeded before flooding is registered. This behavior contrasts with SAR detections in such events, which can still identify surface wetness. In Woodlawn, SWMM generally predicts smaller flood extents than SAR, reflecting how stormwater is routed into drainage infrastructure in the model, potentially underestimating surface water accumulation in poorly drained or unmodeled locations. In Humboldt Park, SWMM identifies specific storm sewer surcharge points concentrated around drainage bottlenecks, producing localized predicted flooding patterns. While SWMM explicitly incorporates impervious surface fractions (derived from detailed datasets from the Metropolitan Water Reclamation District, MWRD) within its modified Green–Ampt infiltration calculations, discrepancies between its predictions and observed flooding in SAR or 311 suggest other important influences, such as localized infrastructure issues like blocked inlets or collapsed pipes, surface runoff conditions that exceed modeled capacity, or minor topographic depressions that are not explicitly represented in the model inputs. These differences underscore the importance of considering both modeled and observational data sources to account for unmodeled or emergent flooding mechanisms in urban systems.

4.3. Signals from 311 reports

The 311 service request data, though smoothed for spatial analysis, remain inherently point-based and reflect localized reports from residents. Rainfall accumulation patterns in figures 3–5 sometimes align with 311 request clusters, but such alignment is far from universal. Differences between rainfall patterns and 311

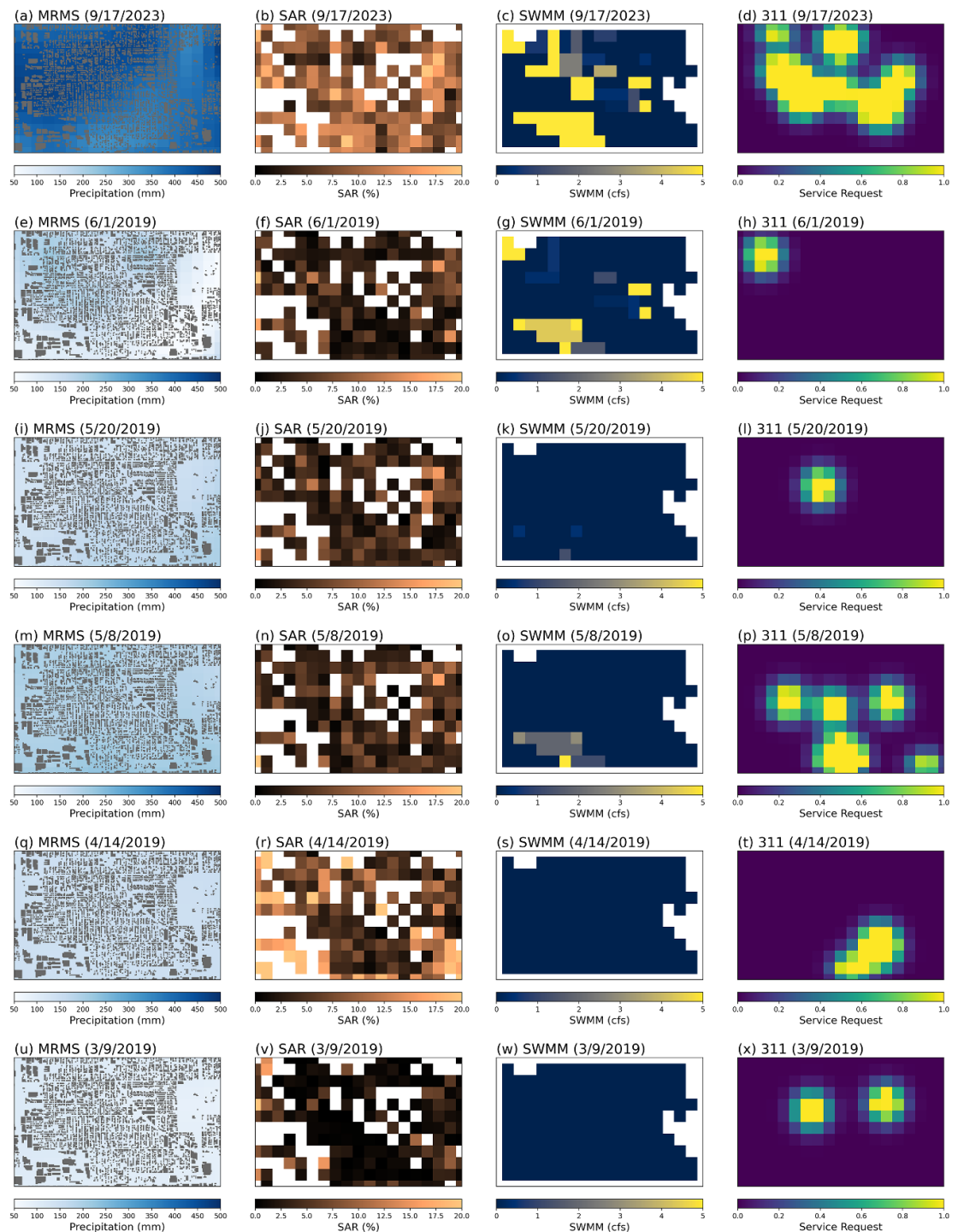
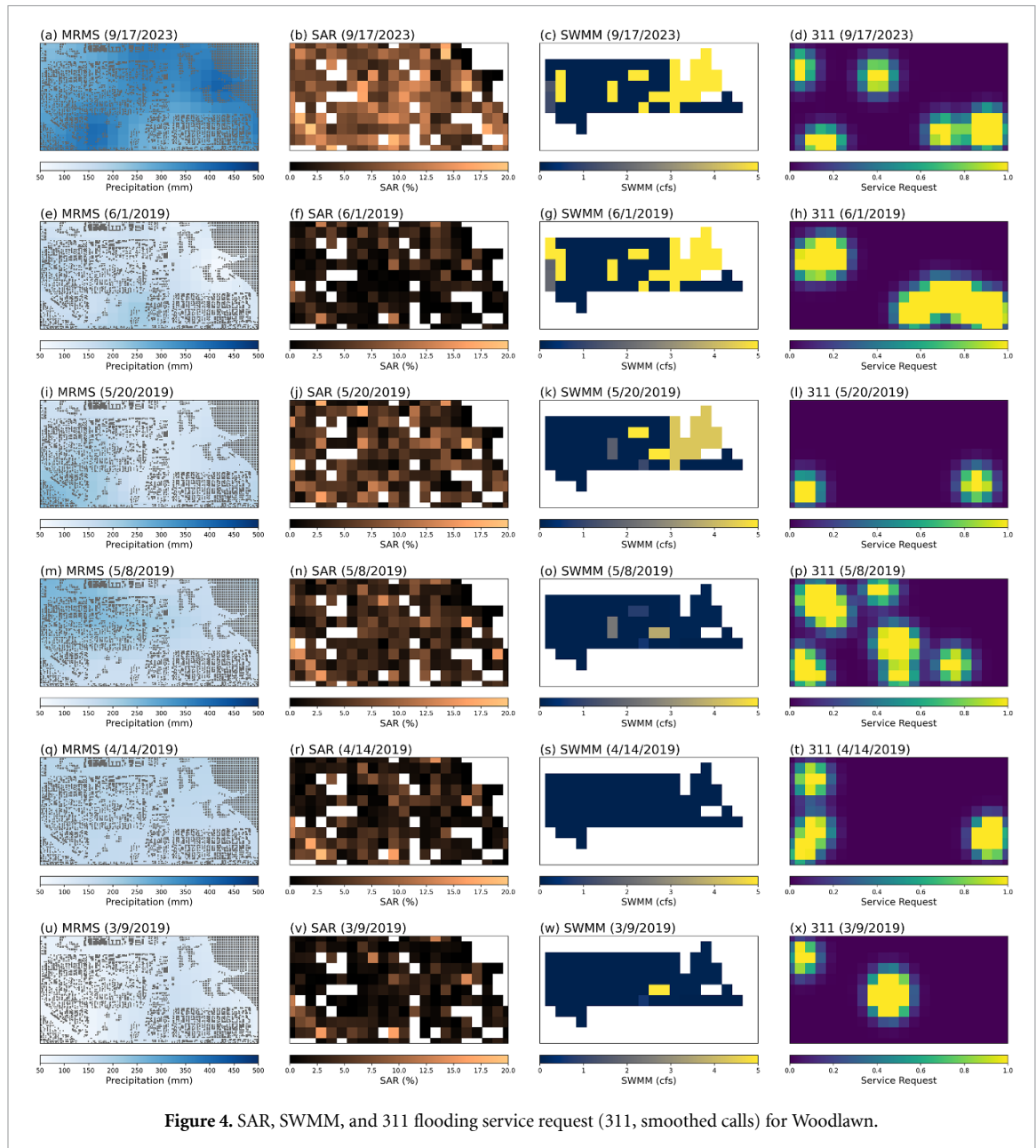


Figure 3. MRMS precipitation accumulation during the storm event (with building footprints), satellite-based flooding observation (SAR, pixel percentage), model-based flooding estimate (SWMM, cubic feet per second) and self-reported flooding service requests (311, smoothed calls) for Humboldt Park. (a)–(d) MRMS, SAR, SWMM, and 311 data for 17 September 2023. (e)–(h), (i)–(l), (m)–(o), and (p)–(t) same as (a)–(c) but for 1 June 2019, 20 May 2019, 8 May 2019, 14 April 2019, and 9 March 2019, respectively.

clusters can be driven by variations in infrastructure performance, the routing of surface runoff, and social or behavioral factors influencing reporting behavior. These factors include population density, public awareness of reporting systems, and accessibility during flood events. The 311 datasets thus capture a community-based perception of flooding impacts rather than purely physical flood occurrence. In Humboldt Park, 311 requests often show flooding beyond SWMM-predicted areas, suggesting vulnerabilities not fully represented in the modeled drainage system. In Chatham, 311 reports correspond more closely to SAR detections than to



SWMM predictions, consistent with known ponding in low-lying areas and the potential underestimation of such flooding by SWMM. While 311 data provide valuable localized information, they are subject to reporting biases and incomplete coverage, meaning they are best interpreted in conjunction with other datasets.

4.4. Within-method and cross-metric comparisons

The quantitative comparison of SAR, SWMM, and 311 flooding datasets reveals clear differences in agreement that depend strongly on event magnitude and spatial context (table 1). During major events, such as 17 September 2023 and 14 April 2019, binary overlap metrics (IoU, Dice) were consistently high for certain dataset pairs. In Humboldt Park on 17 September 2023, SAR-SWMM achieved an IoU of 0.91 and Dice of 0.96, while SWMM-311 recorded IoU and Dice values of 0.96 and 0.98, respectively. These values indicate substantial similarity in mapped flood extents when conditions produce widespread flooding. However, the corresponding continuous metrics tell a different story. For SAR-SWMM in the same event and location, Pearson's r was 0.0012 and Willmott's index was 0.0004, demonstrating that, despite large spatial overlaps, the two datasets differed markedly in how they distributed flooding intensity within the area boundary.

For smaller or more spatially fragmented events, both binary and continuous metrics tended to be low, and cases of undefined MCC were common due to the scarcity of flooded pixels. In Woodlawn on 1 June 2019, SAR-SWMM recorded an IoU of only 0.09 and a Dice of 0.17, showing minimal common flooded

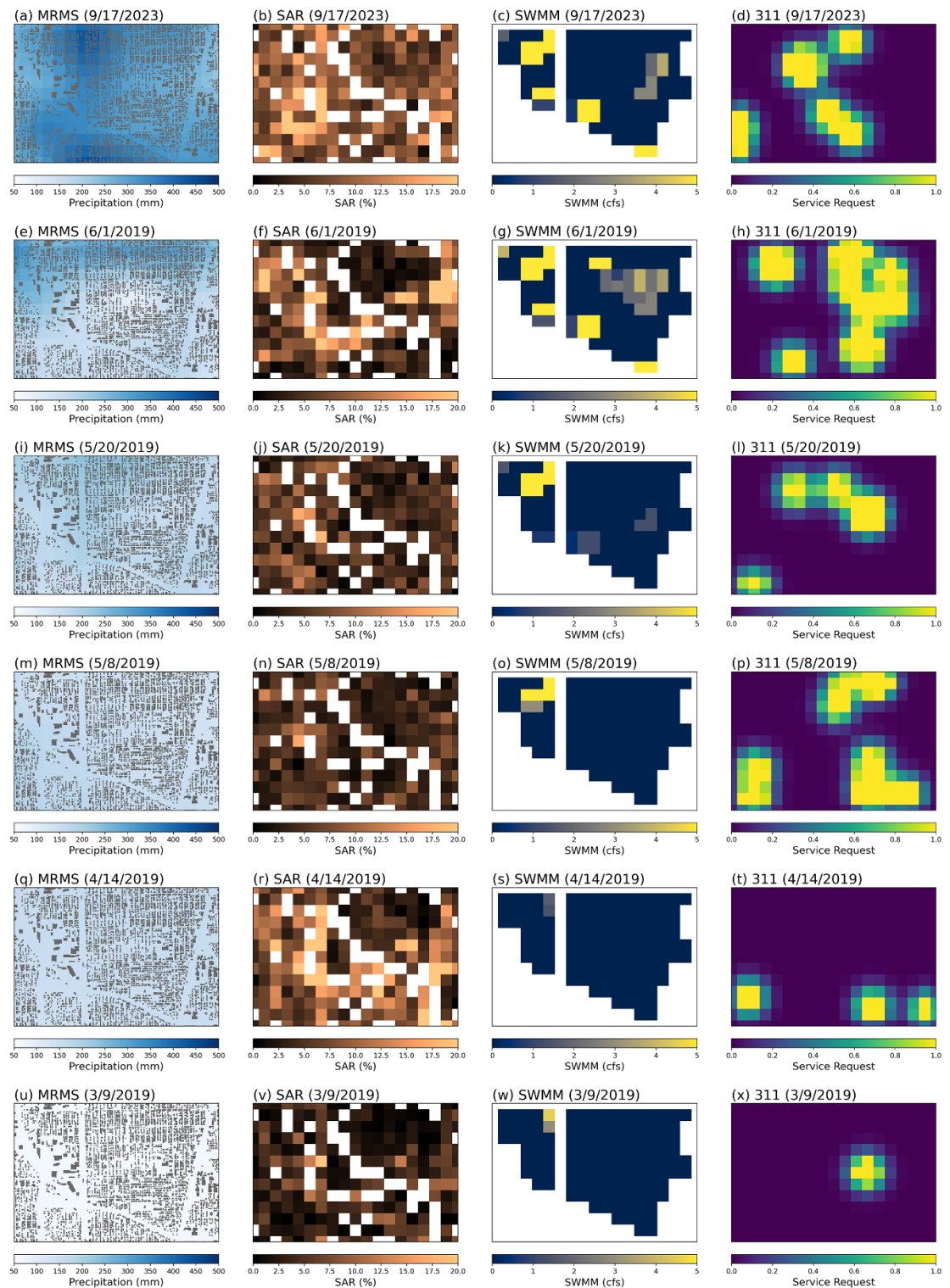


Figure 5. SAR, SWMM, and 311 flooding service request (311, smoothed calls) for Chatham.

area. Where MCC could be computed, such as for SAR-311 in Woodlawn on 17 September 2023, the value was modest (0.17), indicating weak agreement even when some overlap existed.

The continuous metrics often exposed discrepancies not apparent from binary results. In Humboldt Park on 17 September 2023, SAR-311 had an IoU of 0.83 but a negative Pearson's r (-0.1385), reflecting inverse relationships in pixel-level intensities. Willmott's index further underscored these differences: SWMM-311 typically had relatively high scores (e.g. 0.63 in Humboldt Park on 17 September 2023), suggesting similar spatial intensity patterns between modeled surcharging and resident-reported impacts, while SAR-SWMM

Table 1. Statistical comparison for each dataset pair (SAR, SWMM, 311), region (Humboldt Park, Woodlawn, Chatham), and date, using thresholds of 5% (SAR), 0 (SWMM), and 0.01 (311) for IOU calculation.

Pair	Region	Date	MCC	IoU	Dice	Pearson	Kendall	Willmott
SAR-SWMM	Humboldt Park	9 March 2019	—	0.1121	0.2016	—	—	0.5016
		14 April 2019	—	0.6379	0.7789	—	—	0.3675
		8 May 2019	—	0.3707	0.5409	0.1288	−0.0059	0.0883
		20 May 2019	—	0.3276	0.4935	0.1134	0.0776	0.2749
		1 June 2019	—	0.2155	0.3546	−0.0058	0.0569	0.0059
		17 September 2023	—	0.9138	0.955	0.0012	0.0506	0.0004
	Woodlawn	9 March 2019	—	0.0676	0.1266	0.0382	0.1162	0.0815
		14 April 2019	—	0.1757	0.2989	—	—	0.4295
		8 May 2019	—	0.2838	0.4421	0.0429	0.0011	0.0703
		20 May 2019	—	0.4324	0.6038	−0.1015	−0.0311	0.012
		1 June 2019	—	0.0946	0.1728	−0.0371	−0.1212	0.0001
		17 September 2023	—	0.8649	0.9275	0.2752	0.1789	0.001
	Chatham	9 March 2019	—	0.2581	0.4103	−0.0537	−0.0255	0.2218
		14 April 2019	—	0.5591	0.7172	−0.0252	0.0096	0.3244
		8 May 2019	—	0.2903	0.45	0.1135	0.0653	0.0789
		20 May 2019	—	0.4409	0.6119	0.0899	−0.0188	0.0052
		1 June 2019	—	0.5161	0.6809	0.0827	−0.0261	0.001
		17 September 2023	—	0.8065	0.8929	−0.0238	0.0518	0.0032
SAR-311	Humboldt Park	9 March 2019	−0.1999	0.068	0.1273	−0.148	−0.2227	0.4125
		14 April 2019	−0.1851	0.2424	0.3902	−0.183	−0.1807	0.3397
		8 May 2019	−0.0027	0.3472	0.5155	−0.046	−0.0097	0.2722
		20 May 2019	−0.1182	0.125	0.2222	−0.0181	−0.106	0.3398
		1 June 2019	0.0303	0.1406	0.2466	−0.0506	0.039	0.3556
		17 September 2023	−0.027	0.8261	0.9048	−0.1385	−0.0942	0.2138
	Woodlawn	9 March 2019	−0.1737	0.0517	0.0984	−0.0547	−0.0532	0.3134
		14 April 2019	−0.0649	0.1538	0.2667	0.0095	−0.0553	0.3777
		8 May 2019	0.1455	0.3095	0.4727	0.106	0.1085	0.3297
		20 May 2019	0.0992	0.2768	0.4336	0.0414	0.0407	0.3722
		1 June 2019	−0.0637	0.0972	0.1772	−0.1845	−0.0988	0.2119
		17 September 2023	0.1678	0.7374	0.8489	0.1068	0.0041	0.2933
	Chatham	9 March 2019	−0.0117	0.1489	0.2593	0.0416	−0.0427	0.4016
		14 April 2019	0.1936	0.4145	0.586	−0.0005	0.0972	0.3717
		8 May 2019	−0.0853	0.3034	0.4655	−0.1428	−0.1329	0.2233
		20 May 2019	−0.0027	0.3602	0.5297	−0.0691	0.0113	0.2853
		1 June 2019	0.1433	0.5269	0.6901	−0.1274	−0.0613	0.3059
		17 September 2023	0.0515	0.644	0.7834	−0.051	−0.0042	0.2754
SWMM-311	Humboldt Park	9 March 2019	—	0.6561	0.7923	—	—	0
		14 April 2019	—	0.3758	0.5463	—	—	0
		8 May 2019	—	0.9172	0.9568	0.0055	−0.0002	0.5765
		20 May 2019	—	0.3885	0.5596	−0.0369	−0.0386	0.5309
		1 June 2019	—	0.1975	0.3298	0.1247	0.0775	0.6363
		17 September 2023	—	0.9618	0.9805	−0.0668	0.0861	0.6292
	Woodlawn	9 March 2019	—	0.7386	0.8497	0.146	−0.221	0.5707
		14 April 2019	—	0.4545	0.625	—	—	0
		8 May 2019	—	0.9318	0.9647	0.0227	−0.3397	0.5175
		20 May 2019	—	0.2614	0.4144	−0.0908	−0.1697	0.5944
		1 June 2019	—	0.8295	0.9068	−0.2793	−0.1249	0.5859
		17 September 2023	—	0.875	0.9333	−0.0369	−0.2663	0.5884
	Chatham	9 March 2019	—	0.5294	0.6923	−0.046	−0.1207	0.6071
		14 April 2019	—	0.2605	0.4133	−0.0362	−0.0856	0.5923
		8 May 2019	—	0.8235	0.9032	−0.1	−0.2603	0.6214
		20 May 2019	—	0.8487	0.9182	0.0229	−0.042	0.6495
		1 June 2019	—	1	1	0.269	0.0939	0.638
		17 September 2023	—	0.7899	0.8826	−0.037	−0.0316	0.6372

values were extremely low (e.g. 0.0004 for the same case), consistent with fundamentally different intensity distributions.

Sensitivity testing of binary thresholds confirmed that high IoU or Dice values can sometimes result from small absolute overlaps when overall flood coverage is low. In such cases, visual inspection of figures 3–5 often revealed that the apparent agreement was confined to a few scattered pixels, rather than representing meaningful spatial correspondence. This reinforces the need to interpret binary metrics in tandem with continuous measures, which capture differences in magnitude and spatial distribution that binary classification alone cannot reflect.

Overall, the metric results show that high spatial overlap during major floods does not necessarily translate into agreement in flood intensity or distribution. Continuous measures repeatedly revealed mismatches—particularly for SAR–SWMM—where one dataset mapped much broader extents or detected different physical expressions of flooding. This combined metric approach therefore provides a more complete and realistic assessment of agreement, highlighting where methods converge in flood mapping and where they diverge in ways that matter for interpretation and application.

4.5. Multi-source flooding overlap result

The spatial distributions of flood-indicating overlaps among the SAR, SWMM, and 311 datasets across the neighborhoods of Humboldt Park, Woodlawn, and Chatham are illustrated in figure 6. Overlap maps were generated by identifying grid cells that exhibited simultaneous flooding detections in at least two out of four analyzed high-precipitation events (17 September 2023; 1 June 2019; 20 May 2019; and 8 May 2019). We considered three types of pairwise overlaps (SAR–SWMM, SAR–311, SWMM–311) as well as the triple overlap involving all three data sources.

Clear differences in spatial overlap patterns among the datasets were observed. SAR–SWMM overlaps appeared relatively sparse and isolated, highlighting limited agreement between SAR-derived surface



flooding detections and SWMM-predicted surcharges. This limited overlap suggests that the hydrologic model, despite explicit incorporation of impervious surface data, primarily captured flooding events driven by stormwater infrastructure constraints. Consequently, SWMM did not identify many locations of surface inundation identified by SAR, particularly those associated with shallow, widespread surface ponding or areas not directly connected to modeled drainage systems.

In contrast, SAR-311 overlaps were notably broader and more spatially extensive, indicating better alignment between areas identified by residents as flooded and regions identified by SAR imagery. However, visual inspection of flood maps frequently indicated that statistical binary metrics (e.g. IoU, Dice) suggested higher spatial agreement than was visually apparent. This occurred because binary overlap metrics can exaggerate agreement when flooding is spatially fragmented or when one dataset systematically overpredicts flood extents, a condition frequently observed with SAR. This bias is particularly evident in urban settings, where SAR tends to capture shallow moisture and wet surfaces not necessarily considered flooding by residents or infrastructure-focused models like SWMM. Thus, SAR-311 overlaps must be interpreted cautiously, with complementary continuous metrics employed to assess true flood intensity agreement.

SWMM-311 overlaps were intermediate, forming moderate-sized clusters that tended to differ spatially from SAR overlaps. SWMM typically identified flooding associated with drainage bottlenecks, while 311 calls reflected community-reported flooding, influenced not only by visible inundation but also by social and behavioral factors (White and Trump 2018, Lee *et al* 2025). Therefore, the SWMM-311 overlaps emphasized locations where community impacts coincided with modeled drainage limitations, highlighting genuine infrastructure vulnerabilities.

The limited number of locations exhibiting triple overlaps (SAR-SWMM-311) represent robustly identified chronic flood ‘hotspots.’ These areas repeatedly exhibited evidence of significant flooding impacts from all three independent methods, underscoring critical infrastructure or topographic issues necessitating targeted interventions.

To address potential biases in our overlap analysis explicitly, we conducted sensitivity tests varying the binary thresholds used for overlap detection. Results confirmed that binary metrics, while valuable for clearly identifying spatial overlaps, often provided overly optimistic spatial agreement values when flooding was sparse or fragmented. Continuous metrics consistently showed reduced agreement in these conditions, providing critical complementary information on flooding intensity differences.

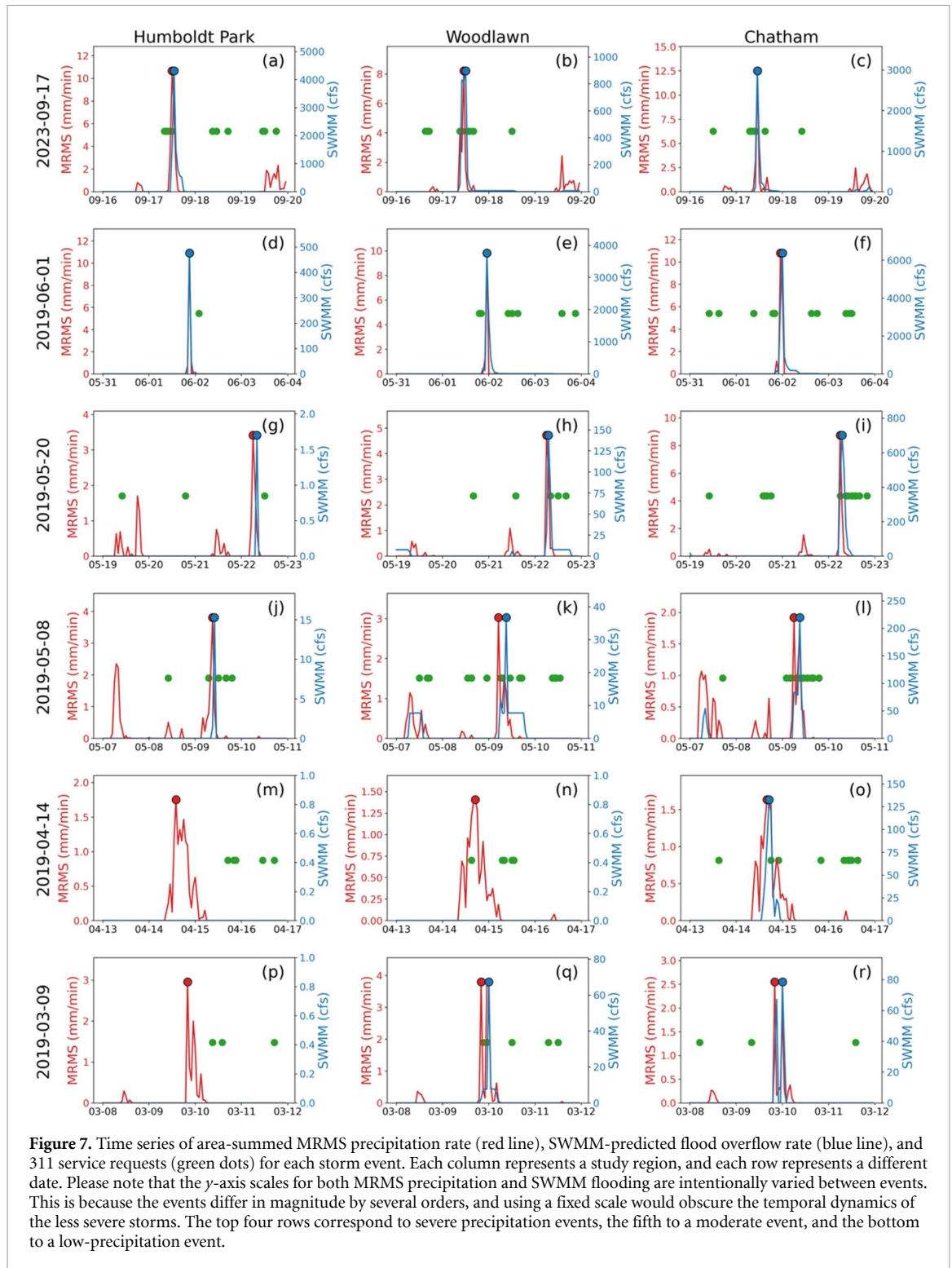
In summary, this integrated multi-source flooding overlap analysis demonstrates that binary spatial metrics alone can misrepresent true flood risk, particularly when visually inspected flood maps suggest weaker spatial correlations. By interpreting binary overlaps in conjunction with continuous similarity metrics, we achieve a more accurate, realistic, and nuanced understanding of urban flood dynamics. These insights strongly support an integrated observational and modeling framework, combining the strengths of SAR imagery, hydrologic modeling, and community-sourced reports for improved flood risk assessment and infrastructure planning.

4.6. Temporal comparison of flooding events

To investigate the temporal dimension of flooding in each dataset, we analyze the temporal progression of three datasets: area total MRMS precipitation, SWMM-estimated flooding, and 311 service requests. SAR is excluded from this analysis because it provides only a single scene with a fixed revisit interval, making it insufficient to capture the evolution of a single event. Figure 7 shows the time series for these datasets across each study region (Humboldt Park, Woodlawn, and Chatham) and each flood event date. Each column corresponds to one site, while each row represents a distinct event. The red line indicates the MRMS precipitation intensity (in mm hr^{-1}), the blue line depicts the SWMM-predicted flow rate (in cubic feet per second), and the green dots mark 311 service requests reported in the respective neighborhood.

A key observation is that intense precipitation events (large peaks in the red line) generally coincide with spikes in the SWMM flow rate (blue line), typically occurring shortly after the heaviest rainfall. This pattern is especially pronounced in the first four rows, which represent severe precipitation. The lag between rainfall peaks and SWMM-estimated flooding highlights the time required for water to be routed through the stormwater system, indicating that pipe flow dynamics and detention storage influence the flood response. By contrast, moderate or low precipitation events (shown in the fifth and last rows) often remain below the SWMM’s flood threshold, resulting in minimal SWMM-predicted flooding. Nonetheless, 311 service requests (green dots) still occur even in these lower-intensity events, illustrating that in certain regions infrastructure or drainage capacity remains below the precipitation-drive flow even during small events.

Another noteworthy pattern emerges when comparing the timing of 311 calls to precipitation peaks and SWMM estimates. Although the SWMM model typically predicts flooding within one to two hours of peak rainfall, 311 calls can persist for a longer period and, in some cases, even precede the main precipitation event. Several factors can explain this discrepancy. For instance, intense rainfall in other parts of the city can



travel through the interconnected sewer system, causing backups and basement flooding that trigger 311 calls before the storm arrives locally. Additionally, some early reports may reflect residual flooding from earlier rainfall, while other mechanisms also play a role: localized storm cells can produce heavy rainfall not fully captured in the area-averaged data, infrastructure limitations like pre-saturated soils or clogged drains can trigger flooding earlier than the storm peak, and residents may report flooding earlier than local precipitation peaks, for example due to upstream inflows, drainage delays, or residual water from prior rainfall.

Overall, the SWMM model's node-based framework generally exhibits a more binary response: once rainfall surpasses its operational threshold, predicted flooding can rise sharply, but remains negligible otherwise. In contrast, 311 reports capture a broader spectrum of on-the-ground realities. This dataset often captures smaller-scale flooding events—such as basement seepage or minor street ponding—that do not

necessarily meet the model's threshold, highlighting the need for integrating human-reported data with physical modeling to refine flood risk assessments. This divergence underscores how real-world flooding dynamics can be more complex than model predictions suggest.

5. Discussion and conclusion

5.1. Evaluation of flood delineation methods

This study evaluated three complementary approaches to urban flood monitoring—satellite remote sensing (using SAR imagery), hydrologic simulation (via the SWMM model), and crowdsourced data (from 311 service requests)—in three flood-prone Chicago neighborhoods (Humboldt Park, Woodlawn, and Chatham). The findings demonstrate the value of a multi-source strategy. Areas where different data sources agree, or overlap, represent zones of high-confidence flooding and serve as crucial anchor points for our analysis. Simultaneously, the unique information from each data source reveals different facets of the flood event, offering a more comprehensive picture than any single method could alone. However, this picture should not be as definitive proof of flood occurrence in the absence of additional independent field validation (e.g. gauges, sensors, or post-event surveys). Rather, it is a comparative synthesis that highlights areas of consensus and disagreement between different indicators, each with its own strengths, limitations, and potential biases.

- **Satellite SAR imagery:** our results confirm that SAR provides broad coverage and weather-independent detection but also highlight its tendency to identify larger flood extents than other methods. For instance, SAR consistently flagged widespread shallow inundation in Chatham and Woodlawn (figures 4 and 5), even when SWMM and 311 data showed more localized impacts. This underscores SAR's sensitivity to surface moisture but also the risk of overestimation in dense urban settings.
- **Hydrologic modeling (SWMM):** SWMM simulations effectively captured drainage bottlenecks and surcharges, as illustrated in Humboldt Park where modeled flooding aligned with 311 reports around infrastructure constraints (figure 3). However, SWMM underrepresented certain locations reported by residents or detected by SAR, particularly where localized inlet failures or small-scale ponding occurred. This supports the conclusion that while SWMM is valuable for system-wide dynamics, it may not resolve highly localized flooding.
- **Crowdsourced flooding reports (311):** the 311 data provided critical insight into community-level impacts, including reports that persisted after precipitation had subsided (figure 7). This temporal persistence captured the duration of standing water, which neither SAR (due to revisit gaps) nor SWMM (due to threshold-based surcharging) represented. However, our overlap analysis (figure 6) also revealed uneven reporting across neighborhoods, reflecting demographic and behavioral biases consistent with prior studies.

5.2. Multi-modal approaches

Despite the inherent limitations and biases in each dataset, the locations where two ('pairwise') or all three ('triple') sources converge offer compelling evidence of chronic flooding issues. When different methodologies consistently point to the same flood-prone zones, this convergence provides stronger evidence that those areas experience recurrent flooding or drainage stress. This convergence often signals underlying infrastructure vulnerabilities, such as insufficient storm-sewer capacity, inadequate drainage design, or topographic depressions that prolong surface water pooling. For instance, if SAR repeatedly detects standing water in a specific area, the SWMM model predicts sewer overflows there, and community members consistently submit 311 flood complaints for the same location, the collective evidence becomes difficult to dismiss as random or one-off occurrences. These multi-dataset overlaps not only confirm that a given spot is prone to recurrent flooding but also help localize the root cause—whether it be storm-sewer surcharging, blocked catch basins, or low-lying terrain—thereby guiding more focused remedial action. However, while agreement across SAR, SWMM, and 311 increases confidence, it should not be interpreted as definitive proof of flood occurrence in the absence of independent validation.

- **SAR-SWMM overlap:** when surface-water detection from SAR aligns with SWMM's simulated surcharges, the consensus suggests that rainfall has overwhelmed sewer capacities. While the absence of 311 calls may indicate that residents do not perceive these events as severe or persistent, this pairwise overlap still highlights possible weak points in the drainage system that may become more problematic under heavier or prolonged storms.
- **SAR-311 overlap:** agreement between radar-detected inundation and resident complaints demonstrates that water has accumulated long enough for people to notice and report it. Although the SWMM model may not confirm flooding in these locations—possibly due to localized issues like precipitation below the threshold

to trigger node-based flooding—this pairing confirms that standing water presents a visible nuisance to affected communities. At the same time, it is important to note that residents in some neighborhoods may be less likely to file 311 reports due to social or behavioral factors, meaning that the absence of reports does not imply absence of flooding.

- **SWMM-311 overlap:** When modeled sewer overflows coincide with 311 complaints, it points to drainage-system constraints that manifest in on-the-ground experiences. These scenarios may involve basement flooding, short-lived street-level surcharges, or backups that do not remain long enough to be captured by SAR. However, it is also important to note that SWMM does not account for flooding caused by clogged inlets, collapsed pipes, or under-designed infrastructure, so certain real-world incidents reported through 311 may not appear in the model output. Even so, the occurrence of both model-predicted and resident-reported flooding underscores structural vulnerabilities that merit further investigation.
- **Triple overlap (SAR, SWMM, and 311):** locations where all three sources converge provide the most conclusive evidence of chronic flood risk. In these ‘hotspots,’ the flooding mechanism is overwhelmingly clear, showing surface ponding (SAR), sewer constraints (SWMM), and community impact (311), all reinforcing one another. These areas should be prioritized for targeted interventions—such as upsizing storm-sewer infrastructure, clearing blocked catch basins, and improving grading or green infrastructure—to mitigate repeated flood occurrences and protect both public safety and property. At the same time, focusing solely on triple overlaps risks overlooking other vulnerable locations. Pairwise overlaps or consistent signals from a single dataset can still highlight critical issues—such as under-reported flooding in disadvantaged communities or localized drainage failures—that may not register across all three measures. Therefore, triple overlaps should be treated as a conservative, high-confidence subset of hotspots, but not as the sole basis for decision-making. Effective flood risk management should integrate multiple lines of evidence, engineering knowledge, and local input to avoid neglecting areas in urgent need of improvements.

5.3. Implications for flooding detection and event response

A key benefit of this integrated multi-source perspective lies in its potential to enhance early warning and emergency response. Near-real-time combination of precipitation measurements with SWMM simulations could provide indicative warnings regarding neighborhoods or drainage sub-catchments that may be at higher risk of flooding during storm events, provided that model assumptions and uncertainties are carefully considered. If these model predictions are quickly corroborated by a surge in resident-submitted 311 reports or validated with post-storm SAR imagery, city agencies can swiftly identify and prioritize emerging flood issues. Such rapid identification allows proactive measures—such as the deployment of portable pumps or preemptive road closures—and enables quicker dispatch of maintenance crews, helping to prevent minor incidents from escalating into larger hazards.

Another practical advantage of this multi-source approach is strategic infrastructure maintenance and targeted flood mitigation planning. Over time, patterns revealed by combined data sources identify locations consistently affected by flooding. For instance, persistent ponding detected by SAR at an intersection, regularly simulated sewer overflows by SWMM at the same spot, and recurrent basement backups reported by residents via 311 suggest a clear priority location for infrastructure upgrades. Leveraging this convergent evidence empowers decision-makers to strategically invest in necessary improvements—such as upsizing drainage pipes, installing green infrastructure, or conducting preventative maintenance—instead of merely reacting to severe or emergency flooding events. Thus, multi-source data can systematically inform long-term capital improvements, leading to increased urban resilience.

Furthermore, the principle of integrating remote sensing, physical models, and community-sourced reports could extend beyond flooding to other urban hazards. Each hazard—such as extreme heat events, water main breaks, sinkholes, or air quality crises—has unique indicators and modeling requirements. However, adopting a similar multifaceted monitoring strategy could produce valuable, actionable insights. For instance, satellite imagery, numerical simulations, and citizen-generated data could collectively track heatwaves or pinpoint recurring air-quality issues. The key insight is that integrating diverse data sources generates a more robust and validated understanding of localized hazards, greatly improving situational awareness and informing more effective interventions across various climate-related risks.

5.4. Summary and future research

This study demonstrates the effectiveness of a combined flood monitoring approach, integrating satellite SAR imagery, SWMM hydrologic modeling, and community-generated 311 reports within three flood-prone neighborhoods of Chicago. Each individual method provides distinct insights into flood dynamics, but their integration significantly enhances the detection, assessment, and response capabilities related to urban flooding. Locations identified by the convergence of multiple sources emerged as clear priority areas, highlighting persistent flooding risks and underlying infrastructure vulnerabilities.

Future research should aim at advancing data frequency, accuracy, and integration techniques to further improve flood prediction, detection, and response. Expanding SAR coverage through additional satellite constellations, refining SWMM simulations using higher-resolution precipitation inputs, and employing advanced machine learning approaches to analyze 311 reporting patterns will enhance urban flood-monitoring capabilities. Additionally, emerging data sources—including traffic cameras, connected vehicle sensors, and social media—could offer more dynamic, real-time insights into urban flood conditions. Finally, successful operationalization of this integrated approach will require supportive policies and institutional frameworks, ensuring that city agencies effectively utilize multi-source data to inform early warning systems, infrastructure planning, and climate-resilience strategies. By continuously refining and expanding these methodologies, cities will become better equipped to anticipate, mitigate, and respond to flooding and other climate-related hazards.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: MRMS precipitation data is available in Iowa State University repository (<https://mtarchive.geol.iastate.edu/>). Sentinel-1 data for SAR calculation can be found in NASA Earthdata Portal (www.earthdata.nasa.gov/data/platforms/space-based-platforms/sentinel-1). 311 service request data can be found in Chicago Data Portal (<https://data.cityofchicago.org/>).

Supplementary data available at <https://doi.org/10.1088/3033-4942/ae0aed/data1>.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research's Urban Integrated Field Laboratories CROCUS project research activity, under Award No. DE-SC0023226. This research is also supported by the National Science Foundation awards 2330565 and 2139316 at Discovery Partners Institute.

ORCID iDs

Jangho Lee  0000-0002-8942-1092
Sun Young Park  0009-0000-4018-1254
Abhinav Wadhwa  0000-0003-4982-3270
Aaron I Packman  0000-0003-3172-4549
Stephen W Nesbitt  0000-0003-0348-0452
Ashish Sharma  0000-0001-9332-8256
Max Berkelhammer  0000-0002-8924-716X
Marcelo H Garcia  0000-0003-0643-489X
Rao Kotamarthi  0000-0002-2612-7590
Deanna Hence  0000-0003-3299-5009
William M Miller  0000-0003-0750-6314

References

- AghaKouchak A, Chiang F, Huning L S, Love C A, Mallakpour I, Mazdiyasni O, Moftakhari H, Papalexiou S M, Ragno E and Sadegh M 2020 Climate extremes and compound hazards in a warming world *Annu. Rev. Earth Plan. Sci.* **48** 519–48
- Agonafir C, Lakhankar T, Khanbilvardi R, Krakauer N, Radell D and Devineni N 2022a A machine learning approach to evaluate the spatial variability of New York City's 311 street flooding complaints *Comput. Environ. Urban Syst.* **97** 101854
- Agonafir C, Pabon A R, Lakhankar T, Khanbilvardi R and Devineni N 2022b Understanding New York City street flooding through 311 complaints *J. Hydrol.* **605** 127300
- Almedeij J and Esen I 2014 Modified Green-Ampt infiltration model for steady rainfall *J. Hydrol. Eng.* **19** 04014011
- Amitrano D, Di Martino G, Di Simone A and Imperatore P 2024 Flood detection with SAR: a review of techniques and datasets *Remote Sens.* **16** 656
- Balaian S K, Sanders B F and Abdolhosseini Qomi M J 2024 How urban form impacts flooding *Nat. Commun.* **15** 6911
- Bauer-Marschallinger B, Cao S, Tupas M E, Roth F, Navacchi C, Melzer T, Freeman V and Wagner W 2022 Satellite-based flood mapping through Bayesian inference from a sentinel-1 SAR datacube *Remote Sens.* **14** 3673
- Bisht D S, Chatterjee C, Kalakoti S, Upadhyay P, Sahoo M and Panda A 2016 Modeling urban floods and drainage using SWMM and MIKE URBAN: a case study *Nat. Hazards* **84** 749–76
- Bulti D T and Abebe B G 2020 A review of flood modeling methods for urban pluvial flood application *Model. Earth Syst. Environ.* **6** 1293–302
- Bytheway J L, Abel M R, Mahoney K and Cifelli R 2019 A multiscale evaluation of multisensor quantitative precipitation estimates in the Russian River basin *J. Hydrometeorol.* **20** 447–66

- Cantone J and Schmidt A 2011 Improved understanding and prediction of the hydrologic response of highly urbanized catchments through development of the Illinois urban hydrologic model *Water Resour. Res.* **47** 8
- Chandran D V, Anitha J and Nair G 2024 Deep learning-based flood detection system using semantic segmentation 2024 7th Int. Conf. on Circuit Power and Computing Technologies (ICCPCT)
- Chicco D and Jurman G 2020 The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation *BMC Genomics* **21** 6
- Choi B-Y, Al-Mansoori M K, Zaman R and Albishri A A 2018 Understanding what residents ask cities: open data 311 call analysis and future directions *Proc. Workshop Program of the 19th Int. Conf. on Distributed Computing and Networking*
- Cousins J J 2017 Infrastructure and institutions: stakeholder perspectives of stormwater governance in Chicago *Cities* **66** 44–52
- Dalton F and Kenny G 1980 The Chicagoland Tunnel and Reservoir Plan (TARP) *ISRM Int. Symp.-Rockstore*
- Darabi H, Rahmati O, Naghibi S A, Mohammadi F, Ahmadisharaf E, Kalantari Z, Torabi Haghighi A, Soleimanpour S M, Tiefenbacher J P and Tien Bui D 2022 Development of a novel hybrid multi-boosting neural network model for spatial prediction of urban flood *Geocarto Int.* **37** 5716–41
- Deidda C, Rahimi L and De Michele C 2021 Causes of dependence between extreme floods *Environ. Res. Lett.* **16** 084002
- Delleur J W 2003 The evolution of urban hydrology: past, present, and future *J. Hydraul. Eng.* **129** 563–73
- Diggle P 1985 A kernel method for smoothing point process data *J. R. Stat. Soc. C* **34** 138–47
- Dixon B, Johns R and Fernandez A 2021 The role of crowdsourced data, participatory decision-making and mapping of flood related events *Appl. Geogr.* **128** 102393
- Elmore K L, Flamig Z L, Lakshmanan V, Kaney B, Farmer V, Reeves H D and Rothfusz L P 2014 MPING: crowd-sourcing weather reports for research *Bull. Am. Meteorol. Soc.* **95** 1335–42
- Elmore K L, Grams H M, Apps D and Reeves H D 2015 Verifying forecast precipitation type with mPING *Weather Forecasting* **30** 656–67
- Fan Q, Chen F, Cheng M, Lou S, Xiao R, Zhang B, Wang C and Li J 2019 Ship detection using a fully convolutional network with compact polarimetric SAR images *Remote Sens.* **11** 2171
- Feldman A B 2021 How a dual drainage approach changes the examination of flood management infrastructure: a case study of the Chatham neighborhood of Chicago, IL (University of Illinois at Urbana-Champaign)
- Fernandez A, Black J, Jones M, Wilson L, Salvador-Carulla L, Astell-Burt T, Black D and Ebi K L 2015 Flooding and mental health: a systematic mapping review *PLoS One* **10** e0119929
- Filipponi F 2019 Sentinel-1 GRD preprocessing workflow *Proceeding*
- Geudtner D, Torres R, Snoeij P, Davidson M and Rommen B 2014 Sentinel-1 system capabilities and applications 2014 IEEE Geoscience and Remote Sensing Symp.
- Habibi H, Awal R, Fares A and Temimi M 2021 Performance of multi-radar multi-sensor (MRMS) product in monitoring precipitation under extreme events in Harris County, Texas *J. Hydrol.* **598** 126385
- Hagen L, Seon Yi H, Pietri S and Keller E T 2019 Processes, potential benefits, and limitations of big data analytics: a case analysis of 311 data from city of Miami *Proc. 20th Annual Int. Conf. on Digital Government Research*
- Hamers E M, Maier H R, Zecchin A C and van Delden H 2023 Effectiveness of nature-based solutions for mitigating the impact of pluvial flooding in urban areas at the regional scale *Water* **15** 642
- Hammond M J, Chen A S, Djordjević S, Butler D and Mark O 2015 Urban flood impact assessment: a state-of-the-art review *Urban Water J.* **12** 14–29
- Helmrich A M et al 2021 Opportunities for crowdsourcing in urban flood monitoring *Environ. Modelling Softw.* **143** 105124
- Huang H, Chen X, Zhu Z, Xie Y, Liu L, Wang X, Wang X and Liu K 2018 The changing pattern of urban flooding in Guangzhou, China *Sci. Total Environ.* **622** 394–401
- Huong H T L and Pathirana A 2013 Urbanization and climate change impacts on future urban flooding in Can Tho city, Vietnam *Hydrol. Earth Syst. Sci.* **17** 379–94
- IFRC 2018 World disasters report 2018. Leaving no one behind report (The International Federation of Red Cross)
- Islam M T and Meng Q 2022 An exploratory study of Sentinel-1 SAR for rapid urban flood mapping on Google Earth engine *Int. J. Appl. Earth Obs. Geoinf.* **113** 103002
- Jiang L, Chen Y and Wang H 2015 Urban flood simulation based on the SWMM model *Proc. Int. Assoc. Hydrol. Sci.* **368** pp 186–91
- Kassen M 2013 A promising phenomenon of open data: a case study of the Chicago open data project *Gov. Inf. Q.* **30** 508–13
- Keenan M B, Shankar P and Haas P 2019 Assessing disparities of urban flood risk for households of color in Chicago *Illinois Munic. Policy J.* **4** 1–18 (available at: <https://cnt.org/sites/default/files/publications/IMPJ-Assessing-Disparities-of-Urban-Flood-Risk-for-Households-of-Color-in-Chicago.pdf>)
- Kontokosta C E and Hong B 2021 Bias in smart city governance: how socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions *Sustain. Cities Soc.* **64** 102503
- Lane K, Charles-Guzman K, Wheeler K, Abid Z, Graber N and Matte T 2013 Health effects of coastal storms and flooding in urban areas: a review and vulnerability assessment *J. Environ. Public Health* **2013** 913064
- Lee J-S, Jurkevich L, Dewaele P, Wambacq P and Oosterlinck A 1994 Speckle filtering of synthetic aperture radar images: a review *Remote Sens. Rev.* **8** 313–40
- Lee J, Berkelhammer M, Park S Y and Wilson M D 2025 Analysis of urban flooding in Chicago based on crowdsourced data: drivers and the need for community-based mitigation strategies *Environ. Res.* **5** 025008
- Lee J, Berkelhammer M, Wilson M D, Love N and Cintron R 2024 Urban land surface temperature downscaling in Chicago: addressing ethnic inequality and gentrification *Remote Sens.* **16** 1639
- Luo H, Collis S M, Crisologo I A, Horton D E, Packman A and Garcia M H 2022a Prediction of urban flooding risks using high-resolution modeling and hybrid rainfall data *Proc. IAHR World Congress*
- Luo H, Oberg N, Landry B J and García M H 2021 Assessing the system performance of an evolving and integrated urban drainage system to control combined sewer overflows using a multiple-layer based coupled modeling approach *J. Hydrol.* **603** 127130
- Luo P, Luo M, Li F, Qi X, Huo A, Wang Z, He B, Takara K, Nover D and Wang Y 2022b Urban flood numerical simulation: research, methods and future perspectives *Environ. Modelling Softw.* **156** 105478
- Macharyulu I S, Satapathy D P, Sahoo A, Samantaray S, Mohanta N R and Ray A 2022 Performance evaluation of MLP and CNN models for flood prediction *Intelligent System Design: Proc. INDIA 2022* (Springer) pp 273–81
- Mao L, Li Y, Hao W, Zhou X, Xu C and Lei T 2016 A new method to estimate soil water infiltration based on a modified Green-Ampt model *Soil Tillage Res.* **161** 31–37
- Markus M, Wuebbles D J, Liang X-Z, Hayhoe K and Kristovich D A 2012 Diagnostic analysis of future climate scenarios applied to urban flooding in the Chicago metropolitan area *Clim. Change* **111** 879–902

- Marsalek J, Barnwell T, Geiger W, Grottker M, Huber W, Saul A, Schilling W and Torno H C 1993 Urban drainage systems: design and operation *Water Sci. Technol.* **27** 31–70
- Mason D C, Speck R, Devereux B, Schumann G J-P, Neal J C and Bates P D 2009 Flood detection in urban areas using TerraSAR-X *IEEE Trans. Geosci. Remote Sens.* **48** 882–94
- Mazzoleni M, Verlaan M, Alfonso L, Monego M, Norbiato D, Ferri M and Solomatine D P 2017 Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? *Hydrol. Earth Syst. Sci.* **21** 839–61
- McGuire M P 2018 Cities as hydro-geologic terrain: design research to transform urban surfaces *Plan J.* **3** 165–90 (available at: <https://repository.library.noaa.gov/view/noaa/42532>)
- Mignot E, Li X and Dewals B 2019 Experimental modelling of urban flooding: a review *J. Hydrol.* **568** 334–42
- Ming X, Liang Q, Dawson R, Xia X and Hou J 2022 A quantitative multi-hazard risk assessment framework for compound flooding considering hazard inter-dependencies and interactions *J. Hydrol.* **607** 127477
- Moazami S and Najafi M 2021 A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada *J. Hydrol.* **594** 125929
- Mohamed M S, Bagwell M, Alqallaf S S and Reddy K R 2025 Bioswales and permeable pavements as green infrastructure for flood mitigation in Chicago neighborhoods *J. Hazard. Toxic Radioact. Waste* **29** 04025004
- Mullissa A, Vollrath A, Odongo-Braun C, Slagter B, Balling J, Gou Y, Gorelick N and Reiche J 2021 Sentinel-1 sar backscatter analysis ready data preparation in google earth engine *Remote Sens.* **13** 1954
- Nanía L S, Leon A S and García M H 2015 Hydrologic-hydraulic model for simulating dual drainage and flooding in urban areas: application to a catchment in the metropolitan area of Chicago *J. Hydrol. Eng.* **20** 04014071
- National Academies of Sciences 2019 *Framing the Challenge of Urban Flooding in the United States* (National Academies Press)
- Negri R, Tsai Y-J, Tan B Y and Ceferino L 2023 Investigating the use of citizen-science data as a proxy for flood risk assessment in New York City
- Nkwunonwo U, Whitworth M and Baily B 2020 A review of the current status of flood modelling for urban flood risk management in the developing countries *Sci. Afr.* **7** e00269
- Noether G E 1981 Why kendall tau *Teach. Stat.* **3** 41–43
- Notti D, Giordan D, Caló F, Pepe A, Zucca F and Galve J P 2018 Potential and limitations of open satellite data for flood mapping *Remote Sens.* **10** 1673
- Nychka D W 2000 Spatial-process estimates as smoothers *Smoothing and Regression: Approaches, Computation, and Application* vol 329 p 393 (available at: https://web.archive.org/web/20120109185310id_/http://www.image.ucar.edu/~nychka/manuscripts/SplinesKriging.pdf)
- Pachaly R L, Vasconcelos J G, Allasia D G, Tassi R and Bocchi J P P 2020 Comparing SWMM 5.1 calculation alternatives to represent unsteady stormwater sewer flows *J. Hydraul. Eng.* **146** 04020046
- Piadeh F, Behzadian K and Alani A M 2022 A critical review of real-time modelling of flood forecasting in urban drainage systems *J. Hydrol.* **607** 127476
- Qin Y 2020 Urban flooding mitigation techniques: a systematic review and future studies *Water* **12** 3579
- Rahmati O et al 2020 Development of novel hybridized models for urban flood susceptibility mapping *Sci. Rep.* **10** 12937
- Rai P K, Chahar B and Dhanya C 2017 GIS-based SWMM model for simulating the catchment response to flood events *Hydrol. Res.* **48** 384–94
- Rainey J L, Pandian K, Sterns L, Atoba K, Mobley W, Highfield W, Blessing R and Brody S D 2021 Using 311-call data to measure flood risk and impacts: the case of Harris County TX (Institute for a Disaster Resilient Texas) p 22
- Rezatoufighi H, Tsoi N, Gwak J, Sadeghian A, Reid I and Savarese S 2019 Generalized intersection over union: a metric and a loss for bounding box regression *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*
- Rivera-Giboyeaux A and Weinbeck S 2024 Evaluation of WSR-88D level III and MRMS rainfall estimates against rain gauge observations at the Savannah river site (<https://doi.org/10.15191/nwajom.2024.1207>)
- Rossman L A and Supply W 2006 Storm water management model, quality assurance report: dynamic wave flow routing (US Environmental Protection Agency, Office of Research and Development)
- Rydén J 2022 Statistical analysis of possible trends for extreme floods in northern Sweden *River Res. Appl.* **38** 1041–50
- Safavi F, Chowdhury T and Rahnemoonfar M 2021 Comparative study between real-time and non-real-time segmentation models on flooding events. 2021 *IEEE Int. Conf. on Big Data (Big Data)*
- Scalise C and Fitzpatrick K 2012 *Chicago deep tunnel design and construction Structures Congress 2012*
- Sedgwick P 2012 Pearson's correlation coefficient *Bmj* **345** e4483
- Shamir R R, Duchin Y, Kim J, Sapiro G and Harel N 2019 Continuous dice coefficient: a method for evaluating probabilistic segmentations (arXiv: 1906.11031)
- Sharma N K and Saharia M 2025 DeepSARFlood: rapid and automated sar-based flood inundation mapping using vision transformer-based deep ensembles with uncertainty estimates *Sci. Remote Sens.* **11** 100203 (available at: <https://www.sciencedirect.com/science/article/pii/S2666017225000094>)
- Shen X, Wang D, Mao K, Anagnostou E and Hong Y 2019 Inundation extent mapping by synthetic aperture radar: a review *Remote Sens.* **11** 879
- Singh P, Sinha V S P, Vijhiani A and Pahuja N 2018 Vulnerability assessment of urban road network from urban flood *Int. J. Disaster Risk Reduct.* **28** 237–50
- Sörensen J and Mobini S 2017 Pluvial, urban flood mechanisms and characteristics—assessment based on insurance claims *J. Hydrol.* **555** 51–67
- Tanim A H, McRae C B, Tavakol-Davani H and Goharian E 2022 Flood detection in urban areas using satellite imagery and machine learning *Water* **14** 1140
- Tingsanchali T 2012 Urban flood disaster management *Proc. Eng.* **32** 25–37
- Tsyganskaya V, Martinis S, Marzahn P and Ludwig R 2018 SAR-based detection of flooded vegetation—a review of characteristics and approaches *Int. J. Remote Sens.* **39** 2255–93
- Uddin K, Matin M A and Meyer F J 2019 Operational flood mapping using multi-temporal Sentinel-1 SAR images: a case study from Bangladesh *Remote Sens.* **11** 1581
- Wang L, Qian C, Kats P, Kontokosta C and Sobolevsky S 2017 Structure of 311 service requests as a signature of urban location *PLoS One* **12** e0186314
- Wang Y et al 2022 Spatial characteristics and driving factors of urban flooding in Chinese megacities *J. Hydrol.* **613** 128464
- White A and Trump K-S 2018 The promises and pitfalls of 311 data *Urban Aff. Rev.* **54** 794–823

- Willmott C J, Robeson S M and Matsuura K 2012 A refined index of model performance *Int. J. Climatol.* **32** 2088–94
- Wing O E J, Pinter N, Bates P D and Kousky C 2020 New insights into US flood vulnerability revealed from flood insurance big data *Nat. Commun.* **11** 1444
- Wu H, Adler R F, Hong Y, Tian Y and Policelli F 2012 Evaluation of global flood detection using satellite-based rainfall and a hydrologic model *J. Hydrometeorol.* **13** 1268–84
- Zhang C, Xu T, Wang T and Zhao Y 2023 Spatial-temporal evolution of influencing mechanism of urban flooding in the Guangdong Hong Kong Macao greater bay area, China *Front. Earth Sci.* **10** 1113997
- Zhang J et al 2016 Multi-radar multi-sensor (MRMS) quantitative precipitation estimation: initial operating capabilities *Bull. Am. Meteorol. Soc.* **97** 621–38
- Zhang W and Villarini G 2020 Deadly compound heat stress-flooding hazard across the Central United States *Geophys. Res. Lett.* **47** e2020GL089185
- Zhong S, Yang L, Toloo S, Wang Z, Tong S, Sun X, Crompton D, FitzGerald G and Huang C 2018 The long-term physical and psychological health impacts of flooding: a systematic mapping *Sci. Total Environ.* **626** 165–94
- Zhu J, Dai Q, Deng Y, Zhang A, Zhang Y and Zhang S 2018 Indirect damage of urban flooding: investigation of flood-induced traffic congestion using dynamic modeling *Water* **10** 622
- Zhu Z, Oberg N, Morales V M, Quijano J C, Landry B J and Garcia M H 2016 Integrated urban hydrologic and hydraulic modelling in Chicago, Illinois *Environ. Modelling Softw.* **77** 63–70