Assessing the Trustworthiness of Crowdsourced Rainfall Networks: A Reputation System Approach

Alexander B. Chen1, Madhur Behl1,2, and Jonathan L. Goodall1

1Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA. 2Department of Computer Science, University of Virginia, Charlottesville, VA, USA

Abstract High resolution and accurate rainfall information is essential to modeling and predicting hydrological processes. Crowdsourced personal weather stations (PWSs) have become increasingly popular in recent years and can provide dense spatial and temporal resolution in rainfall estimates. However, their usefulness could be limited due to less trust in crowdsourced data compared to traditional data sources. Using crowdsourced PWSs data without a robust evaluation of its trustworthiness can result in inaccurate rainfall estimates as PWSs are installed and maintained by non-experts. In this study, we advance the Reputation System for Crowdsourced Rainfall Networks (RSCRN) to bridge this trust gap by assigning dynamic trust scores to PWSs. Based on rainfall data collected from 18 PWSs in two dense clusters in Houston, Texas, USA as a case study, we found that using RSCRN-derived trust scores can increase the accuracy of 15-min PWS rainfall estimates when compared to rainfall observations recorded at the city's high-fidelity rainfall stations. Overall, RSCRN rainfall estimates improved for 77% (48 out of 62) of the analyzed storm events, with a median root-mean-square error (RMSE) improvement of 27.3%. Compared to an existing PWS quality control method, results showed that RSCRN improved rainfall estimates for 71% of the storm events (44 out of 62), with a median RMSE improvement of 18.7%. Using RSCRN-derived trust scores can make the rapidly growing network of PWSs a more useful resource for hydrologic applications, greatly improving knowledge of rainfall patterns in areas with dense PWSs.

1. Introduction

Flooding is becoming commonplace in cities and communities worldwide, causing severe damage and loss of property (Salman & Li, 2018; Wilby & Keenan, 2012). As a result of climate change, rainfall extremes are expected to become more intense and highly heterogeneous (Ohba & Sugimoto, 2019; Sharma et al., 2018). Floods triggered by these increased storms often exhibit large variability both in space and time, especially in urban areas with a large portion of impervious surface (Cristiano et al., 2017; Quinn et al., 2019). Although recent advances in computational power and modeling approaches have made it possible to accurately model flooding at increasingly high resolution (Mosavi et al., 2018; Sakserna et al., 2019; Savage et al., 2016; Shen et al., 2019; Zahura et al., 2020), these models require measured rainfall observations as input at high spatial and temporal resolutions. However, the current resolution of observations through traditional rainfall networks is typically insufficient, or even unavailable, for certain flood-prone regions (Cristiano et al., 2017; Sadler et al., 2018; Zhu et al., 2018).

Traditionally, in situ rainfall observations are obtained from gauges managed by federal or municipal agencies. These rain gauges, which we refer to as high-fidelity rainfall stations in this study, provide accurate measurements as they are installed and maintained by experts, but they are limited in coverage (Overeem et al., 2013; Villarini et al., 2008). An alternative to rain gauges is the use of weather radars. However, radar rainfall is derived indirectly from radar reflectivity observed at certain heights in the atmosphere, which may not accurately represent rainfall at the ground level (Smith et al., 1996) and requires calibration with ground gauges (Krajewski & Smith, 2002). Recent improvements in dual-polarization technologies in weather radar address some of these limitations. However, further improvement of rainfall estimation using dual-polarization radars is needed. For example, range-dependent sampling errors and uncertainties in identifying hydrometeor types with radar measurements may introduce larger bias in rainfall estimates, which cannot be easily corrected with ground gauges (Cunha et al., 2015).
Crowdsourcing could offer a potential solution to the need for high resolution and accurate rainfall estimates. Crowdsourcing is a practice whereby data are obtained through open calls to the general public for data collection, resulting in increased coverage, but introducing the challenges associated with data collection by non-experts (Estellés-Arolas & González-Ladrón-De-Guevara, 2012). Personal Weather Stations (PWSs) are user-friendly and affordable off-the-shelf weather stations installed and maintained by individuals that offer a means for crowdsourcing weather data including rainfall observations (Gharesifard & Wehn, 2016). PWSs data can be easily shared through services such as Weather Underground, which enables real-time data gathering, integration, and visualization of weather data collected across a world-wide network of PWSs via online platforms and mobile applications.

The growing popularity of PWSs in recent years has made crowdsourcing a powerful opportunity to supplement existing rainfall networks (Muller et al., 2015; de Vos et al., 2017; Yang & Ng, 2017; Weeser et al., 2019; Lowery & Fienen, 2013). This crowdsourced data is growing rapidly, making it an increasingly valuable resource for hydrologists (Zhang et al., 2018). Based on our review of the Weather Underground data archive, the number of PWSs in the US increased exponentially from 7,000 to 100,000 from 2010 to 2019 (Figure 1). In Houston, Texas, for example, the number of PWSs has grown from 99 to 382 over the three-year period 2016 to 2019 (Figure 1), which equates to an increased density of from 0.06 to 0.24 PWSs per square kilometer. If such exponential growth continues, the density of PWSs in populated areas in the US could reach one PWS per square kilometer in five years, which exceeds recommended spatial resolutions for rainfall observations required for urban hydrology (Berne et al., 2004; Fletcher et al., 2013).

The increased popularity of PWSs can be attributed to the availability of the crowdsourced networks that allow anyone to act as a data contributor. Such openness, however, introduces challenges in assuring accurate data (Bell et al., 2013; Chapman et al., 2017; Muller et al., 2015; Meier et al., 2017). Crowdsourced networks are typically lightly controlled networks with few constraints and only basic quality-control processes. As a result, people have higher levels of confidence in data collected from high-fidelity rainfall stations, as there are fewer sources of error in their observations compared to crowdsourced data (Cox, 2011; Hunter et al., 2013). Like high-fidelity rainfall stations, PWSs can experience device errors, but they can also suffer from compromised setups, lack of routine maintenance, and other sources of error that are less common in high-fidelity stations (de Vos et al., 2017; Meier et al., 2017; Muller et al., 2015). For example, improper installation of PWSs, such as siting the station under a tree canopy or next to a building, can lead to consistently inaccurate readings. Likewise, the owner of the PWS might not routinely maintain and calibrate the device, which could lead to sensor drift and faulty observations. Beyond these cases, it is also possible in open crowdsourced networks in which people might deliberately manipulate data to produce misleading evidence (Huang et al., 2014; Sanchez et al., 2018). Therefore, a method to evaluate the trustworthiness of crowdsourced PWSs is needed before this rich and growing dataset can be used with confidence in hydrologic studies and decision making.

One way to address the problems with crowdsourced PWS data would be to adopt quality control and quality assurance (QA/QC) methods to detect, flag, or remove doubtful and erroneous data based on certain rules and thresholds (Bäserud et al., 2020; Blenkinsop et al., 2017; de Vos et al., 2019; Estévez et al., 2011; Fiebrich et al., 2010; Niu et al., 2021). If other data from more trusted sources is available, then another method would be to evaluate the quality of crowdsourced rainfall data by direct comparison with them (de Vos et al., 2017; Muller et al., 2015). Existing methods, however, may not adequately address the needs of crowdsourced weather data, and specifically, rainfall observations. QA/QC methods designed for high-fidelity stations tend to focus on outlier detection that presumes a certain source of error (sensor malfunction) and may be less able to detect other sources of error (poor sensor siting or installation). For example, the Weather Underground designates a PWS as a “Gold Star Weather Station” if it passes basic quality control criteria such as data validity, and a sensor failure checks over the previous five days (The Weather Channel, 2018). Direct comparison with trusted data sources presumes that such data is available, but PWSs have reached a density of observation in space and time that cannot be matched with other, more trusted, measurement methods. There is a need and opportunity, therefore, for innovative methods for assessing the data generated by PWSs at scale so that trustworthy stations can be used.

Figure 1. The number of Personal Weather Stations in the Weather Underground network in the US and Houston, Texas, has been growing exponentially in the past 20 years.
more confidently in decision making and, just as importantly, untrustworthy stations can be reported to owners with suggestions for improving data quality so that the overall observation network reaches its full potential.

In this study, we explore the use of reputations systems as an approach for measuring the trustworthiness of rainfall observations from PWSs. Reputation systems are commonly used to build trust between participants and foster good behavior in online crowdsourced systems (Jøsang et al., 2007). For example, online markets such as eBay and Amazon use reputation systems to enhance users’ buying and selling experiences. Such systems aggregate sellers’ past behavior and represent it as a trustworthiness rating for buyers to rely on (Resnick et al., 2000). Reputation systems have also been used for citizen science and crowdsourced data. Yang et al. (2013) designed a reputation system framework for enhancing the data reliability of citizen science environmental acoustic data. Silvertown et al. (2015) used a reputation system to motivate and reward participants of a crowdsourced species-identification website that improved the accuracy of species determinations. Huang et al. (2014) proposed a reputation system framework using the Gompertz function to compute device reputation scores based on the trustworthiness of the contributed data in participatory sensing applications. However, the use of reputation systems for crowdsourced PWS networks has not been widely adopted.

In our previous work, we presented an initial version of a system called the Reputation System for Crowdsourced Rainfall Network (RSCRN) (Chen et al., 2018) to assign trust scores to PWSs. In this paper, we significantly enhance RSCRN and evaluate the method for storm events using Houston, Texas, as a case study. The research questions guiding this work are (a) How can we systematically evaluate the trustworthiness of crowdsourced PWSs? and (b) To what extent could a reputation system approach improve rainfall estimates derived from PWSs?

The remainder of the study is organized as follows. Section 2 describes details of the RSCRN algorithm and methods to evaluate the RSCRN for storm events. Section 3 provides a description of the study area and data used in the study, as well as a storm events selection process. The results and discussion of this study are presented in Sections 4 and 5, followed by conclusions in Section 6.

2. Material and Methods

2.1. Data Preparation

The RSCRN method begins with a crowdsourced rainfall network in a specific region having $N$ PWSs. Given an analysis period of interest (say $X$ time steps), the rainfall observations from these $N$ PWSs can be collected into a matrix $P$

$$P = \begin{pmatrix}
p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\
p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\
\vdots & \vdots & \ddots & \vdots \\
p_{X,1} & p_{X,2} & \cdots & p_{X,N}
\end{pmatrix}$$

where $p_{i,j}$ is a rainfall observation measured at time step $i$ and PWS $j$. This matrix $P$, with $X$ rows for the rainfall observations and $N$ columns of PWSs, will be used as the input for RSCRN.

2.2. RSCRN Algorithm

The RSCRN algorithm consists of three steps: Cluster, Consensus, and Score (Algorithm 1). The objective of RSCRN is to evaluate the trustworthiness of PWSs based on their consensus with rainfall measured at neighboring stations. The Cluster step is to find clusters of neighboring PWSs. Next, the Consensus step is used to identify PWSs with rainfall observations that deviate from a cluster’s consensus. Finally, the Score step uses the degree of deviation from consensus to assign a new trust score to each PWS on a given time step that represents the trustworthiness of that PWS. Further detail for each step in the algorithm follows (Algorithm 1).
Typically, distances of 4 kilometers or less should be selected (Sadler et al., 2017) based on the available PWSs within a region, and any local knowledge of rainfall autocorrection distances. A cluster for the analysis. For other applications of this method, the buffer distance may differ and should be selected depending on the geographic proximity of stations. First, we compute the number of PWSs within a buffered distance of a given PWS, \( p_j \). Second, PWSs that have (a) at least four of PWSs actively reporting rainfall observations during the analysis period of interest, and (b) at least one high-fidelity rainfall station within 2 km that can be used as a cluster for the analysis. For other applications of this method, the buffer distance may differ and should be selected based on the available PWSs within a region, and any local knowledge of rainfall autocorrection distances. Typically, distances of 4 kilometers or less should be selected (Sadler et al., 2017). Starting from the input matrix \( P \), the resulting clustered matrices are denoted by \( D_k^{M_k} \), where \( k \) is the \( k \)th cluster, and \( M_k \) is the number of PWSs in the \( k \)th cluster. These matrices will be the input for the consensus step.

### 2.2.2. Consensus

The input to the consensus step is clustered sub-datasets \( D_k^{M_k} \), where each sub-dataset contains rainfall observations from PWSs that fall within the same cluster. Assuming the \( k \)th clustered sub-dataset has \( m \) PWSs, this sub-dataset will be a matrix \( D_k^{M_k}(i, j) \):

\[
D_k^{M_k}(i, j) = \begin{pmatrix}
  p_{1,1} & p_{1,2} & \cdots & p_{1,m} \\
  p_{2,1} & p_{2,2} & \cdots & p_{2,m} \\
  \vdots & \vdots & \ddots & \vdots \\
  p_{X,1} & p_{X,2} & \cdots & p_{X,m}
\end{pmatrix}
\]
where \( p_{ij} \) represents the rainfall observation of PWS \( j \) within the cluster \( k \) measured at time \( i \). For the clustered sub-datasets \( D^k_m \) \((k = 1, 2, \ldots, K)\), the consensus step computes a cooperative metric (denoted as \( C_{ij} \), which has the same dimension as \( D^k_n \)) based on the rainfall observations for each time-step \( (i = 1, 2, \ldots, X) \) and each PWS \( (j = 1, 2, \ldots, m) \).

We used the robust averaging algorithm (Chou et al., 2013) to estimate a cluster’s consensus. We selected this method for its effectiveness and efficiency in similar applications for wireless sensor networks and participatory sensing (Ganeriwal et al., 2008; Huang et al., 2014). Robust averaging is a type of weighted average method that is less affected by values that deviate from the average. For each time step \( i \), this iterative algorithm works as follows:

1. First, assign an initial (uniform) weight to every PWS \( j \) at iteration \( l = 1 \)
   \[
   w_{ij}^{(1)} = \frac{1}{m}
   \]  
   where \( m \) is the number of PWSs in the clustered sub-dataset \( D^k_m \).
2. Next, compute the robust average \( RA^i_j \), such that
   \[
   RA^i_j = \sum_{j=1}^{m} w_{ij}^l \cdot p_{ij}
   \]  
   where \( p_{ij} \) is the rainfall observation of PWS \( j \) for time step \( i \).
3. Next, compute the squared difference of PWS \( j \)'s rainfall observation \( p_{ij} \) from the robust average \( RA^i_j \)
4. Finally, compute the new robust weight at iteration \( l + 1 \)
   \[
   w_{ij}^{l+1} = \left( \frac{1}{\sum_{j=1}^{m} v_{ij}^{l} + \epsilon} \right) \left( \sum_{j=1}^{m} \frac{1}{\sum_{i=1}^{n} v_{ij}^{l} + \epsilon} \right)
   \]
   \[
   v_{ij}^l = (p_{ij} - RA^i_j)^2
   \]  
   The algorithm continues iterating until the convergence \(|w_{ij}^{l} - w_{ij}^{l+1}| < \nu \) is achieved, i.e., the robust weights converge to a value with difference less than \( \nu \). Note that the \( \epsilon \) is a small positive constant that is set to 0.1, determined by trial and error based on the convergence of the algorithm (Chou et al., 2013).

The cooperative metric is then defined as
\[
C_{ij} = \frac{w_{ij} - \overline{W}_i}{\sigma(W)}
\]
where \( \overline{W}_i \) and \( \sigma(W) \) are the average and standard deviation, respectively, of the \( i \)th row of the robust weight matrix. This metric represents the level of deviation of the final robust weight from the initial weight. A positive cooperative metric indicates agreement with the consensus (robust average) within the cluster, while a negative cooperative metric represents disagreement with the consensus. The resulting cooperative metric is used as the input for score step.

We further extended the algorithm to accommodate two data exception cases: (a) all zero observations and (b) missing observations in certain time steps. In a clustered sub-dataset, the first case occurs when all rainfall observations are zero on a given time step. In this case, cooperative metrics will be invalid because the standard deviation of the robust weight matrix is zero. Therefore, the cooperative metric of every PWS in this case is set to zero. The second case occurs when PWSs have intermittent missing observations. In this case, for those time steps where a PWS has missing observations, this particular PWS is excluded from the robust average calculation, and the cooperative metric of this PWS will be set to zero. Additionally, if there are too many PWSs with missing observations resulting in a low number of active PWSs reporting data on a time step, the cooperative metrics...
of all PWSs on that time step will also be set to zero, because the consensus computed from the robust average algorithm may be unreliable. The definition of this low number can be determined based on the data availability during the analysis period.

2.2.3. Score

As described in Section 2.2.2, the cooperative metric can be interpreted as a measure of the PWS deviation from the robust average for each time step. To evaluate the trustworthiness of the PWSs, this step assumes neutral initial trust scores for every PWS without knowledge of any past behaviors (rainfall observations, in this case), and integrates this cooperative metric to update the trust score for every PWS for each time step.

We used the beta reputation system (Josang & Ismail, 2002) for its advantages of simplicity, flexibility, and ability to counter most arbitrary device faults in wireless sensor networks (Ganeriwal et al., 2008). The beta reputation system uses a statistical approach to provide a mathematical basis for trust management. The idea is that the trust score, which is computed based on the beta probability density function (PDF), is gradually updated as new observations are made available. The beta PDF is a continuous family of distribution indexed by two parameters: $\alpha$ and $\beta$. It is denoted by $\text{beta}(\alpha, \beta)$ and can be expressed using the gamma function $\Gamma$ as

$$\text{beta}(p\vert \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1 - p)^{\beta-1}$$ (6)

where $0 < p <= 1$, $\alpha$ and $\beta > 0$. The expectation value of the beta distribution is given by

$$E(p) = \frac{\alpha}{\alpha + \beta}$$ (7)

where $0 < E(p) < 1$.

In each clustered sub-dataset, the prior distribution is assumed to be a uniform beta PDF with $\alpha_1 = 1$, $\beta_1 = 1$, and $E(p)_i = 0.5$ for every PWS at time step $i = 1$ before any data are collected. This can be interpreted as the neutral trust for these PWSs, which indicates that the relative frequency of reporting trustworthy or untrustworthy observations is equal. After observing new data, the posterior distribution will be the beta PDF with updated $\alpha$ and $\beta$ parameters. In RSCRN, these parameters are updated using the cooperative metrics $C_{ij}$ computed from the consensus step as

$$\alpha_{i+1,j} = \alpha_{i,j} \times \lambda + C_{i,j}, \quad \beta_{i+1,j} = \beta_{i,j} \times \lambda \quad \text{if} \quad C_{i,j} > 0$$

$$\alpha_{i+1,j} = \alpha_{i,j} \times \lambda, \quad \beta_{i+1,j} = \beta_{i,j} \times \lambda + |C_{i,j}| \quad \text{if} \quad C_{i,j} < -1$$

$$\alpha_{i+1,j} = \alpha_{i,j}, \quad \beta_{i+1,j} = \beta_{i,j} \quad \text{if} \quad C_{i,j} = 0 \quad \text{or} \quad -1 < C_{i,j} < 0.$$ (8)

There are four possible outcomes for updating the $\alpha$ and $\beta$ parameters: (a) A positive outcome is defined if the cooperative metric is greater than zero. In a positive outcome, the alpha parameter increases by the value of the cooperative metric. (b) A negative outcome is defined if the cooperative metric is less than $-1$, which implies significant deviation from the consensus because the final robust weight is more than one standard deviation below the average weight. In a negative outcome, the beta parameter increases by the absolute value of the cooperative metric. (c) A zero cooperative metric outcome indicates either all observations at the time step were zero, or there is a missing observation from a single PWS. In this case, both alpha and beta parameters are held constant with the previous time step values. (d) In a minor negative outcome—which we define as when the cooperative metric is less than zero but greater than $-1$—both alpha and beta parameters will also be held constant with the previous time step values because the deviation from the consensus is insignificant. In addition, to focus the evaluation on time steps when rain is reported, the algorithm is set to update trust scores only on time steps when at least one PWS in the cluster is reporting more than one tick (0.25 mm) of rainfall.

A forgetting factor $\lambda$ is introduced in Equation 8 to avoid the trust score being overly weighted on past information. The $\lambda$ parameter, which ranges from 0 to 1, is used to give old information less weight than more recent information. A forgetting factor of 1.0 indicates no forgetting at all, whereas a forgetting factor of 0 indicates forgetting all past information except for the previous time step.
Given the updated alpha and beta parameters by the cooperative metrics, the expected value of the posterior beta PDF becomes

$$E(p_{i+1,j}) = \frac{\alpha_{i+1,j}}{\alpha_{i+1,j} + \beta_{i+1,j}}.$$  \hfill (9)

Finally, the trust score $T_{i,j}$ is computed by re-scaling the expectation value to be between 0 and 10 for each PWS $j$ at time step $i$

$$T_{i,j} = 10 \cdot E(p_{i,j}).$$  \hfill (10)

### 2.3. Comparison With a PWS Quality Control Method

The performance of the RSCRN approach is evaluated against a quality control method recently proposed for PWSs (de Vos et al., 2019). This quality control approach (hereinafter referred as PWS QC method) consists of four major filters to flag PWS rainfall observations. These filters are (a) a high influx (HI) filter to capture PWSs with observations much higher than neighboring stations, (b) a faulty zero (FZ) filter to identify erroneous zeros, (c) a station outlier (SO) filter to flag PWSs with low correlation of rainfall time series with neighboring stations, and (d) a bias correction (BC) filter to compensate for a possible bias in observations compared to the rainfall observations from neighboring stations. Using the same clustered sub-datasets as input for the PWS QC method, individual PWS observations were flagged with HI flags, FZ flags and SO flags, and these flags were used to compare to RSCRN trust scores. The BC filter was later applied in the validation step as described in the next section.

### 2.4. Validation Using High-Fidelity Rainfall Stations on Storm Events

Using a binary trust score threshold, PWSs were classified as trustworthy or untrustworthy PWSs for storm events with durations of X time steps that begin on time step $T_1$ and end on time step $T_2$. The trust score thresholds of the each PWS are defined as follows:

**Trustworthy PWS:**

$$\sum_{i=T_1}^{T_2} T_{i,j} > \gamma$$  \hfill (11)

**Untrustworthy PWS:**

$$\sum_{i=T_1}^{T_2} T_{i,j} < \gamma$$  \hfill (12)

where $\gamma$ is the threshold value that ranges from 0 to 10.

Using $\gamma = 5.0$ as an example, trustworthy PWSs are stations that received average trust scores higher than 5.0 during this storm event. Generally, these PWSs are more likely to report trustworthy data during this storm event because they have been consistently contributing observations that agreed with the consensus from neighboring PWSs before the storm event. On the other hand, untrustworthy PWSs are PWSs that received average trust scores lower than 5.0, which indicates these PWSs have been reporting observations that disagreed with the consensus from neighboring PWSs. These PWSs are, therefore, less likely to report trustworthy data during the storm event.

To validate whether using a trust score threshold method can improve rainfall estimates from the crowdsourced rainfall network, we compute the root-mean-square error (RMSE) of the PWS rainfall observation with the nearest high-fidelity rainfall station for the storm event as

$$RMSE = \sqrt{\frac{1}{X} \sum_{i=T_1}^{T_2} (c_i - h_i)^2}$$  \hfill (13)
where \( c_i \) is the rainfall time series of the PWS, \( h_i \) is the rainfall time series of the high-fidelity rainfall station, and \( X \) is the duration of the storm event. Consider a cluster with \( M \) PWSs: the average RMSE of all PWSs in the cluster (denoted as \( R_{\text{all}} \)) becomes

\[
R_{\text{all}} = \frac{\sum_{j=1}^{M} \text{RMSE}_j}{M}
\]

(14)

where \( \text{RMSE}_j \) is the RMSE of the \( j \)th PWS in the cluster. This \( R_{\text{all}} \) is used to benchmark the improvement made from the RSCRN and PWS QC methods.

Assuming that the RSCRN trust score threshold revealed that in these \( M \) PWSs there are \( U \) trustworthy PWSs that received trust scores above the threshold, the average RMSE of trustworthy PWSs in the cluster (denoted as \( R_{\text{RSCRN}} \)) can be computed as

\[
R_{\text{RSCRN}} = \frac{\sum_{j=1}^{U} \text{RMSE}_j}{U}
\]

(15)

The RMSE of trustworthy PWSs is further compared with the RMSE of QC PWS. In the comparison, the rainfall time series \( c_i \) used in Equation 13 will be replaced with a bias corrected time series \( c'_i \) resulting from the BC filter of the PWS method (de Vos et al., 2019). The RMSE (denoted as \( \text{RMSE}' \)) used in the PWS QC method will then become

\[
\text{RMSE}' = \sqrt{\frac{1}{X} \sum_{i=T_1}^{T_2} (c'_i - h_i)^2}
\]

(16)

In order to compare the RSCRN and the PWS QC method, rather than using the flags resulting from the PWS QC method for individual rainfall measurements, the flags were translated into a binary decision of either at least one flag or no flags. Assuming there are \( V \) unflagged PWSs (stations without any flags filtered by the PWS QC method) during a storm event, the average RMSE of QC rainfall estimates (denoted as \( R_{\text{QC}} \)) can be be computed as

\[
R_{\text{QC}} = \frac{\sum_{j=1}^{V} \text{RMSE}'_j}{V}
\]

(17)

Lower RMSE values indicate agreement with the high-fidelity rainfall station observations. The comparison of \( R_{\text{all}} \), \( R_{\text{RSCRN}} \), and \( R_{\text{QC}} \) can then be used to determine the improvements of rainfall estimates made by each method in providing accurate rainfall estimation from a network of PWSs.

### 3. Case Study

#### 3.1. Study Area

To demonstrate and evaluate RSCRN, we focus on PWSs in Houston, Texas, as a case study. The City of Houston is in a sub-tropical climate with average annual rainfall of 1,250 mm. Flooding has been a recurring issue in Houston because of urbanization and the increase in frequency and intensity of severe storms (Zhang et al., 2018). The growing adoption of PWSs in Houston in recent years significantly increases ground gauge rainfall networks coverage. Extracting trustworthy rainfall information from the PWSs could potentially supply denser point rainfall time series and, thus, improve the knowledge of rainfall patterns to better model and control flooding.

#### 3.2. Data

##### 3.2.1. High-Fidelity Rainfall Network

In Houston, the Harris County Flood Control District (HCFCD) manages a rainfall monitoring network of 174 rainfall stations, with temporal resolutions up to 5 min, that can be used as the ground truth of the rainfall observation to evaluate RSCRN (Figure 2). Additional information about the network as well as QA/QC reports can be accessed from the HCFCD website (https://www.harriscountyfws.org/).
Figure 2. Two clusters of crowdsourced PWSs in Houston, Texas, were selected as case studies for evaluating the RSCRN.
3.2.2. Crowdsourced Rainfall Network

The crowdsourced rainfall network used in this study consists of PWSs available through the Weather Underground. We accessed the data through the API provided by the Weather Underground. The PWS observation sampling interval varies from station to station. Most of the sampling intervals are about 5–10 min per observation. Based on the available PWSs in the Houston area queried from the Weather Underground API, there were 99 PWSs in January 2016 and 382 PWSs just a few years later in April 2019.

3.3. PWS Cluster

In the cluster step, the buffer distance was set to 2 km so that the identified clusters satisfy the criteria described in Section 2.2.1. Two clusters with the most active PWSs available at the beginning of the analysis period were used to evaluate RSCRN (see Figure 2). The first cluster consists of eight PWSs and is located in southwestern Houston, Texas. The second cluster consists of 10 PWSs and is located northwest of downtown Houston. The inter-distance of PWSs in both clusters is less than 3 km (Figure 3). Table 1 shows the available metadata from the Weather Underground API. Each PWS has different start times, which is when the station joined the Weather Underground and started reporting data to its database. Among these PWSs are three major PWS brands: Ambient Weather, AcuRite, and Davis Instrument.

The 15-min rainfall time series from the 18 PWSs in the two clustered sub-datasets for the analysis period from January 1, 2017 to March 28, 2019 were input to the RSCRN. In this study, the minimum number of valid rainfall observations on a time step for computing the cooperative metrics was set to 5, given that there are 5–7 PWSs actively reporting rainfall within a cluster for the majority of the time steps during the analysis period. The forgetting factor $\lambda$ was set to 0.95, which retains approximately 20% of the prior knowledge that is more than 25 time steps (6 hr) old. This is to ensure that the trust score computed by RSCRN will not be overweighted by past observations so that it is able to accommodate temporary behavioral changes, especially during storm events. The sensitivity to this forgetting factor is further explored later in the paper. For the PWS QC method, the neighboring stations for each PWSs were set to all other PWSs in the cluster identified by the RSCRN. Several parameter choices were evaluated, and the best one was chosen based on the data availability and rainfall characteristics of the collected PWS data.

3.4. Storm Events Selection

In this study, a storm event is defined as accumulated rainfall greater than 25.4 mm within a 12-hr rolling window. Rainfall time series from the high fidelity rainfall network (HCFCFCD rainfall stations 445 and 560) were used to identify storm events for cluster 1 and cluster 2, respectively. As shown in Tables 2 and 3, 33 and 29 storm events with various rainfall statistics that occurred in what we refer to as winter (November to March) and summer (April to October) seasons during the analysis period (January 1, 2017 to March 28, 2019) were identified. In these storm events, duration ranged from 0.5 to 10.8 hr, maximum rainfall intensity from 20.3 to 113.8 mm/hr, and total rainfall from 25.4 to 164.6 mm.
4. Results

4.1. Reputation System for Crowdsourced Rainfall Networks

4.1.1. PWS Trust Score Assignment

Figure 4 shows the resulting trust scores based on the RSCRN for an example storm event. In this example, the majority of rainfall observations of the target PWS (KTXHOUST327) matched well with the robust average computed from neighboring stations in its cluster. At the beginning of the storm (marked with circle 1 in Figure 4), the trust scores remained unchanged because the RSCRN updates trust scores only when at least one PWS in the cluster reports rainfall greater than 0.25 mm. Similarly, in the middle and end of the storm (also marked with circle 1), the trust scores remain constant because all reporting rainfall is lower than 0.25 mm. The PWSs began to observe heavier rainfall starting at 02:45. As can be seen in Figure 4, the rainfall time series during the following time interval (marked with circle 2) agreed with the robust average. Therefore, the PWS received several positive outcomes (marked with the black dot on the Alpha axis), and the trust score steadily increased. There were two time steps when the PWS received negative outcomes (marked with circle 3 and black dots on the Beta axis) because its observations disagreed with the consensus, and the trust score decreased accordingly. Note that the change of trust score with regard to the same positive or negative outcomes was higher when using a smaller forgetting factor, because less prior knowledge was retained, and thus the trust score change was more sensitive. As expected, the PWS QC method did not identify any flags for the PWS during this storm event. Thus, both approaches agree this is a trustworthy PWS.

Table 1
The Metadata of PWSs Used in This Study

| ID            | Cluster  | ID  | Elevation | Latitude | Longitude | Start time      | Station type                          |
|---------------|----------|-----|-----------|----------|-----------|-----------------|---------------------------------------|
| KTXHOUST281   | Cluster 1| 24  | 29.65     | −95.46   |            | September 14, 2012 | N/A                                   |
| KTXHOUST617   | Cluster 1| 21  | 29.66     | −95.47   |            | June 7, 2015     | Ambient Weather WS-1400-IP (Wireless) |
| KTXHOUST1971  | Cluster 1| 16  | 29.66     | −95.49   |            | May 7, 2017      | AcuRite Pro Weather Center            |
| KTXHOUST2323  | Cluster 1| 21  | 29.65     | −95.48   |            | March 18, 2018   | AcuRite Pro Weather Center            |
| KTXHOUST327   | Cluster 1| 24  | 29.65     | −95.49   |            | October 27, 2013 | Davis Vantage Pro2 (Cabled)           |
| KTXHOUST1903  | Cluster 1| 18  | 29.66     | −95.48   |            | December 29, 2016| Ambient Weather WS-900-IP (Wireless)  |
| KTXHOUST355   | Cluster 1| 21  | 29.65     | −95.50   |            | June 29, 2014    | N/A                                   |
| KTXHOUST777   | Cluster 1| 21  | 29.66     | −95.46   |            | March 24, 2016   | Ambient Weather WS-1001-WiFi (Wireless)|
| KTXHOUST240   | Cluster 2| 15  | 29.79     | −95.38   |            | October 15, 2010 | Davis Vantage Pro2 Plus (Wireless)    |
| KTXHOUST805   | Cluster 2| 20  | 29.80     | −95.38   |            | May 6, 2016      | AcuRite Pro Weather Center            |
| KTXHOUST443   | Cluster 2| 21  | 29.79     | −95.37   |            | December 26, 2014| AcuRite Pro Weather Center            |
| KTXHOUST2591  | Cluster 2| 21  | 29.79     | −95.37   |            | January 1, 2019  | Ambient Weather WS-2902               |
| KTXHOUST314   | Cluster 2| 28  | 29.78     | −95.37   |            | May 2, 2013      | Davis Vantage Pro2 Plus (Wireless)    |
| KTXHOUST686   | Cluster 2| 25  | 29.78     | −95.39   |            | November 20, 2015| Ambient Weather WS-1001-WiFi (Wireless)|
| KTXHOUST452   | Cluster 2| 24  | 29.80     | −95.40   |            | January 19, 2015 | Ambient Weather WS-1200-IP (Wireless) |
| KTXHOUST2533  | Cluster 2| 26  | 29.80     | −95.38   |            | October 29, 2018 | AcuRite 5-in-1 Weather Station with AcuRite Access |
| KTXHOUST275   | Cluster 2| 20  | 29.79     | −95.40   |            | June 23, 2012    | Davis Vantage Pro 2                   |
| KTXHOUST2258  | Cluster 2| 21  | 29.80     | −95.38   |            | January 27, 2018 | Ambient Weather WS-1001-WiFi (Wireless)|
Figure 5 shows examples of when trust scores decrease for the majority of the time steps during a storm event for two untrustworthy PWSs. In the first example, the rainfall time series of the target PWS (KTXHOUST452) frequently deviated from the robust average. Although this PWS captured some of the peak values of the storm, there were several time steps between those peaks where rainfall observations significantly deviated from the consensus of the neighboring PWSs. For example, the consensus of rainfall observations among the neighboring stations was showing that it had been raining heavily between the time interval 03:00 to 06:00. However, this PWS was either reporting zero rainfall or underreporting rainfall, which resulted in receiving many negative outcomes (black dots on the Beta axis). Therefore, the trust score decreased and remained low for the entire storm. Using the PWS QC method, several time steps were identified with the FZ flag, which agreed with the RSCRN

| No | Storm event date | Season | Duration (hr) | Max. rainfall Intensity (mm/hr) | Total rainfall (mm) | Active PWSs | Median PWS total rainfall (mm) |
|----|------------------|--------|---------------|---------------------------------|---------------------|-------------|-------------------------------|
| 1  | January 2, 2017  | Winter | 0.5           | 81.3                            | 30.5                | 6           | 29.3                          |
| 2  | January 18, 2017 | Winter | 4.8           | 89.4                            | 119.9               | 6           | 100.6                         |
| 3  | January 20, 2017 | Winter | 2.0           | 77.2                            | 35.6                | 6           | 33.8                          |
| 4  | March 5, 2017    | Winter | 8.8           | 24.4                            | 61.0                | 6           | 48.5                          |
| 5  | March 29, 2017   | Winter | 2.5           | 93.5                            | 49.8                | 6           | 40.4                          |
| 6  | April 18, 2017   | Summer | 4.0           | 16.3                            | 25.4                | 6           | 14.7                          |
| 7  | May 22, 2017     | Summer | 3.3           | 52.8                            | 32.5                | 7           | 30.5                          |
| 8  | May 29, 2017     | Summer | 1.8           | 105.7                           | 62.0                | 7           | 56.1                          |
| 9  | June 4, 2017     | Summer | 2.5           | 56.9                            | 53.8                | 7           | 44.5                          |
| 10 | June 24, 2017    | Summer | 1.8           | 56.9                            | 31.5                | 6           | 20.8                          |
| 11 | July 15, 2017    | Summer | 3.3           | 113.8                           | 47.8                | 6           | 28.7                          |
| 12 | August 2, 2017   | Summer | 3.8           | 36.6                            | 42.7                | 7           | 37.3                          |
| 13 | August 8, 2017   | Summer | 2.3           | 48.8                            | 26.4                | 6           | 24.8                          |
| 14 | August 25, 2017  | Summer | 10.8          | 61.0                            | 101.6               | 7           | 67.1                          |
| 15 | September 18, 2017 | Summer | 2.5          | 52.8                            | 53.8                | 7           | 44.7                          |
| 16 | December 3, 2017 | Winter | 1.3           | 69.1                            | 39.6                | 7           | 42.9                          |
| 17 | December 16, 2017 | Winter | 1.8          | 40.6                            | 25.4                | 7           | 23.4                          |
| 18 | February 10, 2018 | Winter | 6.8          | 32.5                            | 87.4                | 7           | 86.9                          |
| 19 | March 29, 2018   | Winter | 4.3           | 52.8                            | 66.0                | 8           | 55.0                          |
| 20 | April 21, 2018   | Summer | 1.5           | 56.9                            | 41.7                | 8           | 44.2                          |
| 21 | May 21, 2018     | Summer | 2.0           | 69.1                            | 35.6                | 7           | 20.8                          |
| 22 | July 4, 2018     | Summer | 6.5           | 89.4                            | 164.6               | 8           | 144.3                         |
| 23 | July 31, 2018    | Summer | 1.0           | 81.3                            | 37.6                | 7           | 35.6                          |
| 24 | September 9, 2018 | Summer | 1.5          | 61.0                            | 29.5                | 6           | 35.8                          |
| 25 | October 15, 2018 | Summer | 0.8           | 89.4                            | 27.4                | 6           | 20.1                          |
| 26 | October 31, 2018 | Summer | 3.3           | 101.6                           | 82.3                | 5           | 94.2                          |
| 27 | December 7, 2018 | Winter | 8.5           | 52.8                            | 124.0               | 7           | 118.1                         |
| 28 | December 13, 2018 | Winter | 2.0          | 28.4                            | 26.4                | 7           | 30.0                          |
| 29 | December 27, 2018 | Winter | 3.3          | 20.3                            | 36.6                | 6           | 39.2                          |
| 30 | January 2, 2019  | Winter | 2.8           | 21.3                            | 28.4                | 7           | 35.6                          |
| 31 | January 19, 2019 | Winter | 0.8           | 65.0                            | 25.4                | 7           | 22.9                          |
| 32 | January 23, 2019 | Winter | 4.8           | 20.3                            | 26.4                | 4           | 26.2                          |
| 33 | February 26, 2019 | Winter | 3.3          | 28.4                            | 27.4                | 6           | 28.6                          |
that this station is likely to be untrustworthy. In the second example, the rainfall observation from the target PWS (KTXHOUST1903) was underreporting (00:00 to 00:30) and reporting zero value rainfall while the neighboring stations showed strong consensus of a certain rainfall magnitude. This station also overreported rainfall at 03:30 while the consensus of the neighboring stations showed that the storm had stopped. Using the PWS QC method, this PWS was first flagged with FZ flags for several intervals, followed by an HI flag where this PWS was reporting 49.27 mm while other neighboring stations all reported zero. Thus, both approaches agree these are untrustworthy PWSs.

4.1.2. PWS Trustworthiness Assessment

The RSCRN trust score evolution over all analyzed storm events is shown in Figure 6. Note that the trust scores of each PWS were computed for every time step during the analysis period (1/1/2017 00:00 to 03/28/2019 23:45), and the trust scores for the analyzed storm events were extracted to assess the trustworthiness of a PWS during a particular storm event. In this figure, each dot represents the average trust score for a storm event. The trust score evolution shows that some PWSs were assigned high trust scores throughout the analyzed storm events.

| No. | Storm event date | Season | Duration (hr) | Max. rainfall intensity (mm/hr) | Total rainfall (mm) | Active PWSs | Median PWS total rainfall (mm) |
|-----|------------------|--------|---------------|-------------------------------|---------------------|-------------|-------------------------------|
| 1   | January 2, 2017  | Winter | 1.0           | 56.9                          | 25.4                | 5           | 21.3                          |
| 2   | January 18, 2017 | Winter | 6.0           | 81.3                          | 157.5               | 6           | 165.4                         |
| 3   | January 20, 2017 | Winter | 2.8           | 73.2                          | 42.7                | 6           | 39.2                          |
| 4   | January 20, 2017 | Winter | 6.3           | 12.2                          | 37.6                | 5           | 40.4                          |
| 5   | March 5, 2017    | Winter | 7.5           | 16.3                          | 38.6                | 5           | 38.6                          |
| 6   | March 29, 2017   | Winter | 2.0           | 101.6                         | 47.8                | 5           | 46.5                          |
| 7   | April 11, 2017   | Summer | 4.3           | 48.8                          | 36.6                | 6           | 35.6                          |
| 8   | May 22, 2017     | Summer | 2.5           | 69.1                          | 37.6                | 7           | 39.1                          |
| 9   | June 4, 2017     | Summer | 3.0           | 105.7                         | 61.0                | 7           | 65.3                          |
| 10  | June 25, 2017    | Summer | 2.3           | 81.3                          | 57.9                | 7           | 25.4                          |
| 11  | July 13, 2017    | Summer | 1.0           | 81.3                          | 30.5                | 7           | 27.2                          |
| 12  | August 7, 2017   | Summer | 1.8           | 48.8                          | 27.4                | 7           | 22.4                          |
| 13  | August 26, 2017  | Summer | 18.8          | 113.8                         | 389.1              | 6           | 417.7                         |
| 14  | September 18, 2017 | Summer | 1.8           | 48.8                          | 25.4                | 6           | 30.5                          |
| 15  | January 8, 2018  | Winter | 2.5           | 40.6                          | 27.4                | 7           | 16.5                          |
| 16  | February 10, 2018 | Winter | 5.5           | 20.3                          | 57.9                | 8           | 59.6                          |
| 17  | March 29, 2018   | Winter | 5.5           | 52.8                          | 67.1                | 8           | 62.4                          |
| 18  | April 21, 2018   | Summer | 1.8           | 65.0                          | 45.7                | 8           | 46.4                          |
| 19  | May 20, 2018     | Summer | 2.0           | 44.7                          | 26.4                | 8           | 26.7                          |
| 20  | July 4, 2018     | Summer | 5.8           | 40.6                          | 110.7               | 6           | 148.6                         |
| 21  | September 9, 2018 | Summer | 1.3           | 81.3                          | 45.7                | 7           | 37.9                          |
| 22  | September 22, 2018 | Summer | 2.0           | 56.9                          | 25.4                | 7           | 16.5                          |
| 23  | October 15, 2018 | Summer | 1.3           | 61.0                          | 27.4                | 7           | 31.2                          |
| 24  | October 31, 2018 | Summer | 3.0           | 44.7                          | 34.5                | 8           | 27.8                          |
| 25  | December 7, 2018 | Winter | 9.0           | 44.7                          | 116.8               | 8           | 109.0                         |
| 26  | December 13, 2018 | Winter | 2.0           | 36.6                          | 27.4                | 7           | 27.9                          |
| 27  | December 27, 2018 | Winter | 3.5           | 36.6                          | 43.7                | 7           | 39.1                          |
| 28  | January 2, 2019  | Winter | 2.3           | 28.4                          | 30.5                | 8           | 26.4                          |
| 29  | January 23, 2019 | Winter | 5.0           | 16.3                          | 26.4                | 8           | 25.7                          |
(e.g., KTXHOUST327 in cluster 1 and KTXHOUST805 in cluster 2), while other PWSs consistently received low trust scores (e.g., KTXHOUST617 in cluster 1 and KTXHOUST452 in cluster 2). However, some PWSs had trust scores that fluctuated over time, which indicates that perhaps these stations had state changes over the analysis period.

Table 4 shows the overall assessment of PWS trustworthiness for the analyzed storm events. Based on the results for cluster 1, KTXHOUST617 was the least trustworthy PWS. If we assume 4.0 as the trust score threshold $\gamma$, of the 23 active storm events for which this PWS reported valid rainfall observations, 18 (78%) were classified as untrustworthy. If we use a more restrictive trust score threshold $\gamma = 5.0$, 20 (87%) storm events were classified as untrustworthy. KTXHOUST281 was the second least trustworthy PWS in this cluster, as its trust score fluctuated between 4.0 and 6.0, and eventually dropped below 2.5. Of the 31 active storm events this PWS actively reported, 10 (32%) were classified as untrustworthy with trust score threshold $\gamma = 4.0$. Notably, as shown in Figure 6, KTXHOUST1903 initially received high trust scores, but dropped below 5.0 during several storm events. However, its trust score was restored to above 5.0 after storm event 20170715, and remained mostly trustworthy for the rest of the time. Other PWSs, such as KTXHOUST1971, KTXHOUST327, and KTXHOUST355, received relatively higher trust scores and were classified as trustworthy for most of the storm events (Figure 6). In cluster 2, KTXHOUST452 and KTXHOUST240 were the least trustworthy PWSs with an average trust score less than the threshold $\gamma = 4.0$ for 83% and 61% of the storm events, respectively. KTXHOUST443, with 29% of the storm events evaluated as untrustworthy received a high trust score at the beginning of the analysis period but decreased over time and eventually dropped below 2.5. Other PWSs were mostly trustworthy during the storm events based on the average trust scores they received.

Figure 4. Example of a trustworthy PWS for a storm event. Trust score steadily increases during a storm event when the observed rainfall of a PWS (black dashed line) matches well with the consensus (robust average, shown in red line) of neighboring stations’ reported rainfall (gray lines) consensus. No flags were identified by the PWS QC method in this storm event. 

![Figure 4](image-url)
4.1.3. Comparison With PWS Quality Control Method

A comparison with PWS QC method (Table 4) shows that for each PWS, the number of untrustworthy events (defined as storm events that average trust scores below a threshold) identified by RSCRN generally agreed with the number of flagged events (defined as storm events that had at least one observation flagged by the PWS QC method) for the analyzed storm events. In cluster 1, PWSs with a large number of untrustworthy events were...
Figure 6. RSCRN trust score evolution during analyzed storm events. The gray markers indicate the PWS was not reporting any data during the storm events, and thus the trust score remained constant.
also frequently flagged by the PWS QC method, whereas PWSs that were assigned higher trust scores usually received fewer or no flagged events. In cluster 2, PWSs with a higher percentage of untrustworthy events also received several flags from the PWS QC method. Using a different trust score threshold for classifying PWSs, the comparison showed that the number of untrustworthy events agreed most with the flagged events from the PWS QC method when using $\gamma = 4.0$.

As most of the agreements between the RSCRN and PWS QC method were for high influx and faulty zero flags (Figure 5), there were cases where RSCRN identified additional untrustworthy behavior, whereas there were no flags determined by the PWS QC method. Using the storm event January 20, 2017, as an example (Figure 7), the rainfall observed from this PWS (KTXHOUST281) received several negative outcomes from the RSCRN. At 17:30 and 18:15, the reported rainfall was 1.02 and 4.06 mm, while the robust average computed from the neighboring PWSs were 5.06 and 0.82 mm, respectively. This caused the trust score of the PWS to drop lower than the threshold value for this storm event. However, no observations were flagged by the PWS QC method in this event because none of the observations in this storm event met the predefined filter threshold of FZ, HI, and SO flags.

In a second example using PWS KTXHOUST443 (cluster 2) and storm event June 25, 2017 (Figure 8), the rainfall reported from this station was much higher than the neighboring consensus, which resulted in a decrease in the trust score because of some time steps of negative outcomes identified by the RSCRN. At 17:30 and 18:15, the reported rainfall was 1.02 and 4.06 mm, while the robust average computed from the neighboring PWSs were 5.06 and 0.82 mm, respectively. This caused the trust score of the PWS to drop lower than the threshold value for this storm event. However, no observations were flagged by the PWS QC method in this event because none of the observations in this storm event met the predefined filter threshold of FZ, HI, and SO flags.

In a second example using PWS KTXHOUST443 (cluster 2) and storm event June 25, 2017 (Figure 8), the rainfall reported from this station was much higher than the neighboring consensus, which resulted in a decrease in the trust score because of some time steps of negative outcomes identified by the RSCRN. At 17:30 and 18:15, the reported rainfall was 1.02 and 4.06 mm, while the robust average computed from the neighboring PWSs were 5.06 and 0.82 mm, respectively. This caused the trust score of the PWS to drop lower than the threshold value for this storm event. However, no observations were flagged by the PWS QC method in this event because none of the observations in this storm event met the predefined filter threshold of FZ, HI, and SO flags.
4.2. Validation Using High-Fidelity Rainfall Stations

To validate whether RSCRN can result in higher accuracy of PWS-derived rainfall estimates, the RMSE between rainfall observations from PWSs to high-fidelity rainfall stations (HCFC) at storm events was computed. As shown in Figure 2, the HCFC rain gauges (445 and 435 in cluster 1, 520 and 560 in cluster 2) were in close proximity (mostly less than 1 km) with PWSs in the clusters and thus were used as the ground truth of actual rainfall observations for validation. The forgetting factor \( \lambda \) was set to 0.95, and the trust score threshold \( \gamma \) was set to 5.0 in this validation, determined by the best performance of the resulting RMSE.

Tables 5 and 6 show the RMSE comparison of PWS rainfall estimates for the analyzed storm events. In these comparisons, the RMSEs were computed using all PWSs (Equation 14, denoted as \( R_{\text{all}} \)), trustworthy PWSs (Equation 15, denoted as \( R_{\text{RSCRN}} \)), and QC PWSs (Equation 17, denoted as \( R_{\text{QC}} \)). The resulting \( R_{\text{all}} \) ranged from 0.43 to 3.41 mm across the analyzed storm events, except for the storm event March 5, 2017, in cluster 1 for which a single PWS (KTXHOUST355) reported an extreme value of 1462.53 mm, which resulted in much higher \( R_{\text{RSCRN}} \) for this particular storm event (column 1 in Table 5). It is worth noting that, because the RSCRN method did not rely on only a single observation to determine the trustworthiness of a PWS, it did not identify this station as untrustworthy for this storm event, which resulted in worse performance at this particular storm event. This suggests, however, that RSCRN could be used along with basic outlier detection methods to improve its results.

The overall performances for both methods are shown in Table 7. Using the RSCRN method, the results showed that of the 33 analyzed storm events in cluster 1, \( R_{\text{RSCRN}} \) outperformed \( R_{\text{all}} \) for 25 (76%) of the events, with a median RMSE improvement (ΔRMSE) of 0.35 (24.5%) (bold values in column 4 of Table 5, which is computed by subtracting \( R_{\text{RSCRN}} \) with \( R_{\text{all}} \)). Of the 29 analyzed storm events in cluster 2, \( R_{\text{RSCRN}} \) outperformed \( R_{\text{all}} \) for 23 (79%) of the events, with a median RMSE improvement of 0.41 (29.8%) (bold values in column 4 of Table 6). Using the PWS QC method, results showed that \( R_{\text{QC}} \) improved for 21 (64%) of the events in cluster 1, and 17 (59%) of
the events in cluster 2 (bold values in column 5 of Tables 5 and 6). This demonstrates that both approaches made significant improvements in PWS rainfall estimates for the majority of the storm events.

Comparison of \( R_{\text{RSCRN}} \) and \( R_{\text{QC}} \) showed that RSCRN generally outperformed the PWS QC method (Table 7). In cluster 1, the results showed that 2 (6%) storm events have the same performance. In the remaining 31 storm events, RSCRN outperformed the PWS QC method in 22 (71%) of the storm events, with a median RMSE improvement of 0.16 (11.5%) (shown in bold values in column 6 of Table 5), while the PWS QC method outperformed RSCRN in 9 (29%) of the storm events (shown in italic values in the column 6 of Table 5). In cluster 2, the results showed that 1 (3%) storm event had the same performance. In the remaining 28 storm events, RSCRN outperformed the PWS QC method in 22 (79%) storm events, with a median RMSE improvement of 0.26 (21.7%), while the PWS QC method outperformed RSCRN in 6 (21%) storm events. This suggests that the RSCRN approach identified additional untrustworthy PWSs that were not flagged by the PWS QC method, and thus improved the rainfall estimates from PWS network.

5. Discussion

5.1. The Efficacy of RSCRN

The proposed RSCRN provides a framework for evaluating the trustworthiness of crowdsourced PWSs. The trust scores derived from RSCRN were demonstrated to reflect both the historical and real-time level of trustworthiness of a PWS for capturing rainfall during a storm event, which could be used to improve rainfall estimates from crowdsourced PWSs. However, the efficacy of the RSCRN can be affected by several factors. For example, RSCRN would generally perform better in populated regions with a larger number of PWSs in close proximity, while the algorithm would be less effective in regions with larger PWS inter-distances and fewer PWSs. In addition, the parameter choices can also affect the obtained results from the RSCRN algorithm. As discussed in

![Figure 8. RSCRN algorithm identified several negative outcomes (mostly overreporting rainfall) in a storm event with large spatial and temporal rainfall variability.](image-url)
Sections 4.1.1 and 4.1.2, using a higher forgetting factor \( \lambda \) ensures longer history of the observations were accounted for, but the resulting trust scores would be less sensitive to sudden changes. A higher trust score threshold \( \gamma \) would result in more PWSs being classified as untrustworthy, but one might risk excluding trustworthy PWSs that are actually contributing trustworthy rainfall observations. On the other hand, choosing a lower \( \gamma \) might result in failing to identify a PWS that is actually untrustworthy. Furthermore, the choice of these parameters will also

### Table 5

| Storm event date      | (1) \( R_{\text{all}} \) | (2) \( R_{\text{RSCRN}} \) | (3) \( R_{\text{QC}} \) | (4) \( R_{\text{all}} - R_{\text{RSCRN}} \) | (5) \( R_{\text{all}} - R_{\text{QC}} \) | (6) \( R_{\text{QC}} - R_{\text{RSCRN}} \) |
|-----------------------|---------------------------|----------------------------|------------------------|-----------------------------|-----------------------------|-----------------------------|
| January 2, 2017       | 1.78                      | 0.87 (3)                   | 1.78 (0)               | 0.92                        | 0.00                        | 0.92                        |
| January 18, 2017      | 1.47                      | 1.34 (1)                   | 1.42 (0)               | 0.13                        | 0.05                        | 0.08                        |
| January 20, 2017      | 1.22                      | 1.13 (2)                   | 1.25 (0)               | 0.09                        | -0.03                       | 0.13                        |
| March 5, 2017         | 27.82                     | 41.13 (2)                  | 0.67 (3)               | -13.31                      | 27.15                       | -40.46                      |
| March 29, 2017        | 2.77                      | 2.52 (1)                   | 2.60 (3)               | 0.25                        | 0.17                        | 0.08                        |
| April 28, 2017        | 1.27                      | 1.72 (2)                   | 0.50 (3)               | -0.46                       | 0.77                        | -1.23                       |
| May 22, 2017          | 1.56                      | 1.10 (4)                   | 1.18 (2)               | 0.46                        | 0.38                        | 0.08                        |
| May 29, 2017          | 3.41                      | 1.15 (3)                   | 1.82 (2)               | 2.26                        | 1.59                        | 0.67                        |
| June 4, 2017          | 2.86                      | 1.20 (3)                   | 1.23 (3)               | 1.66                        | 1.63                        | 0.03                        |
| July 24, 2017         | 0.87                      | 0.80 (1)                   | 0.85 (2)               | 0.07                        | 0.02                        | 0.05                        |
| August 15, 2017       | 2.10                      | 2.14 (1)                   | 2.78 (1)               | -0.04                       | -0.68                       | 0.64                        |
| September 2, 2017     | 1.26                      | 1.33 (1)                   | 1.53 (1)               | -0.08                       | -0.27                       | 0.20                        |
| September 8, 2017     | 1.23                      | 1.05 (1)                   | 0.82 (1)               | 0.18                        | 0.41                        | -0.23                       |
| September 25, 2017    | 2.80                      | 2.22 (1)                   | 2.42 (1)               | 0.58                        | 0.38                        | 0.20                        |
| October 18, 2017      | 1.99                      | 1.92 (2)                   | 1.73 (1)               | 0.07                        | 0.25                        | -0.18                       |
| December 3, 2017      | 1.36                      | 0.88 (3)                   | 1.13 (1)               | 0.48                        | 0.23                        | 0.25                        |
| December 16, 2017     | 0.83                      | 0.54 (2)                   | 0.52 (2)               | 0.29                        | 0.31                        | -0.02                       |
| February 10, 2018     | 1.19                      | 0.88 (2)                   | 0.883 (1)              | 0.31                        | 0.31                        | 0.003                       |
| March 29, 2018        | 2.20                      | 2.00 (2)                   | 2.37 (2)               | 0.20                        | -0.17                       | 0.37                        |
| April 21, 2018        | 1.73                      | 1.37 (1)                   | 1.30 (1)               | 0.35                        | 0.43                        | -0.07                       |
| May 21, 2018          | 1.23                      | 0.75 (3)                   | 1.27 (1)               | 0.48                        | -0.04                       | 0.52                        |
| July 4, 2018          | 1.93                      | 1.44 (1)                   | 1.24 (1)               | 0.48                        | 0.68                        | -0.20                       |
| July 31, 2018         | 1.00                      | 0.80 (3)                   | 1.08 (1)               | 0.20                        | -0.08                       | 0.28                        |
| September 9, 2018     | 1.18                      | 0.78 (1)                   | 2.16 (1)               | 0.40                        | -0.98                       | 1.38                        |
| October 15, 2018      | 1.77                      | 1.33 (3)                   | 1.90 (1)               | 0.43                        | -0.13                       | 0.57                        |
| October 31, 2018      | 1.25                      | 1.08 (1)                   | 1.18 (2)               | 0.17                        | 0.08                        | 0.10                        |
| December 7, 2018      | 1.51                      | 0.82 (2)                   | 0.76 (2)               | 0.69                        | 0.75                        | -0.06                       |
| December 13, 2018     | 0.60                      | 0.60 (2)                   | 0.58 (1)               | 0.00                        | 0.02                        | -0.02                       |
| December 27, 2018     | 0.47                      | 0.48 (1)                   | 0.58 (1)               | -0.01                       | -0.11                       | 0.11                        |
| January 2, 2019       | 2.37                      | 0.45 (2)                   | 0.46 (1)               | 1.92                        | 1.91                        | 0.01                        |
| January 19, 2019      | 0.80                      | 0.80 (2)                   | 0.85 (1)               | 0.00                        | -0.05                       | 0.05                        |
| January 23, 2019      | 0.43                      | 0.40 (1)                   | 0.40 (0)               | 0.03                        | 0.03                        | 0.00                        |
| February 26, 2019     | 0.55                      | 0.55 (2)                   | 0.55 (0)               | 0.00                        | 0.00                        | 0.00                        |

**Note.** Columns 1 to 3 are the average RMSE of rainfall estimates computed from all PWSs, trustworthy PWSs only and QC PWSs only, respectively. The values in the parentheses in Columns 2 and 3 are the number of untrustworthy and flagged PWSs, respectively. Column 4 shows the improvements made from trustworthy PWSs (\( R_{\text{RSCRN}} - R_{\text{all}} \)); Column 5 shows the improvements made from QC PWSs (\( R_{\text{QC}} - R_{\text{all}} \)); Column 6 shows the improvement between RSCRN and PWS QC method (\( R_{\text{QC}} - R_{\text{RSCRN}} \)). Bold and italic values represent decrease and increase in RMSE, respectively.
Table 6
Comparison of RMSE Improvements for PWS Rainfall Estimates Across Storm Events Using RSCRN and PWS QC Method for Cluster 2

| Storm event date | $R_{all}$ | $R_{RSCRN}$ | $R_{QC}$ | $R_{all} - R_{RSCRN}$ | $R_{all} - R_{QC}$ | $R_{QC} - R_{RSCRN}$ |
|------------------|-----------|-------------|---------|------------------------|------------------|---------------------|
| January 2, 2017  | 0.74      | 0.88 (1)    | 0.74 (0)| −0.14                  | 0.00             | −0.14               |
| January 18, 2017 | 3.15      | 2.78 (2)    | 3.24 (1)| 0.38                   | −0.09            | 0.47                |
| January 20, 2017 | 2.07      | 1.93 (3)    | 2.17 (0)| 0.13                   | −0.10            | 0.23                |
| February 20, 2017| 0.80      | 0.50 (2)    | 0.82 (0)| 0.30                   | −0.02            | 0.32                |
| March 5, 2017    | 1.10      | 0.63 (2)    | 0.94 (0)| 0.47                   | 0.16             | 0.31                |
| March 29, 2017   | 3.50      | 0.75 (3)    | 2.73 (1)| 2.75                   | 0.78             | 1.98                |
| April 11, 2017   | 0.97      | 0.68 (2)    | 0.86 (1)| 0.29                   | 0.11             | 0.19                |
| May 22, 2017     | 1.19      | 1.53 (3)    | 1.58 (2)| −0.34                  | −0.39            | 0.06                |
| June 4, 2017     | 1.69      | 0.98 (2)    | 1.22 (1)| 0.71                   | 0.47             | 0.24                |
| June 25, 2017    | 1.99      | 2.06 (2)    | 2.18 (1)| −0.07                  | −0.20            | 0.12                |
| July 13, 2017    | 2.51      | 1.64 (2)    | 1.88 (2)| 0.87                   | 0.63             | 0.24                |
| August 7, 2017   | 1.07      | 0.90 (2)    | 1.22 (1)| 0.17                   | −0.15            | 0.32                |
| August 26, 2017  | 7.40      | 4.10 (2)    | 4.90 (2)| 3.30                   | 3.50             | −0.20               |
| September 18, 2017| 1.70     | 1.85 (2)    | 1.72 (1)| −0.15                  | −0.02            | 0.13                |
| January 8, 2018  | 2.17      | 2.76 (2)    | 1.65 (1)| −0.59                  | 0.52             | −1.11               |
| February 10, 2018| 1.21      | 1.10 (1)    | 1.01 (1)| 0.11                   | 0.20             | −0.09               |
| March 29, 2018   | 1.81      | 1.50 (3)    | 1.69 (1)| 0.31                   | 0.13             | 0.19                |
| April 21, 2018   | 2.00      | 1.61 (1)    | 1.69 (1)| 0.39                   | 0.31             | 0.07                |
| May 20, 2018     | 1.20      | 1.20 (2)    | 1.10 (2)| 0.00                   | 0.10             | −0.10               |
| July 4, 2018     | 3.10      | 2.18 (2)    | 3.03 (0)| 0.93                   | 0.07             | 0.86                |
| September 9, 2018| 2.46      | 0.98 (2)    | 1.98 (2)| 1.48                   | 0.48             | 1.00                |
| September 22, 2018| 1.16     | 0.62 (2)    | 0.62 (2)| 0.54                   | 0.54             | 0.00                |
| October 15, 2018 | 2.33      | 1.92 (2)    | 3.58 (1)| 0.41                   | −1.25            | 1.66                |
| October 31, 2018 | 1.40      | 0.46 (3)    | 1.40 (2)| 0.94                   | 0.00             | 0.94                |
| December 7, 2018 | 1.55      | 1.06 (2)    | 1.65 (4)| 0.49                   | −0.10            | 0.59                |
| December 13, 2018| 1.44      | 1.02 (2)    | 1.40 (2)| 0.42                   | 0.04             | 0.38                |
| December 27, 2018| 1.11      | 0.88 (3)    | 1.08 (1)| 0.24                   | 0.03             | 0.21                |
| January 2, 2019  | 0.67      | 0.48 (3)    | 0.61 (1)| 0.19                   | 0.06             | 0.13                |
| January 23, 2019 | 0.60      | 0.50 (3)    | 0.60 (1)| 0.10                   | 0.00             | 0.10                |

Note. Bold and italic values represent decrease and increase in RMSE, respectively.

5.2. The Assumption of Consensus in Rainfall Observations

One of the premises behind RSCRN is that consensus in crowdsourced rainfall observations exists at some scale in space and time and can, therefore, be used to judge trustworthiness of stations within a cluster. Such consensus-based ideas are widely used across disciplines to identify errors in data (Foody et al., 2018; Strobl et al., 2019; Zhang et al., 2017). Strong consensus in rainfall observations occurs when rain gauges are located in close proximity, but the exact distance and other factors that should be used for defining a cluster are uncertain. This idea is not new, however. For example, the United States Climate Reference Network (USCRN), an extremely high-fidelity rainfall network, uses three distinct tipping bucket sensors installed next to each other on the same site for immediate detection of single sensor failure (Diamond et al., 2013). Better accounting for factors that influence consensus in rainfall observations (e.g., geography, climate, observation frequency) are possible extensions to the approach used in this study. In addition, in this work, a cluster was identified based on a group of PWSs that were close to each other. However, large rainfall variability may exist even over short distances, especially for high frequency rainfall observations or if stations have large elevation differences. For example, this particular case study focused on Houston, Texas, which has only slight topographical variation across the region; thus elevation was not a factor in station clustering. However, for regions with greater variation in elevation such as mountainous areas, the clustering results should include elevation to reflect where the consensus actually exists (Buytaert et al., 2006). Rainfall types can also be one of the factors that affect the consensus. For example, a convective storm may produce rainfall over a small area that is only captured by a single PWS in a cluster if clusters are not carefully created. In these cases, incorporating additional variables of the PWS location into the clustering method may better capture the consensus and thus result in more meaningful trust scores.

5.3. Feedback to Data Collectors for Improved Crowdsourced Data Quality

People are motivated by various kinds of incentives to adopt a PWS. Examples of these incentives include obtaining useful weather data for personal purposes or having a sense of belonging to a community of friends with shared interests (Garesifard & Wehn, 2016). As the need for higher spatial and temporal resolution of rainfall data increases, the role of PWS data may be shifted from serving personal interests to benefiting society at large (Garesifard & Wehn, 2016). In this case, because people who need to use the data are interested in knowing the quality of the data they contributed, PWS owners might become what Jøsang et al. (2007) described as service providers. To manage their provision trust, they may be willing to demonstrate their competence in collecting data and arguably welcome any feedback to improve their data quality. As a result, RSCRN could assist by making the trust score information available to the PWS owners. PWS owners could be notified of a drop in the trust score, and actions could be taken to correct the erroneous observations (e.g., cleaning the clogged rain gauge). Such efforts not only help
restore the trust scores, which maintain their provision trust, but also greatly improve the overall data quality of the crowdsourced rainfall network in the long term. Future improvements to RSCRN could focus on identifying particular types of errors to more effectively advise users on steps to improve their trust score.

5.4. Limitation of Binary Trust Score Threshold

Trust scores derived from the RSCRN represent the relative frequency of a PWS reporting trustworthy rainfall observations in the future. This continuous form is computationally efficient for reputation systems to calculate and update over time (Ruan & Durresi, 2016). However, to better enable reputation-based decision making, a discrete format of trust scores is often used (Mousa et al., 2015), as humans are generally better able to understand discrete verbal statements than continuous measures (Jøsang et al., 2007). In this study, we used a RSCRN-derived trust score threshold approach to classify PWSs as either trustworthy or untrustworthy. This can be thought of converting a continuous measure of trust metric into a binary discrete format. While using a binary trust score threshold is simple and straightforward for enabling decision making (e.g., include or ignore rainfall from a specific PWS), it does not represent the varying trustworthiness of PWSs (Ruan & Durresi, 2016). For example, a PWS with an extremely low trust score and a PWS with a trust score just below the threshold will both be categorized as an untrustworthy PWS, despite the difference in the extent of their untrustworthiness. Alternatives can be dynamically adjusting the binary trust score threshold to optimize decision-making or use of multinomial discrete values such as very trustworthy, trustworthy, untrustworthy, very untrustworthy to account for a broader extent of trustworthiness across PWSs (Ruan & Durresi, 2016). Future work could explore extensions like this so that RSCRN is able to weight information from PWSs based on their trust score rather than simply including or excluding measurements using a threshold method.

5.5. The Potential of Crowdsourced Rainfall Data

While this study focuses on crowdsourced rainfall data collected from PWSs, the proposed RSCRN can be beneficial to ensure the trustworthiness of other emerging crowdsourcing rainfall data collection methods as well. Beyond the case of PWSs, recent improvement in crowdsourcing methods has further enabled rainfall observations to be collected from connected vehicles (Bartos et al., 2019), surveillance cameras (Jiang et al., 2019), and mobile phones (Guo et al., 2018). The availability of these crowdsourcing methods greatly facilitates more crowd-participation, but also raises concerns of increased uncertainty associated with data contributors, highlighting the need for evaluating the trustworthiness of crowdsourced data (Gharesifard & Wehn, 2016; Hunter et al., 2013). In particular for fixed point observations such as PWSs and surveillance cameras, RSCRN can serve as a starting point for creating algorithms capable of systematically assigning the trustworthiness of these data based on physical principles able to be applied at scale for quickly growing networks.

5.6. The Availability and Reliability of Crowdsourced Data

PWS adoption has been growing rapidly thanks to the advancement of technologies that have made PWSs easy to install and affordable, as well as development of software able to connect and share the data through online
platforms. This increase in PWS data openly shared on the Internet has transformed the value of PWSs from serving the owners’ interests to appealing to anyone in the broader community who might benefit from the information (Gharesifard & Wehn, 2016). However, PWS data accessibility depends heavily on the platform the PWSs are connected to. For example, the Weather Underground recently ended a freely available service of the Weather Underground API and replaced it with a new API service that allows only PWS contributors to utilize the service (WXForum, 2018a). The Weather Underground has also stopped the automatic connection of PWSs of certain brands (e.g., Netatmo) to their platform (WXForum, 2018b), resulting in abrupt changes to the number of sensors available in the system. These kinds of sudden changes might happen in any crowdsourced platform without warning, which could further compromise accessibility and reduce the utility of crowdsourced data. The community would benefit from more standardization of open networks and data sharing agreements to make the most of this emerging data resource.

6. Conclusion

In this study, we presented a RSCRN for ensuring the trustworthiness of PWSs in a crowdsourced rainfall network. The RSCRN assigned trust scores to PWSs are calculated by (a) clustering the PWSs into groups with similar rainfall characteristics, (b) computing the rainfall observation consensus within each cluster using a robust average method, and (c) deriving trust scores using a beta reputation system.

Using PWS rainfall data collected from Houston, Texas, as a case study, we demonstrated how RSCRN is able to identify PWSs with untrustworthy rainfall data. By ignoring rainfall from untrustworthy PWSs using a RSCRN-derived trust score threshold, the accuracy of the resulting 15-min rainfall estimates better matched rainfall observations observed from high-fidelity rainfall stations for 77% (48 out of 62) of the analyzed storm events, with a median RMSE improvement of 27.3%. Compared to an existing PWS quality control method, results showed that while 3 (5.0%) storm events had the same performance, RSCRN improved rainfall estimates for 74% of the remaining storm events 44 out of 59, with a median RMSE improvement of 18.7%.

We return now to the research questions from Section 1 that guided this work to provide answers based on the research outcome.

1. How can we evaluate the trustworthiness of crowdsourced PWSs?

   This study demonstrated that a reputation system approach could be useful in evaluating the trustworthiness of crowdsourced PWSs. Unlike a traditional QA/QC method, the reputation system approach collectively evaluates the trustworthiness of a PWS itself over time rather than from single observations collected at a gauge. The RSCRN presented in this study assigns trust scores to PWSs based on their agreement or disagreement with current and historical rainfall observations from neighboring PWS, and is able to converge to a confident trust score in 20–30 time steps, as well as accommodate sudden changes in trust levels.

2. To what extent could a reputation system approach improve rainfall estimates from PWSs?

   The reputation system can be used to improve rainfall estimates in direct and indirect ways. First, the reputation system approach ensures the rainfall estimates were produced from trustworthy PWSs. Using threshold for RSCRN-derived trust scores, PWSs were classified as trustworthy or untrustworthy. By including only trustworthy PWSs in the rainfall estimation process, the resulting trustworthy rainfall estimates were greatly improved in accuracy when compared to rainfall observed from high-fidelity rainfall stations. Second, the reputation system approach has the potential to encourage PWS owners to maintain and contribute high quality data, which indirectly improves rainfall estimates from PWSs in the long term.

Future work could be aimed at (a) a larger analysis of crowdsourced rainfall networks to identify and quantify the extent of untrustworthy PWSs across cities and regions in the world, (b) enhancing the reputation system algorithm to account for rainfall variability in complex topography and finer-temporal scales, and (c) leveraging crowdsourced rainfall estimates to improve hydrological modeling such as rainfall-runoff and flood prediction. With a reputation system able to ensure the trustworthiness of PWSs and improve the quality of data collected through crowdsourced rainfall networks, this growing data resource can be more confidently adopted and trusted for water resources management and decision-making.
Data Availability Statement
Python codes of the RSCRN and datasets used in this study can be accessed from Hydroshare (https://www.hydroshare.org/resource/ec7796cedeae42818d9bd7f95f8e1872/).

Acknowledgments
This research is supported by the National Science Foundation under Grant CBET-1735587. We gratefully acknowledge the Weather Underground and Harris County Flood Control District for access to their data.

References
Bartos, M., Park, H., Zhou, T., Kerkez, B., & Vasudevan, R. (2019). Windshield wipers on connected vehicles produce high-accuracy rainfall maps. Scientific Reports, 9(1), 170. https://doi.org/10.1038/s41598-018-36282-7
Bäserud, L., Lussana, C., Nipen, T. N., Seierstad, I. A., Orm, L., & Aspelien, T. (2020). Titan automatic spatial quality control of meteorological in-situ observations. Advances in Science and Research, 17, 153–163. https://doi.org/10.5194/asr-17-153-2020
Bell, S., Cornell, D., & Bastin, L. (2013). The state of automated amateur weather observations. Weather, 68(2), 36–41. https://doi.org/10.1002/wea.1980
Berne, A., Delrieu, G., Creutin, J.-D., & Obled, C. (2004). Temporal and spatial resolution of rainfall measurements required for urban hydrology. Journal of Hydrology, 299(3–4), 166–179. https://doi.org/10.1016/j.jhydrol.2004.03.063
Blenkisop, S., Lewis, E., Chan, S. C., & Fowler, H. J. (2017). Quality-control of an hourly rainfall dataset and climatology of extremes for the UK. International Journal of Climatology, 37(2), 722–740. https://doi.org/10.1002/joc.4733
Buytaert, W., Celleri, R., Willems, P., De Bie, B., & Wyseure, G. (2006). Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. Journal of Hydrology, 329(3–4), 413–421. https://doi.org/10.1016/j.jhydrol.2006.02.031
Chapman, L., Bell, C., & Bell, S. (2017). Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. International Journal of Climatology, 37(9), 3597–3605. https://doi.org/10.1002/joc.4940
Chen, A. B., Behl, M., & Goodall, J. L. (2018). Trust me, my neighbors say it’s raining outside: Ensuring data trustworthiness for crowdsourced weather stations. In Proceedings of the 5th conference on systems for built environments (pp. 25–28).
Chou, C. T., Ignjatovic, A., & Hu, W. (2013). Efficient computation of robust average of compressive sensing data in wireless sensor networks in the presence of sensor faults. IEEE Transactions on Parallel and Distributed Systems, 24(8), 1525–1534. https://doi.org/10.1109/tpds.2012.260
Cox, L. P. (2011). Truth in crowdsourcing. IEEE Security & Privacy, 9(5), 74–76. https://doi.org/10.1109/msp.2011.145
Cristiano, E., Veldhuis, M.-c. t., & Giesen, N. V. D. (2017). Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas—a review. Hydrology and Earth System Sciences, 21(7), 3859–3878. https://doi.org/10.5194/hess-21-3859-2017
Cunha, L. K., Smith, J. A., Krajewski, W. F., Baeck, M. L., & Seo, B.-C. (2015). Nexrad nws polarimetric precipitation product evaluation for flood forecasting. Journal of Hydrometeorology, 16(4), 1676–1699. https://doi.org/10.1175/jhm-d-14-0148.1
de Vos, L. W., Leijne, H., Overeem, A., & Uijlenhoet, R. (2017). The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam. Hydrology and Earth System Sciences, 21(2), 765–777. https://doi.org/10.5194/hess-21-765-2017
de Vos, L. W., Leijne, H., Overeem, A., & Uijlenhoet, R. (2019). Quality control for crowdsourced personal weather stations to enable operational rainfall monitoring. Geophysical Research Letters, 46(15), 8820–8829. https://doi.org/10.1029/2019gl083731
Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., et al. (2013). US climate reference network after one decade of operations: Status and assessment. Bulletin of the American Meteorological Society, 94(4), 485–498. https://doi.org/10.1175/bams-d-12-00170.1
Estellés-Arolas, E., & González-Ladrón-De-Guevara, F. (2012). Towards an integrated crowdsourcing definition. Journal of Information Science, 38(2), 189–200. https://doi.org/10.1177/0192942712453768
Estévez, J., Gaviñán, P., & Giráldez, J. V. (2011). Guidelines on validation procedures for meteorological data from automatic weather stations. Journal of Hydrology, 402(1–2), 144–154. https://doi.org/10.1016/j.jhydrol.2011.02.031
Fiebrich, C. A., Morgan, C. R., McCombs, A. G., Hall, P. K., & McPherson, R. A. (2010). Quality assurance procedures for mesoscale meteorological data. Journal of Atmospheric and Oceanic Technology, 27(10), 1565–1582. https://doi.org/10.1175/2010jtecha1433.1
Fletcher, T. D., Andrieu, H., & Hamel, P. (2013). Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. Advances in Water Resources, 51, 261–279. https://doi.org/10.1016/j.advwatres.2012.09.001
Foody, G., See, L., Frizz, S., Moorthy, I., Perger, C., Schill, C., & Boyd, D. (2018). Increasing the accuracy of crowdsourced information on land cover via a voting procedure weighted by information inferred from the contributed data. ISPRS International Journal of Geo-Information, 7(3), 80. https://doi.org/10.3390/ijgi7030080
Ganeriwal, S., Balzano, L. K., & Srivastava, M. B. (2008). Reputation-based framework for high integrity sensor networks. ACM Transactions on Sensor Networks, 4(3), 15. https://doi.org/10.1145/1362542.1362546
Gharesifard, M., & Wehm, U. (2016). To share or not to share: Drivers and barriers for sharing data via online amateur weather networks. Journal of Hydrology, 535, 181–190. https://doi.org/10.1016/j.jhydrol.2016.01.036
Guo, H., Huang, H., Sun, Y.-E., Zhang, Y., Chen, S., & Huang, L. (2018). Chaac: Real-time and fine-grained rain detection and measurement using smartphones. IEEE Internet of Things Journal, 6(1), 997–1009.
Huang, K. L., Kanhere, S. S., & Hu, W. (2014). On the need for a reputation system in mobile phone based sensing. Ad Hoc Networks, 12, 130–149. https://doi.org/10.1016/j.adhoc.2011.12.002
Hunter, J., Alabri, A., & van Ingen, C. (2013). Assessing the quality and trustworthiness of citizen science data. Concurrency and Computation: Practice and Experience, 25(4), 454–466. https://doi.org/10.1002/cpe.2923
Jiang, S., Babovic, V., Zheng, Y., & Xiong, J. (2019). Advancing opportunistic sensing in hydrology: A novel approach to measuring rainfall with ordinary surveillance cameras. Water Resources Research, 55(4), 3004–3027. https://doi.org/10.1029/2018wr024480
Josang, A., & Ismail, R. (2002). The beta reputation system. In Proceedings of the 15th Bled Electronic Commerce Conference. (Vol. 5, pp. 2502–2511).
Josang, A., Ismail, R., & Boyd, C. (2007). A survey of trust and reputation systems for online service provision. Decision Support Systems, 43(2), 618–644.
Krajewski, W., & Smith, J. A. (2002). Radar hydrology: Rainfall estimation. Advances in Water Resources, 25(8–12), 1387–1394. https://doi.org/10.1016/s0309-1708(02)00062-3
Lowry, C. S., & Fienen, M. N. (2013). Crowdhydrology: Crowdsourcing hydrologic data and engaging citizen scientists. Ground Water, 51(1), 151–156. https://doi.org/10.1111/j.1745-6584.2012.00956.x
Meier, F., Fenner, D., Grassmann, T., Otto, M., & Scherer, D. (2017). Crowdsourcing air temperature from citizen weather stations for urban climate research. *Urban Climate, 19*, 170–191. https://doi.org/10.1016/j.urcl.2017.01.006

Mosavi, O., Ozturk, P., & Chau, K.-w. (2018). Flood prediction using machine learning models: Literature review. *Water, 10*(11), 1536. https://doi.org/10.3390/w10111536

Mousa, H., Mohiktar, B. B., Hasan, O., Younes, O., Hadhoud, M., & Brunie, L. (2015). Trust management and reputation systems in mobile participatory sensing applications: A survey. *Computer Networks, 90*, 49–73. https://doi.org/10.1016/j.comnet.2015.07.011

Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingsworth, S., Foody, G., et al. (2015). Crowdsourcing for climate and atmospheric sciences: Current status and future potential. *International Journal of Climatology, 35*(11), 3185–3203. https://doi.org/10.1002/joc.4210

Niu, G., Yang, P., Zheng, Y., Cai, X., & Qin, H. (2021). Automatic quality control of crowdsourced rainfall data with multiple noises: A machine learning approach, e2020WR029121. *Water Resources Research.*

Obba, M., & Sugimoto, S. (2019). Differences in climate change impacts between weather patterns: Possible effects on spatial heterogeneous changes in future extreme rainfall. *Climate Dynamics, 52*(7–8), 4177–4191. https://doi.org/10.1007/s00382-018-4374-1

Overeem, A., Leijnse, H., & Uijlenhoet, R. (2013). Country-wide rainfall maps from cellular communication networks. *Proceedings of the National Academy of Sciences, 110*(8), 2741–2745. https://doi.org/10.1073/pnas.1217961110

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., et al. (2019). The spatial dependence of flood hazard and risk in the United States. *Water Resources Research, 55*(3), 1890–1911. https://doi.org/10.1029/2018wr024205

Resnick, P., Kurowtka, Z., Zeckhauser, R., & Friedman, E. (2000). Reputation systems. *Communications of the ACM, 43*(12), 45–48. https://doi.org/10.1145/355112.355122

Ruan, Y., & Durrasi, A. (2016). A survey of trust management systems for online social communities–trust modeling, trust inference and attacks. *Knowledge-Based Systems, 106*, 150–163. https://doi.org/10.1016/j.knosys.2016.05.042

Sadler, J. M., Goodall, J. L., & Morys, M. M. (2017). Effect of rain gauge proximity on rainfall estimation for problematic urban coastal watersheds in Virginia Beach, Virginia. *Journal of Hydrologic Engineering, 22*(9), 04017036. https://doi.org/10.1061/(asce)he.1943-5584.0001563

Sadler, J. M., Goodall, J. L., Morys, M., & Spencer, K. (2018). Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. *Journal of Hydrology, 559*, 43–55. https://doi.org/10.1016/j.jhydrol.2018.01.044

Saksono, S. D., Day, S., Merwade, V., & Singhoen, P. J. (2019). A computationally efficient and physically based approach for urban flood modeling using a flexible spatiotemporal structure. *Water Resources Research.*

Salman, A. M., & Li, Y. (2018). Flood risk assessment, future trend modeling, and risk communication: A review of ongoing research. *Natural Hazards Review, 19*(3), 04018011. https://doi.org/10.1061/(asce)nh.1527-6996.0000294

Sanchez, L., Rousa, E., & Hidalgo, N. (2018). Crowdsourcing under attack: Detecting malicious behaviors in waze. In *IFIP international conference on trust management (pp. 91–106).* https://doi.org/10.1007/978-3-319-95276-5_7

Savage, J. T. S., Pianosi, F., Bates, P., Freer, J., & Wagener, T. (2016). Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research, 52*(11), 9146–9163. https://doi.org/10.1002/2015wr018198

Sharma, A., Wasko, A., & Lettenmaier, D. P. (2018). If precipitation extremes are increasing, why aren’t floods? *Water Resources Research, 54*(11), 8545–8551. https://doi.org/10.1002/2018wr023749

Shen, Y., Morys, M. M., Huxley, C., Tahvildari, N., & Goodall, J. L. (2019). Flood risk assessment and increased resilience for coastal urban watersheds under the combined impact of storm tide and heavy rainfall. *Journal of Hydrology, 579*, 124159. https://doi.org/10.1016/j.jhydrol.2019.124159

Silverton, J., Harvey, M., Greenwood, R., Dodd, M., Rosewell, J., Rebelo, T., et al. (2015). Crowdsourcing the identification of organisms: A case-study of isopt. ZooKeys, 480, 125–146. https://doi.org/10.3897/zookeys.480.8803

Smith, J. A., Seo, D. J., Baeck, M. L., & Hadlow, M. D. (1996). An intercomparison study of nesrd precipitation estimates. *Water Resources Research, 32*(7), 2035–2045. https://doi.org/10.1029/96wr00270

Strobl, B., Ettler, S., van Meerveld, I., & Seibert, J. (2019). The crowdwater game: A playful way to improve the accuracy of crowdsourced water level class data. *PLoS One, 14*(9). https://doi.org/10.1371/journal.pone.0222579

The Weather Channel. (2018). What is a gold star weather station? Retrieved from https://support.weather.com/s/article/What-is-a-Gold-Star-weather-station?language=en_US

Villarini, G., Mandapaka, P. P., Krajewski, W. F., & Moore, R. J. (2008). Rainfall and sampling uncertainties: A rain gauge perspective. *Reviews of Geophysics, 46*(1), 1–18. https://doi.org/10.1029/2007rg000214

Weeber, B., Jacobs, S., Kraft, P., Rufino, M., & Breuer, L. (2019). Rainfall-runoff modelling using crowdsourced water level data. *Water Resources Research.*

Wilby, R. L., & Keenan, R. (2012). Adapting to flood risk under climate change. *Progress in Physical Geography, 36*(3), 348–378. https://doi.org/10.1177/030913331143238908

WXForum (2018a). Netatmo weather stations do not appear on weatherunderground anymore. Retrieved from https://www.wxforum.net/index.php?topic=35059.0

WXForum (2018b). Netatmo weather stations do not appear on weatherunderground anymore. Retrieved from https://www.wxforum.net/index.php?topic=34815.0

Yang, H., Zhang, J., & Roe, P. (2013). Reputation modelling in citizen science for environmental acoustic data analysis. *Social Network Analysis and Mining, 3*(3), 419–435. https://doi.org/10.1007/s12295-012-0087-3

Yang, P., & Ng, T. L. (2017). Gauging through the crowd: A crowd-sourcing approach to urban rainfall measurement and storm water modeling implications. *Water Resources Research, 53*(11), 9462–9478. https://doi.org/10.1002/2017wr020682

Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morys, M. M., & Behl, M. (2020). Training machine learning surrogate models from a high-fidelity physics-based model: Application for real-time street-scale flood prediction in an urban coastal community. *Water Resources Research, 56*(10), e2019WR027038. https://doi.org/10.1029/2019wr027038

Zhang, J., Zheng, Y., Li, Q., Wu, J., & Wu, X. (2017). Consensus algorithms for biased labeling in crowdsourcing. *Information Sciences, 382*, 254–273. https://doi.org/10.1016/j.ins.2016.12.026

Zhang, W., Villarini, G., Vecchi, G. A., & Smith, J. A. (2018). Urbanization exacerbated the rainfall and flooding caused by Hurricane Harvey in Houston. *Nature, 563*(7731), 384–388. https://doi.org/10.1038/s41586-018-0167-z

Zheng, F., Tao, R., Maier, H. R., See, L., Savic, D., Zhang, T., et al. (2018). Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions. *Reviews of Geophysics, 56*(4), 698–740. https://doi.org/10.1029/2018rg000616

Zhu, Z., Wright, D. B., & Yu, G. (2018). The impact of rainfall space-time structure in flood frequency analysis. *Water Resources Research, 54*(11), 8983–8998. https://doi.org/10.1029/2018wr023550