

Article

Stream Stage Monitoring with Community Science-Contributed Stage Data

Ingrid Luffman ^{1,*}  and Daniel Connors ²¹ Department of Geosciences, East Tennessee State University, Johnson City, TN 37614, USA² Department of Engineering, Engineering Technology and Surveying, East Tennessee State University, Johnson City, TN 37614, USA; connors@etsu.edu

* Correspondence: luffman@etsu.edu

Abstract: Volunteered Geographic Information, data contributed by community scientists, is an increasingly popular tool to collect scientific data, involve the community in scientific research, and provide information and education about a prominent issue. Johnson City, Tennessee, USA has a long history of downtown flooding, and recent redevelopment of two land parcels has created new city parks that mitigate flooding through floodwater storage, additional channel capacity, and reduced impervious surfaces. At Founders Park, a project to collect stage data using text messages from community scientists has collected 1479 stage measurements from 597 participants from May 2017 through July 2021. Text messages were parsed to extract the stage and merged with local precipitation data to assess the stream's response to precipitation. Of 1479 observations, 96.7% were correctly parsed. Only 3% of observations were false positives (parser extracted incorrect stage value) or false negatives (parser unable to extract correct value but usable data were reported). Less than 2% of observations were received between 11 p.m. and 7 a.m., creating an overnight data gap, and fewer than 7% of observations were made during or immediately following precipitation. Regression models for stage using antecedent precipitation explained 21.6% of the variability in stream stage. Increased participation and development of an automated system to record stage data at regular intervals will provide data to validate community observations and develop more robust rainfall–runoff models.

Keywords: community science; stage; rainfall–runoff response; text message; parser

Citation: Luffman, I.; Connors, D. Stream Stage Monitoring with Community Science-Contributed Stage Data. *Hydrology* **2022**, *9*, 11. <https://doi.org/10.3390/hydrology9010011>

Academic Editors: Valentina Colaiuda, Annalina Lombardi and Barbara Tomassetti

Received: 22 December 2021

Accepted: 6 January 2022

Published: 10 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Community science research relies on data collected and submitted by community volunteers and is an increasingly popular tool to collect scientific data, involve the community in scientific research, and provide information and education about a prominent issue [1,2]. The process was arguably first employed to engage the public and extend the reach of science at the turn of the 20th century with the Cooperative Weather Service (since 1890) and the National Audubon Society Christmas Bird Count (since 1900) [2]. Since then, community scientists have been active in long-term ecologic population counts, phenologic studies to assess impacts of climate change, and inventories of invasive and rare plants [3]. In the US, community scientists have focused more on monitoring environmental quality over single species monitoring [4]. More recently, one-time events and collective problem-solving activities have become popular, including data blitzes that generate substantial data over a short time period and online participatory events that engage the larger global community, including those with mobility constraints [5].

In hydrology, recent community science projects include collection of stage data [6–9], flood risk analysis [10], and capture of community perspectives of watershed issues [11], among others. In one study, youth were tasked with documenting local watershed issues of perceived importance using Photovoice, and photos were later data-mined to distill

the youth perspective [11]. Such community science research shares a common purpose: community members collaborate in scientific research to meet real-world goals that support conservation, raise environmental awareness, or inform policy decisions [5]. A thorough review of community science applications to hydrology, including opportunities for water quality monitoring, discharge monitoring, mapping, and conflict resolution, is available in [1]. A more recent review examines benefits and obstacles in community science, classifying projects as related to geohazards, observation/classification, education/outreach, and other foci [12].

Benefits of community science activities include increasing the reach of scientific monitoring and supporting science outreach and education while producing reliable research data. One concern with use of contributed data is data quality [13] and assessing uncertainty in data [12]. Using data from the CrowdHydrology network [6], a decision tree approach was applied to evaluate and clean submitted data, yielding estimates of data uncertainty [14]. In a separate study, data quality in community science were examined and it was concluded that datasets produced by volunteers can be of comparable quality to those produced by professionals, provided methods to boost accuracy and account for bias are in place [2]. They suggest employing iterative design, using simple and systematic tasks, recording of metadata, accounting for collection effort variability, and comparing with data collected by professionals as appropriate methods to ensure data quality. Therefore, dual goals of education/outreach and scientific research are achievable in collaborative community science projects.

One difficulty in processing volunteered geographic information is in adding reported data to a database. This may be resolved through use of freely available data collection apps developed for a specific purpose, e.g., iNaturalist and eBird for biodiversity data [15,16]. It may also be resolved through use of web-based mapping utilities [17] such as ArcGIS products used to map visual pollution in urban settings [18] or contribute data to a global Landslide inventory [19]. The necessity of smartphone use may, however, be a barrier to participation for some aspiring community scientists. One solution is to accept data submitted via text message, as with CrowdHydrology [6], following with an automated process to parse the data and enter them into the project database. This is the approach used in the present study.

In this study we describe a free automated system to collect stream stage data contributed by community scientists and match it with weather data to build a database of rainfall and stream stage. These data may then be used to assess both the stream's hydrologic response to rainfall and patterns in community science data contribution. In particular, the question of who contributes data is of interest. Lowry et al. (2019), reporting on data submission patterns in the CrowdHydrology project [20], found 86% of participants contributed data only once, with 0.1% of participants classed as "champions" who contributed over 100 observations each, and accounting for 19% of contributed data. We seek to assess whether participation in this project breaks down in an equivalent manner.

The aims of the study are, therefore, to (1) identify temporal and individual patterns of community science observations and (2) develop a statistical model of stream stage using antecedent rainfall data and stage observations from community scientists.

2. Materials and Methods

2.1. Study Area and Flood History

The research was conducted on Brush Creek in Johnson City, Tennessee, USA (HUC 060101030505). The stream's 42 km² watershed (with a contributing area of 19.2 km² upstream of the staff gauge used in this study) has an average gradient of 11.4 m/km and flows through the historic downtown. The Time of Concentration to the measurement point at the staff gauge is 84 h.

The watershed is in the Valley and Ridge physiographic province, consisting of a series of northeast-southwest trending shale and sandstone ridges and carbonate valleys. Bedrock in the Brush Creek watershed consists of carbonates of the Knox and Honaker dolomites

(Ock and Chk, respectively) with a small ridge of the shale and dolomite Conasauga formation (Ccu) running parallel to the strike (Figure 1). Elevation ranges from 617 m in the southern ridges to 494 m at the staff gauge and the longest flow path is 9 km. The region experiences a humid continental (Köppen Cfa) climate with year-round precipitation and hot summers. Average annual precipitation is 1070 mm and annual average temperature ranges from 1.1 °C in January to 23.3 °C in July.

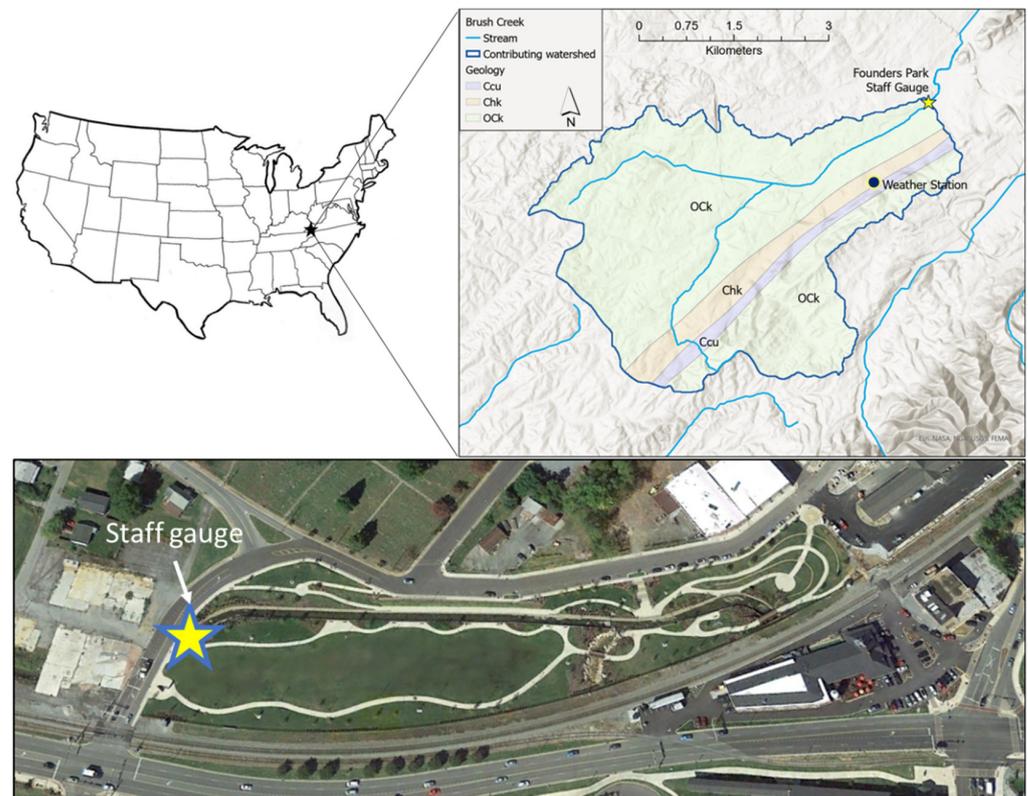


Figure 1. Contributing watershed for Brush Creek Community Science project, with location of ETSU Geosciences weather station, staff gauge, and geology, Johnson City, TN. Ock refers to the Knox Dolomite, Ohk to the Honaker dolomite, and Ccu to the shale-dolomite of the Conasauga formation.

Brush Creek has an extended history of flooding dating back over one hundred years [21]. From 1901 through 1956, sixteen flooding events were recorded, with the largest events in 1908 (56.1 m³/s) and 1938 (62.3 m³/s) [21]. These largest recorded events were several orders of magnitude greater than the typical discharge of 0.33 m³/s (per a US Geological Survey stream gage 3486500 operational from 1932–1934). There are currently no gauging sites on Brush Creek.

To address continued flooding following the 1959 report, in 2001 Johnson City set out to study the causes and solutions to flooding along Brush Creek in downtown Johnson City [22]. Water surface elevations resulting from the 10-year (0.1 annual probability) and 100-year (0.01 annual probability) 24-h storm events were modeled using the EPA Stormwater Management Model (SWMM) Version 4.4, from rainfall intensity, soils, and land use data. Hydrographs for the design storms were generated from model output, and the model was calibrated and validated using data from local rainfall events, including a 2003 flood during which a total of 119 mm of rain was measured at East Tennessee State University (ETSU), 1.9 km upstream. Various mitigation measures to address flooding were presented to the city and to residents during public meetings, and the recommended option was to develop a downtown Greenway to spur economic development and to address flooding.

In 2008, AMEC Earth & Environmental, Inc. was contracted by the City of Johnson City to complete a second downtown drainage study to assess whether a “relatively minor capital improvement project” could help to alleviate some of the flooding problems that have plagued the downtown area since the late 1800s [23]. The 2008 study confirmed the findings of the 2005 CDM study [22], concluding that a major capital outlay would be needed to provide flood relief for the downtown area. AMEC used the SWMM model developed in the 2005 CDM study to perform hydraulic simulations of water levels using 24-h rainfall totals with the goal of assessing the efficacy of various flood mitigation concept scenarios in controlling flooding in the downtown corridor. The report recommended a phased approach that combined (1) storage ponds and bypass culverts for a tributary stream, King Creek (5-year flood protection), (2) a new open channel to replace the old Brush Creek culvert (25-year flood protection), and (3) construction of three regional detention ponds upstream of the downtown area to provide protection from the 0.02 to 0.01 probability flooding event (50- to 100-year flood). The estimated cost for all phases totaled US \$25.7 million [23].

The city recently implemented phases 1 and 2 by restoring an old warehouse property. First a culvert was removed and Brush Creek, which originally flowed under the warehouse, was daylighted. The property was developed into a community space known as Founders Park. The space plays an important flood mitigation role through increased channel capacity, a large lawn for infiltration, and an amphitheater that doubles as flood storage (Figure 1). Community buy-in for efforts to control flooding is important, and a community science project was developed in collaboration with ETSU, the City of Johnson City, and a local watershed group, Boone Watershed Partnership, to increase public awareness and collect stream stage data along Brush Creek. One important component of the data collection was to make the data available in real time to avoid delays associated with physical data retrieval from a logger or autographic stage recorder. Ultimately, access to real-time stage data can provide necessary foundation data for implementation of a flood warning system.

2.2. Field Data Collection

In May 2017, a staff gauge was installed on a bridge abutment at the upstream entrance to Founders Park, and a nearby sign provides details of Brush Creek’s flooding history. Instructions describe how to read the staff gauge and submit the water level reading via text message (Figure 2). Concurrently, a Davis Vantage Pro weather station was installed on the ETSU campus within the Brush Creek watershed, 1.9 km upstream from the staff gauge. This weather station records precipitation, temperature, pressure, and wind data at 5-min intervals.

Contributed stage data observed at the staff gauge by participating community scientists are sent via text message to a Google Voice phone number set to forward to a Gmail account. Python scripts on a computer on the ETSU campus run on an automated schedule using Windows Task Manager to download the email messages. The parser uses regular expressions to extract the reported stage level from the body of the email message, and the stage, timestamp, and source phone number are written into a PostgreSQL database, populating a table of stage data. Figure 3 shows how a representative text message appears as an email when going through the Google Voice/Gmail process. The sender’s phone number and the timestamp are easy to extract from this email. The reported stage level is extracted using two regular expressions. The first regular expression shown in Step 1 in Figure 3 extracts the body of the actual text message along with some additional header/trailer text added by the Google Voice/Gmail process. The second regular expression shown in Step 2 in Figure 3 is then applied to the text extracted from the first step to find the submitted stage level. Each stage data record is assigned a unique ID.

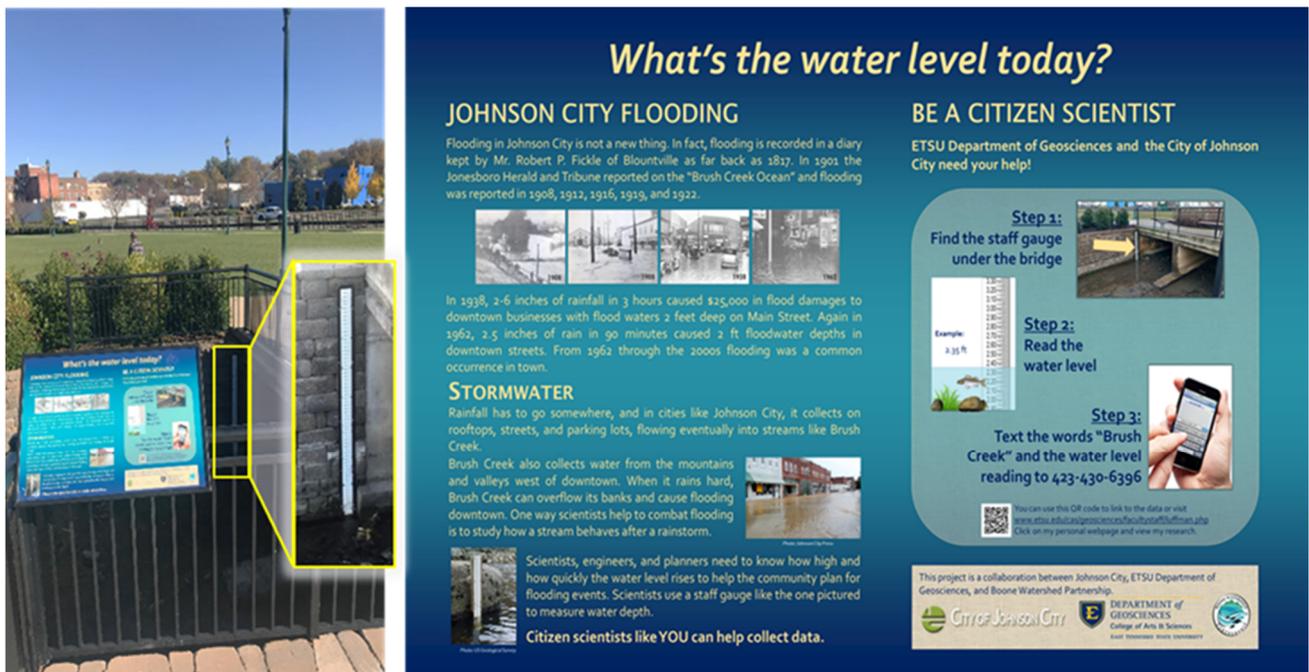


Figure 2. Educational sign provides instructions for data collection and contribution via text message.

```

SUBJECT: New text message from (423) 555-5555
Mail 1, FROM: (423) 555-5555, DATE: Thu, 06 Sep 2018 19:00:37 +0000
*****
*****
Content-Type: text/plain; charset="UTF-8"; format=flowed; delp=yes

<https://voice.google.com>
Brush Creek 99.9
YOUR ACCOUNT <https://voice.google.com> HELP CENTER
<https://support.google.com/voice/topic=1707989> HELP FORUM
<https://productforums.google.com/forum/#!forum/voice>
To edit your email preferences for text messages, go to the email
notification settings <https://voice.google.com/settings#messaging> in your
account.
Google LLC
1600 Amphitheatre Pkwy
Mountain View CA 94043 USA
*****
The Parsed Level: 99.9
                
```

Parser extracts water level from Gmail message text using two regular expressions:

1. Extract block of text between "https://voice.google.com" and "To respond|YOUR ACCOUNT"


```
p = re.compile(
(r'https://voice.google.com>
(.+)(?:To respond|YOUR ACCOUNT) ',
re.DOTALL)
```
2. Within this block, look for a pattern of digits optionally followed by a decimal point and more digits. If found, we call it the water level.


```
p2 = re.compile(r'\d*(?:\.)?\d+',
re.DOTALL)
```
3. Parsed Gmail message in Fig. 3 returns a value of 99.9 feet as the recorded stage (note these are test data).

Figure 3. Parser code and description of parsing process.

Weather data are downloaded from the ETSU weather station hourly, and Python scripts running on an automated schedule add a unique ID to each weather record. Weather data are loaded into the PostgreSQL database, populating a table of weather data. Using the timestamps, the stage data and weather data are matched, and data reports of timestamped stage and weather data are written to .csv files hourly. The complete automated workflow is shown in Figure 4.

The software tools such as Google Voice, Gmail, Python and PostgreSQL were chosen because they are readily and freely available allowing the software part of the system to be implemented at no cost, making this system cost attractive for those wishing to replicate it.

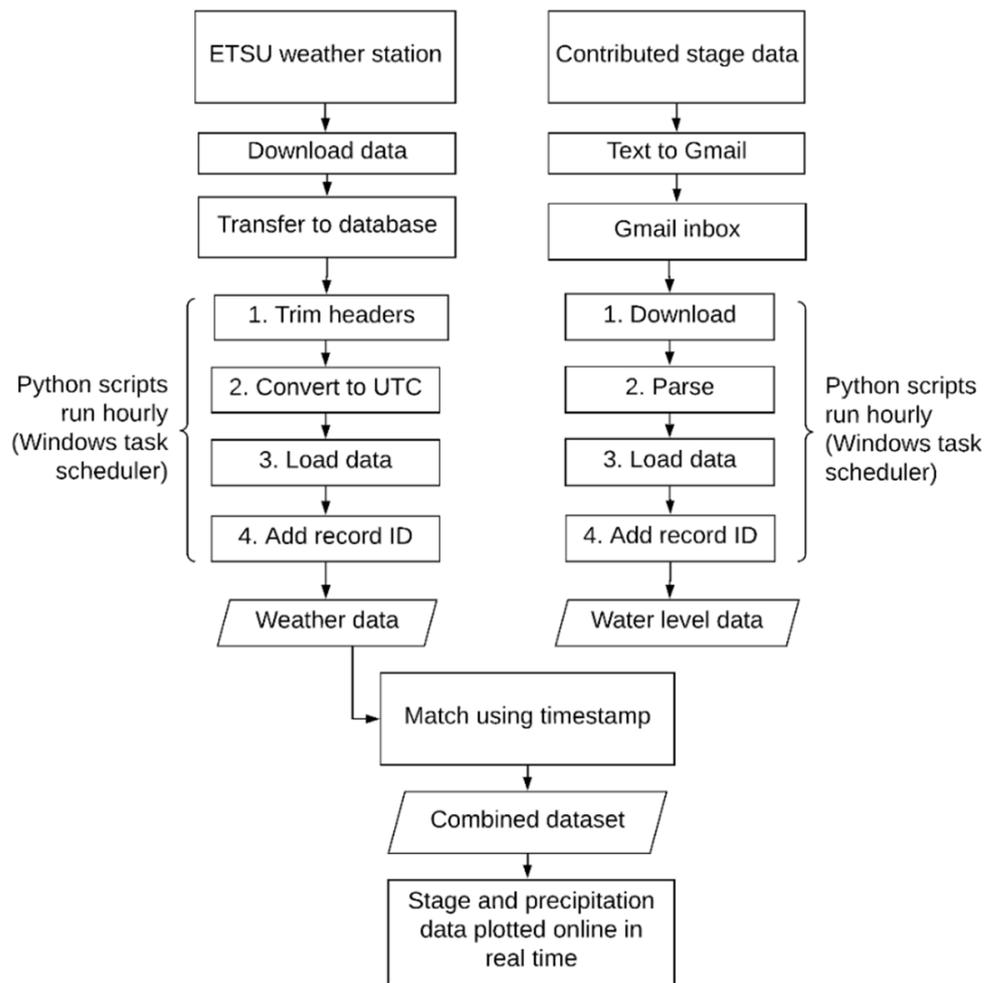


Figure 4. Automated workflow to extract stage data from text messages and load stage and weather data into database.

The Python plotting program Pyplot was used to plot preliminary stage and precipitation data from the hourly reports, generating five plots in real time and reflecting different time frames: all data, past year, past month, past week, and past 24 h. Prior to plotting, contributed stage data outside the range of 0.15 m to 2.1 m are excluded as these values are below baseflow and above the length of the staff gauge, respectively. Prior to formal data analyses, contributed data are reviewed and corrected. Recognizing that data accessibility is paramount to promoting and sustaining a community science project [12], the preliminary plots are updated and uploaded hourly to the ETSU Geosciences ArcGIS server. They are available publicly online at <https://iluffman.wixsite.com/ingrid-luffman/research> (accessed on 15 December 2021). A sample plot showing data from 2017-09-11 to 2021-12-15 is displayed in Figure 5.

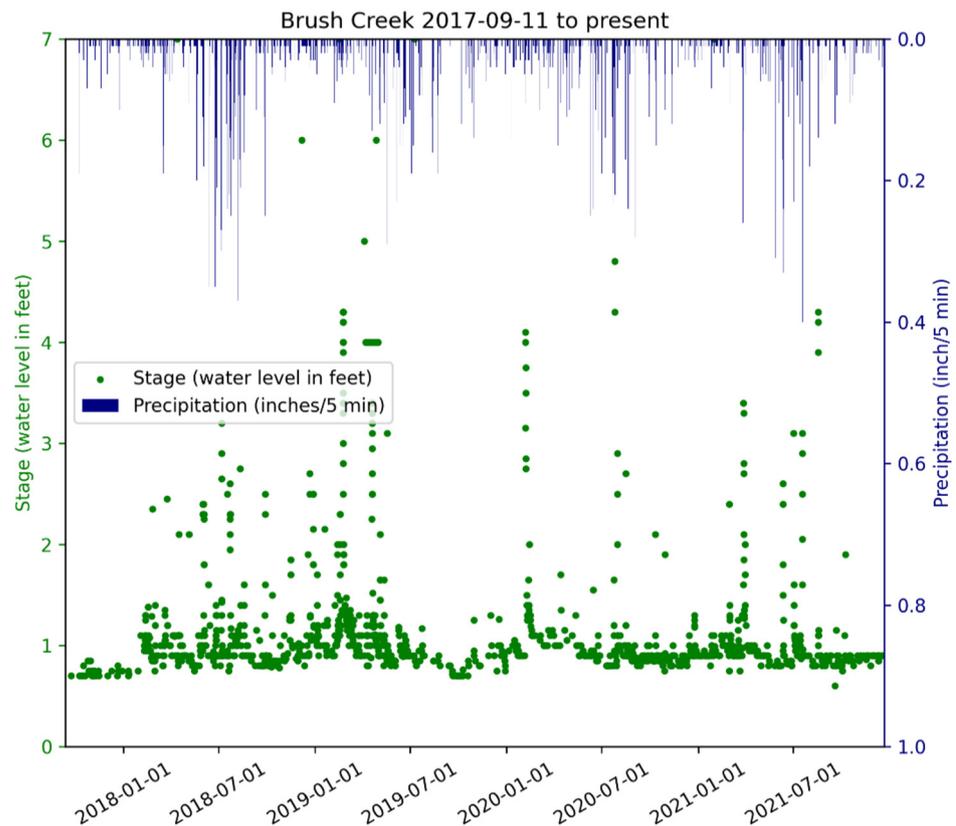


Figure 5. Example of automated plot of stage and precipitation data showing all contributed data. All plots are updated hourly and freely available online. Units are in US format (inches/feet) in all data plots intended for public use.

2.3. Assessment of Contributed Community Science Data

To assess parser performance and data quality, stage data contributed from 2017-05-17 through 2021-07-31 were compared to the parsed data and classified into three groups: (1) data were correctly parsed, (2) false positive (incorrect data extracted), and (3) false negative (parser unable to extract data even though usable data were reported) (Figure 6). Submissions from individual participants were tallied to identify top data contributors and the frequency distribution of number of observations per participant was calculated.

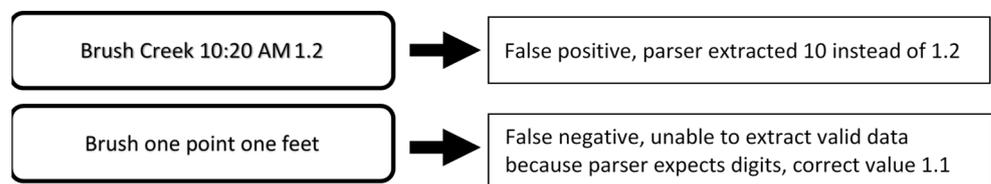


Figure 6. Examples of parser failures (false positive and false negative).

To statistically correlate precipitation accumulation to streamflow as a rainfall–runoff response, a time series dataset of both rainfall and runoff (here, using stage as a proxy) is required. The frequency of contributed stage data was assessed by cross-tabulating by day of week and time of day to identify peak days and times for community participation.

2.4. Rainfall–Runoff Response

We calculated antecedent precipitation for lags of 15 min, 30 min, 1 h, 2 h, 3 h, 6 h, 12 h, 1 day, 3 days, and 7 days prior to each stage observation. The Spearman correlation coefficient between antecedent precipitation and stage was calculated for each stage measurement. Ordinary Least Squares (OLS) regression models were developed

for stage using antecedent precipitation at various lags to identify the timing of rainfall-associated runoff in Brush Creek.

3. Results

3.1. Assessment of Contributed Data

From 2017-05-17 through 2021-07-31, we received 1496 contributed stage measurements (observations) from 597 different participants. Of the nearly 1500 text messages received, the parser extracted correct data for 1430 (96.7%) messages. Only 3.3% of text messages were parser failures; false positives accounted for 3.1% of messages while false negatives accounted for 0.1%.

One challenge in relying on community scientists for data collection is that data are contributed only when participants are present, predominantly when weather is fair and during normal waking hours. Rainfall occurs at any time of day or night, yet participants contribute more during daily peak times between 7–9 a.m., 1–3 p.m., and 6–10 p.m. (Figure 7). Only 27 observations (1.8%) were received between 11 p.m. and 7 a.m., a pattern that is consistent throughout week and weekend nights (Figure 8). Friday and Saturday evenings and Sunday afternoons were the most frequent data contribution times during the week. Cumulative data were compared at three timesteps during the study period (inception on 2017-05-17 to 2019-01-24, to 2019-10-08, and to 2021-07-31); the weekly and daily patterns are similar for each (Figure 8). Further, while stream stage data should be collected during all weather conditions, stream stage during rainy and stormy weather was not well represented in the contributed data. Only 7% of observations were made when rain was recorded in the prior 15 min and 10% of observations were made with rain in the prior 30 min.

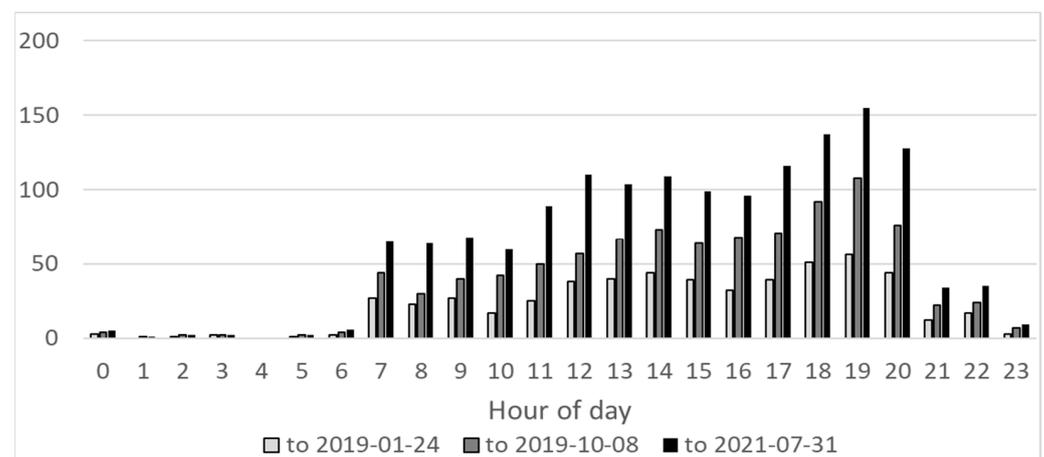


Figure 7. Frequency distribution of stream stage observations by hour of day shows limited activity after 11 p.m. and prior to 7 a.m. Patterns are consistent over separate cumulative time periods.

Most participants (445 participants, or 75%) contributed stage observations a single time during the study period, while eight participants (1%) were considered regulars by contributing more than 20 observations. Only one “champion”, with contributions in excess of 100 observations, was identified (Figure 9). The distribution of contributions by participant closely followed the pattern observed in the CrowdHydrology project [20].

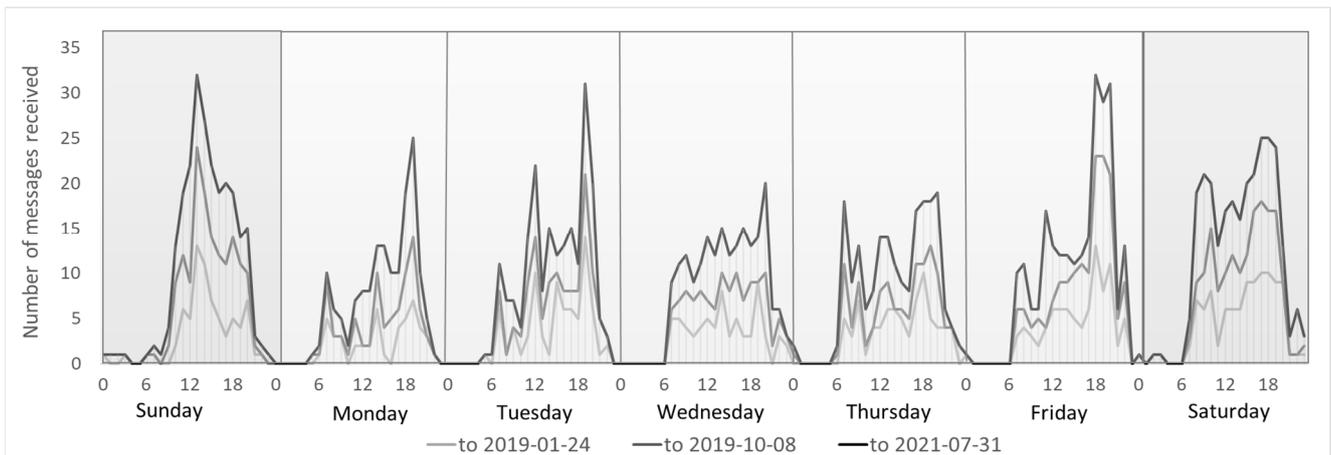


Figure 8. Frequency distribution of incoming messages by hour of day and day of week shows peak activity on Friday and Saturday evenings and Sunday afternoons. Patterns are consistent over separate cumulative time periods.

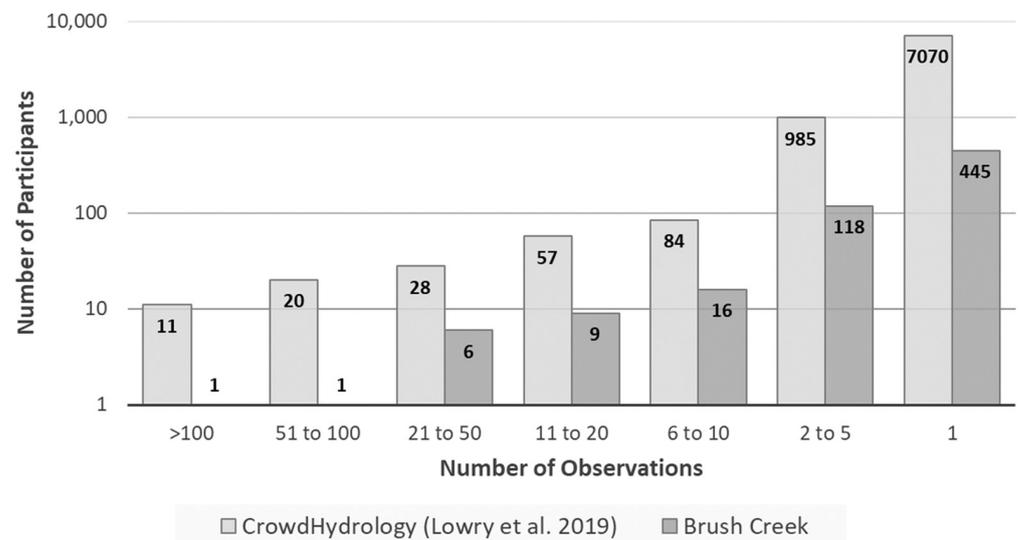


Figure 9. Distribution of frequency of observations by individual participant for current research (Brush Creek) compared to findings of Lowry et al. (2019).

3.2. Analysis of Precipitation-Stage Relationships

Contributed stage data were matched to precipitation data by calculating the depth of precipitation (in mm) received during various antecedent precipitation periods ranging from 15 min to 7 days. The Spearman correlation coefficients between stage and precipitation were highest within 6 h of the stage observation (Table 1). As expected, high correlation existed between the various antecedent precipitation periods.

The OLS stepwise regression model for stream stage using antecedent precipitation depth as the independent variables explained 21.6% of the variability in stage data ($R^2 = 0.216$). Retained variables are listed in Table 2. No multicollinearity problems were indicated based on reviewing the Variance Inflation Factor. Examination of standardized coefficients shows that precipitation in the preceding 15 min has the highest influence on stream stage, followed by precipitation in the prior 2 and 24 h. A small negative influence on stage was noted for antecedent precipitation in the prior 6 h and three days.

Table 1. Spearman correlation of precipitation and stage. All coefficients are significant at $\alpha = 0.01$.

	Stage	15_min	30_min	60_min	2_h	3_h	4_h	5_h	6_h	12_h	24_h	3_Days	7_Days
Stage	1												
15_min	0.254	1											
30_min	0.277	0.867	1										
60_min	0.279	0.78	0.9	1									
2_h	0.313	0.668	0.771	0.856	1								
3_h	0.297	0.598	0.69	0.766	0.881	1							
4_h	0.3	0.541	0.624	0.692	0.798	0.882	1						
5_h	0.292	0.512	0.591	0.655	0.755	0.839	0.909	1					
6_h	0.275	0.489	0.565	0.626	0.72	0.802	0.872	0.905	1				
12_h	0.239	0.42	0.484	0.536	0.614	0.681	0.742	0.771	0.795	1			
24_h	0.214	0.349	0.392	0.433	0.487	0.538	0.584	0.608	0.629	0.741	1		
3_days	0.112	0.243	0.263	0.283	0.299	0.321	0.345	0.359	0.373	0.448	0.569	1	
7_days	0.09	0.197	0.194	0.192	0.187	0.19	0.204	0.207	0.211	0.254	0.328	0.545	1

Table 2. OLS regression results for stream stage (dependent variable) with antecedent precipitation depth show that stage is most influenced by rainfall within the prior 15 min.

	B	Std. Error	Standardized Coefficients	t	Sig.	Tolerance	VIF
Constant	1.035	0.019	-	55.585	0	-	-
24_h	0.006	0.001	0.194	5.53	0	0.459	2.178
15_min	0.257	0.033	0.274	7.755	0	0.452	2.214
3_days	-0.002	0.001	-0.106	-3.035	0.002	0.461	2.17
7_days	0.001	0	0.07	2.242	0.025	0.583	1.717
2_h	0.039	0.009	0.195	4.417	0	0.289	3.465
6_h	-0.008	0.003	-0.103	-2.658	0.008	0.378	2.649

4. Discussion

4.1. Assessment of Contributed Data

Successful data collection by community scientists requires context and clear unambiguous instructions. Junk data, missing data, and use of Emojis can result in false positives (wrong data extracted) and false negatives (parser unable to extract data even though usable data were reported). In this project, community scientists were encouraged to submit stream stage observations via text messaging. Simple illustrated instructions for contributing readings were provided on the educational sign (Figure 2). When participants followed instructions for submitting stream stage data, the parser successfully parsed their texts. However, community scientists did not always follow instructions and embellished their submissions with additional text, time and date stamps, emojis and other non-Unicode symbols that can cause the regular expression-based parser to fail to properly extract the reported stream stage. Failures consisted of false positive results where the parser reported the incorrect stage and false negatives where the parser was unable to extract the stage observation from the text message. Despite the 96.7% parser success rate over the course of the project, we seek to reduce parser failures. Manual analysis of the submitted texts has identified some patterns of user submissions that lead to parser errors.

First, some community scientists entered the date or time of the observation followed by the stream stage data. The simple parser extracts the first numbers it encounters in the text message as the stream's stage, anticipating that participants will follow the instructions provided. To address this, we recommend that the regular expression be extended to disregard numbers that look like dates or time stamps when extracting the stream stage data; numbers containing colons, dashes, slashes or a.m. or p.m. indicators could be ignored.

Second, some participants spelled out their submission, for example "one point four." Some gave ranges, "between 1.4 and 1.5" or relative values, "a bit under 2". Creating

regular expressions to handle these latter examples is very difficult. More sophisticated natural language processing techniques would be needed to properly parse these examples.

4.2. Improving Participation

Community scientists have been encouraged to participate in this project through public outreach. With the support of Johnson City, signage was installed to publicize the project and to instruct the community on how to participate. Undergraduate students in general science education and geosciences courses at ETSU were introduced to the project and encouraged to submit data. Educational field trips have been conducted for local youth (including the Girl Scouts of America), environmentally focused organizations, and conference attendees and other academics.

Personal conversations with community members revealed that potential participants were skeptical that their stage observations were, in fact, recorded and used for research. To encourage repeat participation, in summer 2020, a “Thank you” text message was instituted. Now, within 15 min of submitting an observation, participants receive a response text thanking them for submitting data to the project. Following implementation of the “Thank you” text, participation rate doubled from 13.75 observations per month (July 2019–June 2020) to 26.88 observations per month (July 2020–June 2021).

Automated real-time plots available online and updated hourly were another effort to improve participation through engagement. Within one hour of contributing an observation, participants can view their contributed data on the data plots.

Other suggestions for improving participation focused on generating interest through rewards or badges for regular participants [12], or by installing hands-on devices like a bicycle instrumented to serve as a generator that passersby could ride to power nearby spotlights to aid in nighttime observations, or to power internet-of-things projects.

One advantage of our methodology is the use of simple instructions (read the gauge, send a text), a widespread communication utility (text messaging), and freely available tools to handle observations (Google Voice and Gmail) and generate reports and plots (PostgreSQL and Python). This methodology is replicable in other parts of the watershed and in other communities. The sole portion of the process that uses proprietary software is our use of the ArcGIS server at ETSU to store images of the stage/precipitation plots (Figure 4). We used this server as it was readily available, however, another server could be used.

Another advantage of our methodology is how data are transferred. The system we implemented permits the host computer to request data from Google Voice and Gmail and pull it to the host computer, process it, and push it back out to the cloud in the form of data plots. This eliminates security risks associated with pushing data from the cloud to the host computer.

4.3. Assessing Rainfall–Runoff Response

A limitation of this study is the gaps in contributed stage data that fail to capture peak flows. Nevertheless, data are sufficient to identify patterns of recession following high flows and normal base flows (Figure 5). Occasionally, high stage values were reported by community scientists without any recorded precipitation at the ETSU rain gauge, for example, in February of 2020 (Figure 5). This is likely the result of rainfall elsewhere in the watershed. Given the karst geology of the region, springs may also contribute water from outside the watershed, but within the springshed for Brush Creek. We recommend that data from additional rain gauges be located and included in future analyses.

Because of the sporadic timing of contributed stage data, we selected the approach of using the antecedent precipitation associated with each stage observation to evaluate the rainfall–runoff relationship. Regression models identified that precipitation in the preceding 15 min had the greatest influence on stream stage, indicating a tendency toward flashy behavior. The negative coefficient for 6-h antecedent precipitation likely reflects a quick drop in stage following peak flow associated with a rain event. Field observation of

Brush Creek discharge during and immediately following precipitation supports the “quick recovery” explanation. Further, the negative coefficient for 3-day antecedent precipitation is consistent with the Time of Concentration of 84 h (approximately 3.5 days), suggesting that all runoff from a storm three days prior had moved through the system by 84-h post-storm.

4.4. Future Work

We recognize that contributed stage data are sparse and not uniform in time. Further, stage data for some major precipitation events may go unrecorded due to lack of reporting during times of inclement weather or overnight. Increased participation is not likely to improve this temporal pattern. We also recognize that contributed data have not been validated for accuracy. We therefore recommend installation of automated sensors to fill temporal data gaps and provide a dataset that can validate contributed stage observations. To address this limitation, a team of engineering students at ETSU is building an internet-connected ultrasonic sensor system to report the stream stage data at periodic regular intervals, such as every 5 min, 24 h a day. Data reported by this system may be used to validate the sparsely submitted community scientist observations. A secondary use of data reported by this system is refinement of current regression models, and development of other rainfall–runoff models in which sensor data may be used for model calibration and validation.

5. Conclusions

This research describes patterns of stream stage data contributed by community scientists, and development of regression models for stream stage using antecedent precipitation. Stage observations were contributed overwhelmingly during waking hours (7 a.m. to 11 p.m., 98% of observations) and during periods of fair weather (only 7% received within 15 min of precipitation). Ordinary least squares regression models for stream stage using antecedent precipitation explained 21.6% of the variability in stage data, with the 15-min antecedent precipitation variable having the most influence on stage, indicating flashy behavior of the monitored stream. We conclude that community scientist observations to date are insufficient to develop predictive models for water flow due to data gaps. To validate contributed data and fill data gaps, an automated sensor system is under development to provide continuous measurement of stage.

Author Contributions: Conceptualization, I.L. and D.C.; methodology, I.L. and D.C.; software, D.C.; validation, I.L.; formal analysis, I.L.; resources, I.L.; data curation, I.L.; writing—original draft preparation, I.L. and D.C.; writing—review and editing, I.L. and D.C.; visualization, D.C.; supervision, I.L.; project administration, I.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not yet publicly available due to ongoing and planned future data analyses.

Acknowledgments: We appreciate assistance from Chelsie Perkins, Timothy Muncy, and Alex Mann for contributions to coding and data processing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Buytaert, W.; Zulkafli, Z.; Grainger, S.; Acosta, L.; Alemie, C.T.; Bastiaensen, J.; De Bievre, B.; Bhusal, J.; Clark, J.; Dewulf, A.; et al. Citizen science in hydrology and water resources: Opportunities for knowledge generation, ecosystem service management, and sustainable development. *Front. Earth Sci.* **2014**, *2*, 26. [[CrossRef](#)]
2. Kosmala, M.; Wiggins, A.; Swanson, A.; Simmons, B. Assessing data quality in citizen science. *Front. Ecol. Environ.* **2016**, *14*, 551–560. [[CrossRef](#)]
3. Dickinson, J.L.; Shirk, J.; Bonter, D.; Bonney, R.; Crain, R.L.; Martin, J.; Phillips, T.; Purcell, K. The current state of citizen science as a tool for ecological research and public engagement. *Front. Ecol. Environ.* **2012**, *10*, 291–297. [[CrossRef](#)]
4. Conrad, C.C.; Hilchey, K.G. A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environ. Monit. Assess.* **2011**, 273–291. [[CrossRef](#)] [[PubMed](#)]
5. Bowser, A.; Shanley, L. New visions in citizen science. *Case Study Ser.* **2013**, *3*, 1–53.
6. Lowry, C.S.; Fienen, M.N. CrowdHydrology: Crowdsourcing hydrologic data and engaging citizen scientists. *GroundWater* **2013**, *51*, 151–156. [[CrossRef](#)] [[PubMed](#)]
7. Seibert, J.; Strobl, B.; Etter, S.; Hummer, P.; van Meerveld, H.J.I. Virtual staff gauges for crowd-based stream level observations. *Front. Earth Sci.* **2019**, *7*, 70. [[CrossRef](#)]
8. Etter, S.; Strobl, B.; van Meerveld, I.; Seibert, J. Quality and timing of crowd-based water level class observations. *Hydrol. Processes* **2020**, *34*, 4365–4378. [[CrossRef](#)]
9. Royem, A.A.; Mui, C.K.; Fuka, D.R.; Walter, M.T. Proposing a low-tech, affordable, accurate stream stage monitoring system. *Trans. ASABE* **2012**, *55*, 2237–2242. [[CrossRef](#)]
10. Lane, S.N.; Odoni, N.; Landstrom, C.; Whatmore, S.J.; Ward, N.; Bradley, S. Doing flood risk science differently: An experiment in radical scientific method. *Trans. Inst. Br. Geogr.* **2011**, *36*, 15–36. [[CrossRef](#)]
11. Chanse, V.; Mohamed, A.; Wilson, S.; Dalemarré, L.; Leisnham, P.T.; Rockler, A.; Shirmohammadi, A.; Montas, H. New approaches to facilitate learning from youth: Exploring the use of Photovoice in identifying local watershed issues. *J. Environ. Educ.* **2016**, *48*, 109–120. [[CrossRef](#)]
12. Lee, K.A.; Lee, J.R.; Bell, P. A review of Citizen Science within the Earth Sciences: Potential benefits and obstacles. *Proc. Geol. Assoc.* **2020**, *131*, 605–617. [[CrossRef](#)]
13. Elwood, S.; Goodchild, M.F.; Sui, D.Z. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 571–590. [[CrossRef](#)]
14. Wu, D.; Del Rosario, E.A.; Lowry, C. Exploring the Use of Decision Tree Methodology in Hydrology Using Crowdsourced Data. *J. Am. Water Resour. Assoc.* **2021**, *57*, 256–266. [[CrossRef](#)]
15. Wilson, J.S.; Pan, A.D.; General, D.E.M.; Koch, J.B. More eyes on the prize: An observation of a very rare, threatened species of Philippine Bumble bee, *Bombus irisanensis*, on iNaturalist and the importance of citizen science in conservation biology. *J. Insect Conserv.* **2020**, *24*, 727–729. [[CrossRef](#)]
16. Aristeidou, M.; Herodotou, C.; Ballard, H.L.; Higgins, L.; Johnson, R.F.; Miller, A.E.; Young, A.N.; Robinson, L.D. How do young community and citizen science volunteers support scientific research on biodiversity? The case of iNaturalist. *Diversity* **2021**, *13*, 318. [[CrossRef](#)]
17. Lamoureux, Z.; Fast, V. The tools of citizen science: An evaluation of map-based crowdsourcing platforms. *CEUR Workshop Proc.* **2019**, 2323, 12.
18. Chmielewski, S.; Samulowska, M.; Lupa, M.; Lee, D.; Zagajewski, B. Citizen science and WebGIS for outdoor advertisement visual pollution assessment. *Comput. Environ. Urban Syst.* **2018**, *67*, 97–109. [[CrossRef](#)]
19. Juang, C.S.; Stanley, T.A.; Kirschbaum, D.B. Using citizen science to expand the global map of landslides: Introducing the cooperative open online landslide repository (COOLR). *PLoS ONE* **2019**, *14*, e0218657. [[CrossRef](#)] [[PubMed](#)]
20. Lowry, C.S.; Fienen, M.N.; Hall, D.M.; Stepenuck, K.F. Growing pains of crowdsourced stream stage monitoring using mobile phones: The development of crowdhydrology. *Front. Earth Sci.* **2019**, *7*, 128. [[CrossRef](#)]
21. Tennessee Valley Authority. *Floods on Brush Creek in Vicinity of Johnson City Tennessee*; Report No. 0-5868; Tennessee Valley Authority: Knoxville, TN, USA, 1959.
22. Lamar Dunn & Associates. *Downtown Drainage Study, City of Johnson City, Tennessee*; Lamar Dunn & Associates: Johnson City, TN, USA, 2005.
23. AMEC Earth & Environmental. *Downtown Drainage Study, Johnson City, Tennessee*; AMEC Earth & Environmental: Knoxville, TN, USA, 2008.