A Temporal Gauge Quality Control Algorithm as a Method for Identifying Potential Instrumentation Malfunctions

STEVEN M. MARTINAITIS,[a,b] SCOTT LINCOLN,[c] DAVID SCHLOTZHAUER,[d] STEPHEN B. COCKS,[a,b] AND JIAN ZHANG[b]

[a] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*
[b] *NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*
[c] *National Weather Service Forecast Office, Chicago, Illinois*
[d] *National Weather Service/Lower Mississippi River Forecast Center, Slidell, Louisiana*

ABSTRACT: There are multiple reasons as to why a precipitation gauge would report erroneous observations. Systematic errors relating to the measuring apparatus or resulting from observational limitations due to environmental factors (e.g., wind-induced undercatch or wetting losses) can be quantified and potentially corrected within a gauge dataset. Other challenges can arise from instrumentation malfunctions, such as clogging, poor siting, and software issues. Instrumentation malfunctions are challenging to quantify as most gauge quality control (QC) schemes focus on the current observation and not on whether the gauge has an inherent issue that would likely require maintenance of the gauge. This study focuses on the development of a temporal QC scheme to identify the likelihood of an instrumentation malfunction through the examination of hourly gauge observations and associated QC designations. The analyzed gauge performance resulted in a temporal QC classification using one of three categories: GOOD, SUSP, and BAD. The temporal QC scheme also accounts for and provides an additional designation when a significant percentage of gauge observations and associated hourly QC were influenced by meteorological factors (e.g., the inability to properly measure winter precipitation). Findings showed a consistent percentage of gauges that were classified as BAD through the running 7-day (2.9%) and 30-day (4.4%) analyses. Verification of select gauges demonstrated how the temporal QC algorithm captured different forms of instrumental-based systematic errors that influenced gauge observations. Results from this study can benefit the identification of degraded performance at gauge sites prior to scheduled routine maintenance.

SIGNIFICANCE STATEMENT: This study proposes a scheme that quality controls rain gauges based on its performance over a running history of hourly observational data and quality control flags to identify gauges that likely have an instrumentation malfunction. Findings from this study show the potential of identifying gauges that are impacted by issues such as clogging, software errors, and poor gauge siting. This study also highlights the challenges of distinguishing between erroneous gauge observations based on an instrumentation malfunction versus erroneous observations that were the result of an environmental factor that influence the gauge observation or its quality control classification, such as winter precipitation or virga.

KEYWORDS: Data quality control; Gauges; Quality assurance/control; Instrumentation/sensors

1. Introduction

Numerous hydrometeorological applications depend on observations recorded by surface precipitation gauges. Real-time dissemination of gauge observations contributes to the information needed in the decision-making processes for various flood-type operations (NOAA 2019; Laber 2020). Precipitation gauges are routinely used for the evaluation and verification of gridded quantitative precipitation estimation (QPE) techniques derived from weather radars (Tabary et al. 2007; Gourley et al. 2009; Ryzhkov et al. 2014; Cocks et al. 2019; Wijayarathne et al. 2020) and satellites (Derin et al. 2016; Manz et al. 2017; Gowan and Horel 2020); moreover, gauge observations are employed in bias correction methodologies for gridded QPEs (Brandes 1975; Steiner et al. 1999; Seo and Breidenbach 2002; Ushio et al. 2013; Rabiei and Haberlandt

2015; Zhang et al. 2016). Multidecadal observational records benefit precipitation climatologies (Hulme and New 1997; Chen et al. 2002; Carvalho 2020) and derived diagnostic products, such as average recurrence intervals (Perica et al. 2013).

The wide-ranging set of applications for precipitation gauges demonstrate how crucial it is to ensure recorded observations are free of inaccuracies. There are various factors that can result in erroneous gauge observations, and these factors can be generally sorted into two different categories: systematic errors and instrumentation malfunctions. A systematic error is defined as a consistent, nonrandom inaccuracy inherent in equipment or an approach. Systematic errors related to gauge observations focus on the limitations of the gauge instrumentation design or the challenges with obtaining accurate liquid accumulations based on environmental factors.

Previous studies have assessed various systematic inaccuracies and observational limitations with precipitation gauges. Systematic errors relating to the measuring apparatus are dependent on the instrumentation type. Tipping-bucket gauges are subject to splash-out and tip counting errors (Parsons 1941;

Upton and Rahimi 2003; Molini et al. 2005) as well as sampling errors at small temporal scales (Habib et al. 2001; Boudala et al. 2017). The time to conduct the tipping process can result in precipitation not being measured (Duchon and Essenberg 2001; Duchon and Biddle 2010). Observations from weighing gauges can be influenced by frictional drag (Groisman and Legates 1994), oscillatory motions (Goodison et al. 1998), and load sensor noise (Duchon 2008; Leeper et al. 2015). Systematic errors resulting from environmental factors can also influence the accuracy of the recorded observation. Wind-induced undercatch can vary significantly based on the precipitation type and use of a wind shield (Allerup and Madsen 1980; Sevruk 1989; Førland et al. 1996; Yang et al. 1998; Nešpor and Sevruk 1999; Chubb et al. 2015; Kochendorfer et al. 2017; Pollock et al. 2018). Evaporation or sublimation via wetting losses can occur from moisture adhering to the internal walls (Groisman and Legates 1994; Goodison et al. 1998; Savina et al. 2012) or from the application of heating mechanisms (Metcalfe and Goodison 1992; Larson 1993). Blockages from winter precipitation (Goodison et al. 1998; Rasmussen et al. 2012; Martinaitis et al. 2015) and subsequent postevent thawing (Martinaitis et al. 2015) can also provide erroneous observations both during and after winter precipitation events.

Systematic errors are generally quantifiable; moreover, the ability to quantify systematic errors allow for corrective equations and collection efficiency curves to compensate for the systematic biases in observations. This can be conducted through gauge intercomparisons, laboratory calibrations, particle and turbulence modeling, and computational fluid dynamic simulations (e.g., Lanza and Cauteruccio 2022). Past research demonstrated the successful quantification and correction of both mechanical and environmental systematic errors, including that for tipping mechanisms (e.g., Maksimović et al. 1991; Sevruk 1996; Duchon et al. 2014; Sypka 2019; Liao et al. 2020), wind-induced undercatch (e.g., Sevruk 1982; Førland et al. 1996; Yang et al. 1998; Kochendorfer et al. 2017; Pollock et al. 2018), and wetting and evaporative losses (e.g., Sevruk 1982; Yang et al. 1998; Ye et al. 2004).

Gauge-based impacts related to instrumentation malfunctions fall outside the realm of systematic errors and can also result in significant errors with gauge observations. Data collection challenges can arise from biological interference and other material that can clog a gauge cylinder or orifice (e.g., Nystuen et al. 1996; Nystuen 1998; Steiner et al. 1999). Other challenges that can influence a gauge observation can include poor gauge siting that can result in interference with the measuring process (Steiner et al. 1999; Upton and Rahimi 2003). Malfunctions with the gauge instrumentation can result in improper calibration of the measuring apparatus, damage to the instrumentation, power outages, and data software and transmission issues (e.g., Groisman and Legates 1994; Kondragunta and Shrestha 2006; Sieck et al. 2007). Metcalfe et al. (1997) observed that errors from instrumentation malfunctions can be significant and "virtually impossible to quantify," yet the need exists to improve the identification of gauge sites impacted by instrumental factors.

Gauge quality control (QC) methodologies can easily identify and flag an erroneous gauge observation, yet QC schemes generally focus on the current observation and generally do not provide an indication of the likelihood of an inherent nonsystematic issue with the gauge. Characterizing the overall quality of a gauge should be investigated based upon how a gauge performs over an extended period. Forecasters at the National Weather Service (NWS) Lower Mississippi River Forecast Center (LMRFC) noted that a more formalized structure was needed to identify gauge sites that were consistently flagged as unreliable for operational precipitation product generation. Real-time gauge assessments at the LMRFC focused on nonzero observations ≥ 6.35 mm (0.25 in.) using a three-tiered classification based on gauge-to-QPE ratio values for different temporal accumulations (Lincoln et al. 2017). Forecasters can then graphically view the history of a gauge across a user-specified period and determine if the gauge should be stricken from further operational use. Collaborative efforts between the LMRFC and the National Severe Storms Laboratory (NSSL) sought to implement a similar practice across the NWS to better identify a gauge site experiencing an instrumentation malfunction.

This study describes the development of a best-guess automated identification and classification scheme for instrumentation malfunctions with rain gauges. Determination of gauge errors from instrumentation malfunctions in this study relied on the evaluation of the cumulative QC declarations of hourly gauge observations over defined temporal periods. Potential impacts from instrumentation issues that yielded erroneous observations were assessed through the statistical analyses of gauge observations and their corresponding QC designations. The results from this study can provide the foundation for a situational awareness methodology to identify gauge sites that are likely in need of maintenance.

2. Temporal QC logic

a. Application of MRMS gauge QC

The first step of the temporal QC scheme was processing hourly gauge observations through the Multi-Radar Multi-Sensor (MRMS) system (Zhang et al. 2016) gauge QC algorithm (Martinaitis et al. 2021). The MRMS gauge QC logic compares hourly gauge observations to 1) gridded MRMS dual-polarization synthetic radar QPEs (Cocks et al. 2019; Wang et al. 2019; Zhang et al. 2020a) with an evaporation correction scheme (Martinaitis et al. 2018) and/or 2) quantitative precipitation forecasts (QPFs) from the High-Resolution Rapid Refresh (HRRR; Benjamin et al. 2016) model to determine if a gauge observation was erroneous. How a gauge is analyzed in the MRMS QC scheme is based on a proxy for radar coverage via the hourly Radar Accumulated Quality Index (RAQI) product (Zhang et al. 2020b). The RAQI product is the radar quality index value, which is derived from radar beam height, the radar beam location in relation to the melting layer, and radar beam blockage (e.g., Martinaitis et al. 2018) averaged over a given period. One notable change was made to the hourly gauge QC logic for this study. The time window check to indicate when an observation time was significantly offset from the top of the hour was turned off

TABLE 1. Description of the hourly MRMS gauge QC flags utilized in this study, per Martinaitis et al. (2021).

MRMS QC flag	Gauge QC description	Retain in MRMS
−10	Conditional pass—Zero observation (with sources present)	Yes
−11	Conditional pass—Zero observation (with sources missing)	Yes
−12	Conditional pass—Nonzero observation (with sources present)	Yes
−13	Conditional pass—Nonzero observation (with sources missing)	Yes
0	Pass—Zero value	Yes
1	Pass—Nonzero value	Yes
10	False zero observation (rain only)	No
20	False nonzero observation (rain only)	No
30	Outlier observation—High value (rain only)	No
31	Outlier observation—High value (snow only)	No
40	Outlier observation—Low value (rain only)	No
50	Winter impacts—Zero observation	No
51	Winter impacts—Nonzero observation	No
60	Suspect observation—Nonzero observation (with sources present)	No
61	Suspect observation—Nonzero observation (with sources missing)	No

(i.e., all gauges were analyzed regardless of the valid time of the observation). The resulting MRMS gauge QC process would assign one of 15 flag classifications to each hourly gauge observation (Table 1).

Multiple considerations attributed to the use of the MRMS gauge QC scheme as the foundational input into the temporal QC classification. The application of radar-derived QPEs in the QC of gauge observations have shown to benefit the identification of false zero or false nonzero observations (e.g., Båserud et al. 2020; Ośródka et al. 2022) and values that are significant outliers (e.g., Yeung et al. 2014; Båserud et al. 2020). The MRMS radar-derived QPE employed in the QC of gauges undergoes its own QC assurances, including clutter mitigation, removal of nonmeteorological contaminants, reduced impacts from beam blockage, and mitigation of bright-band influences on reflectivity values (Zhang and Qi 2010; Zhang et al. 2012; Qi et al. 2013a,b; Tang et al. 2014; Zhang et al. 2016; Tang et al. 2020). The MRMS gauge QC approach also includes model variables to assist in 1) identifying false zero and false nonzero observations within regions of substandard radar coverage and 2) delineating impacts on gauges based on winter precipitation impacts (Martinaitis et al. 2015, 2021). The MRMS gauge QC scheme has a comprehensive classification for declaring why a gauge passed or failed the QC algorithm (Table 1) and its performance compared favorably to manual forecaster QC (Martinaitis et al. 2021).

b. Initial hourly temporal QC scheme designation

Assigned hourly MRMS gauge QC classifications were the foundation for the initial temporal QC classifications. One of three initial classifications were assigned to each hourly observation:

- PASS: The hourly gauge observation passed MRMS QC and is considered of good quality.
- FAIL: The hourly gauge observation failed MRMS QC and is considered of poor quality.
- WXCN: The hourly gauge observation failed the MRMS QC but was potentially influenced by some weather-based condition.

All hourly observations that were passed or conditionally passed by the MRMS gauge QC (i.e., retained for use in the MRMS system) were declared as PASS. All hourly observations that were failed by the MRMS gauge QC were further examined to see if the QC was likely influenced by environmental factors.

Gauge observations marked as false zeroes (MRMS gauge QC flag = 10) were reexamined based on the collocated nonzero gridded radar QPE (Fig. 1). Collocated radar QPE ≤ 1.27 mm had a higher probability of the QC being influenced by precipitation being evaporated before reaching the ground (Martinaitis et al. 2018); thus, the hourly temporal QC classification was set to WXCN. The hourly temporal QC classification was set to FAIL with collocated radar QPE values > 1.27 mm. Gauge observations declared as a false nonzero observation (MRMS gauge QC flag = 20) were compared to model surface wet-bulb temperatures (T_{wb}). The surface T_{wb} can account for environments with above-freezing ambient temperatures and nonsaturated relative humidity values that could sustain solid winter precipitation (Matsuo and Sasyo 1981). Declared false nonzero observations in model surface $T_{wb} \leq 5.0°$C environments had a greater likelihood of being influenced by postwinter precipitation thawing and were classified as WXCN, whereas the FAIL classification was used in environments characterized by $T_{wb} > 5.0°$C. A lower bound for the surface T_{wb} comparison was not designated in this study to account for gauge sites equipped with a heating element that could produce nonzero values from melting winter precipitation in subzero T_{wb} conditions. MRMS gauge QC flags characterized with winter weather impacts (QC flags = 50, 51) account for gauge sites that likely have the gauge orifice partially or completely blocked and were assigned the WXCN classification (Fig. 1). All other nonpassing MRMS gauge QC flags (i.e., outlier and suspect values) were classified as FAIL.

c. Temporal analysis for all observations

Each gauge site accumulated a history of hourly temporal QC classifications along with supplemental information (e.g., gauge and radar QPE values, average RAQI). The temporal
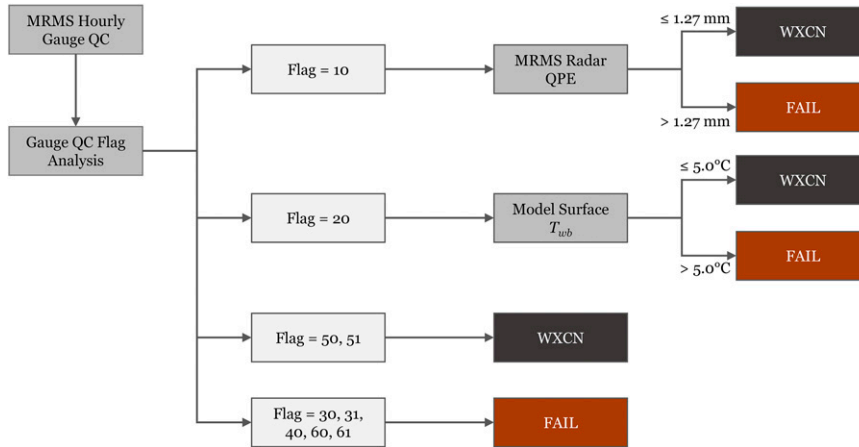
FIG. 1. Temporal QC logic decision tree focused on the hourly classifications for the gauge observations that failed the gauge QC applied within the MRMS system.

QC scheme then separately evaluated the history of each gauge over a running 7- and 30-day period every hour. The selection of the two different temporal periods were designed to provide two different perspectives on gauge performance: 1) a short-term evaluation for early identification of an instrumentation malfunction and 2) a longer-term evaluation for a more comprehensive assessment.

The temporal QC analysis for each gauge and evaluation period was conducted via a two-part classification scheme. The first phase examined all available observations for gauges with ≥10% data availability within the evaluation period (Fig. 2). Emphasis was placed on the percentage of hourly observations with a FAIL classification over the evaluated period. A three-tiered classification was utilized in this phase of the temporal QC:

- GOOD: The gauge site performance is considered to be of good quality.
- SUSP: The gauge site performance is considered to be of suspect quality and warrants further inspection.
- BAD: The gauge site performance is considered to be of poor quality and has a high potential of having an instrumentation malfunction. The gauge site warrants further inspection.

The first phase of the temporal QC analysis focused on the examination of the hourly FAIL percentage for gauges with sufficient availability. A gauge was assigned a BAD classification when ≥12% of available hourly observations were declared as FAIL. A GOOD classification was given when <4% of available hourly observations were declared as FAIL. Percentages in between resulted in a SUSP classification. A separate long-term weather impact classification, designated as WXIMP, is considered for each gauge during the 7- and 30-day periods within this phase of the temporal QC logic. The weather impact analysis was designed to provide information on which sites may have experienced significant periods of environmental influences on the gauge observations. Gauges with ≥5% of hourly observations assigned as WXCN resulted in a WXIMP = yes classification. Gauge sites with <5% of available hourly observations defined as WXCN resulted in WXIMP = no.

d. Temporal analysis for nonzero precipitation observations

The second phase of the temporal QC scheme examining the history of hourly observations focused only on the hours when precipitation was observed (i.e., both the gauge and
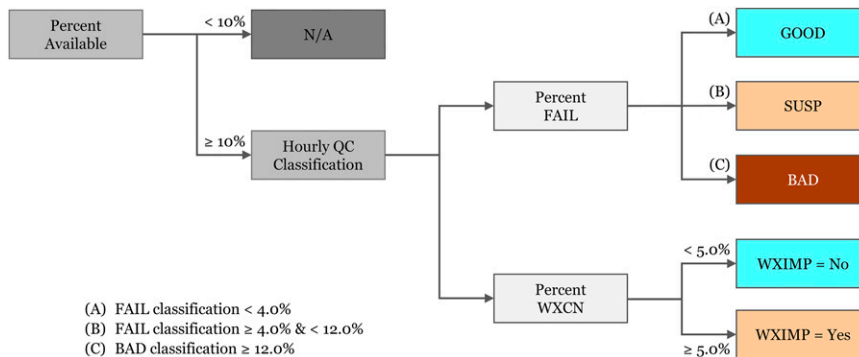

FIG. 2. Temporal QC decision tree for all available observations. Decision tree is used for both the running 7- and 30-day analyses.
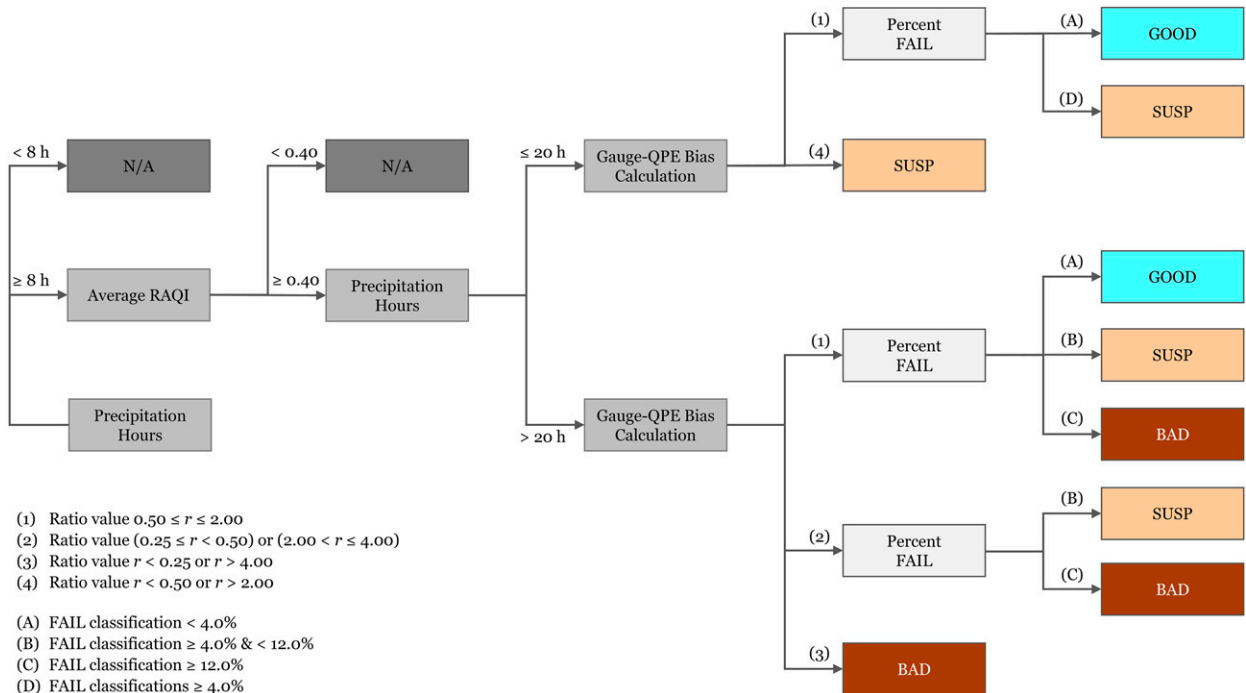
FIG. 3. As in Fig. 2, but for precipitation-only observations (i.e., when the gauge and collocated gridded radar QPE are nonzero).

collocated gridded radar QPE were nonzero). This phase of the temporal QC analysis was conducted if at least 8 h recorded observed precipitation and if the gauge site was in a region of adequate radar coverage (Fig. 3). Defining at least eight precipitation hours allowed for a minimal sample size to conduct further analyses. An average RAQI value ≥ 0.40 during the temporal analysis period was used to denote gauges that could be compared to the MRMS dual-polarization synthetic radar QPE, which was based on the RAQI guidance from the MRMS gauge QC algorithm (Martinaitis et al. 2021).

Gauge sites meeting the sample-size and RAQI requirements were then evaluated using a gauge-QPE bias check and the percentage of precipitation observations that were classified as FAIL (Fig. 3). The evaluation of gauges having 8–20 h of observed precipitation resulted in only a GOOD or SUSP temporal QC classification, since there were enough observations to analyze it but not enough to effectively determine if an instrumentation malfunction existed. The evaluation of gauges with >20 h of observed precipitation could utilize the BAD classification given a more sufficient number of hours to examine. Gauge sites with an overall bias ratio value of <0.25 or >4.00 were assigned a BAD classification. All other gauges leveraged both the bias ratio and the percentage of FAIL observations during precipitation-only hours to assign a classification for this component.

e. Final temporal QC classification

The last step of the temporal QC scheme combines the two QC classification components into a final overall classification for each gauge per analysis period (Fig. 4). The lower classification level between the two analyses becomes the final temporal QC classification. If only one of the two temporal QC

analysis schemes produced a temporal QC classification (i.e., one analysis yielded an N/A), then the remaining non-N/A classification would be declared as the final classification. The final temporal QC flags from the 7- and 30-day observational histories were not combined (i.e., each remained as an independent evaluation) to provide the ability to interrogate gauge performance across different temporal analysis periods.

3. Data and methodology

a. Input hourly gauge observations

Hourly gauge observations utilized in this study primarily consisted of two national collections of gauge networks: the Hydrometeorological Automated Data System (HADS; Kim et al. 2009) and the Meteorological Assimilation Data Ingest System (MADIS; Helms et al. 2009). Additional regional networks included the Snow Telemetry (SNOTEL; e.g., Serreze et al. 1999), the Flood Control District of Maricopa County (FCDMC; e.g., Mascaro 2017), and the Harris County Flood Control District (HCFCD; e.g., Wolff et al. 2019) networks. Hourly gauge observations were available from 1 June 2020 to 30 September 2021 over the MRMS CONUS domain, which covers the conterminous United States as well as portions of southern Canada and northern Mexico (Fig. 5a) Over 37 000 unique gauge sites were identified within the domain area with 69.4% of gauge sites having a greater than 10% data availability across the entire study period (Fig. 5b).

b. Application of temporal QC scheme

The June 2020 observations were used to spin up both the running 7- and 30-day temporal analyses, allowing for hourly

QC Analysis: All Available Observations	QC Analysis: Precipitation-Only		Final Temporal QC Classification
GOOD	GOOD	→	GOOD
GOOD	SUSP	→	SUSP
GOOD	BAD	→	BAD
GOOD	N/A	→	GOOD
BAD	GOOD	→	BAD
BAD	SUSP	→	BAD
BAD	BAD	→	BAD
BAD	N/A	→	BAD

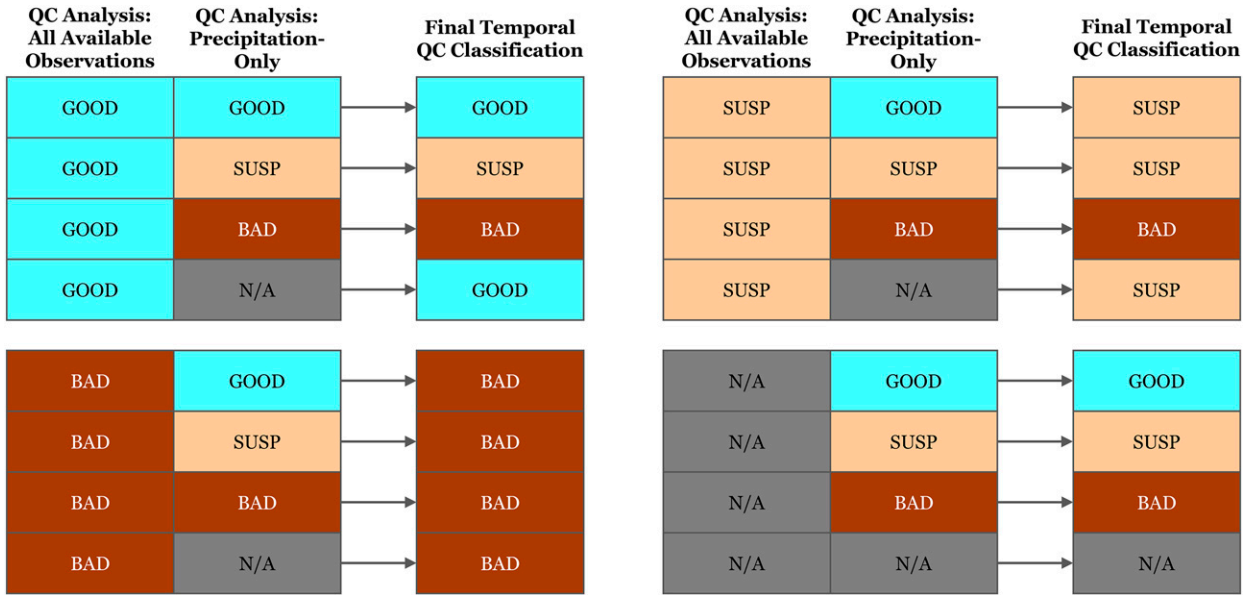| QC Analysis: All Available Observations	QC Analysis: Precipitation-Only		Final Temporal QC Classification
SUSP	GOOD	→	SUSP
SUSP	SUSP	→	SUSP
SUSP	BAD	→	BAD
SUSP	N/A	→	SUSP
N/A	GOOD	→	GOOD
N/A	SUSP	→	SUSP
N/A	BAD	→	BAD
N/A	N/A	→	N/A

FIG. 4. Graphical representation of how the final temporal QC classification is assigned based on the analyses of both all available observations and precipitation-only observations.

assessments of the temporal QC logic to begin at 0000 UTC 1 July 2020. Examinations of the hourly temporal QC distributions over the 15-month study period focused on the overall temporal QC classifications and its two temporal QC classification components. Further examinations of the BAD temporal QC classifications and observations adversely affected by meteorological conditions were also conducted. Focus was placed on the hourly MRMS QC flags and the associated environmental and precipitation properties to examine the distribution and seasonal variations of impacts that resulted in the hourly FAIL and WXCN classifications that influenced the final temporal QC classification. Select gauge sites that received a BAD or SUSP classification and were maintained by the U.S. Geological Survey (USGS) were further investigated to confirm whether the temporal QC classification was reasonable based on documented gauge maintenance. This included an analysis of data patterns and statistical evaluations of clogged gauges during precipitation events. Correspondence with local USGS field offices were used to verify the identification of an instrumentation malfunction.

a) Percent Availability of Gauge Observations

b)

Availability	Percent	Availability	Percent
100–90%	28.82	50–40%	2.97
90–80%	13.14	40–30%	2.12
80–70%	9.69	30–20%	2.14
70–60%	4.40	20–10%	1.68
60–50%	4.40	< 10%	30.64

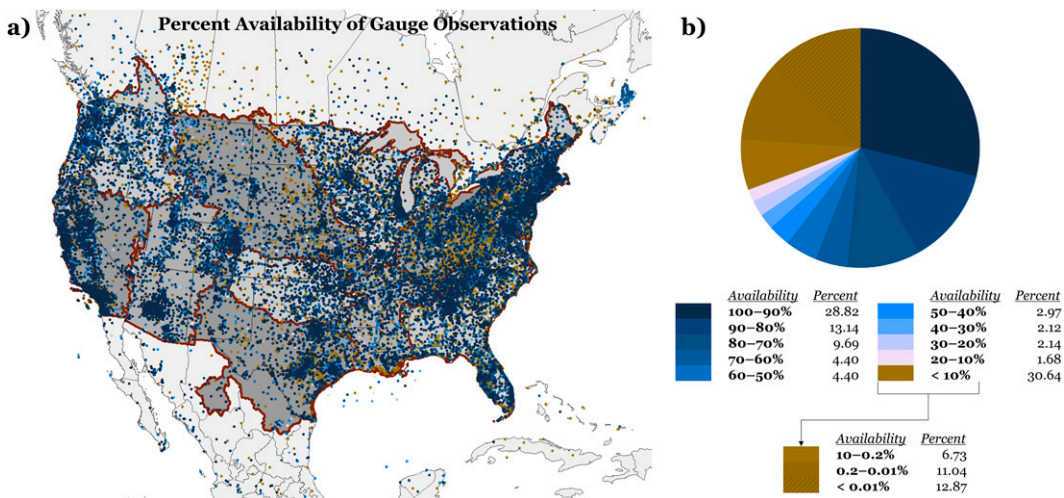| Availability	Percent
10–0.2%	6.73
0.2–0.01%	11.04
< 0.01%	12.87

FIG. 5. (a) Map of all gauge sites that had at least one observation ingested by the MRMS system during the study period for the MRMS CONUS domain area, which includes southern Canada and northern Mexico. Highlighted in brown contours are the NWS River Forecast Center areas of interest. (b) Each gauge shown is color-coded to its percentage availability with the distribution of the percentage availability.
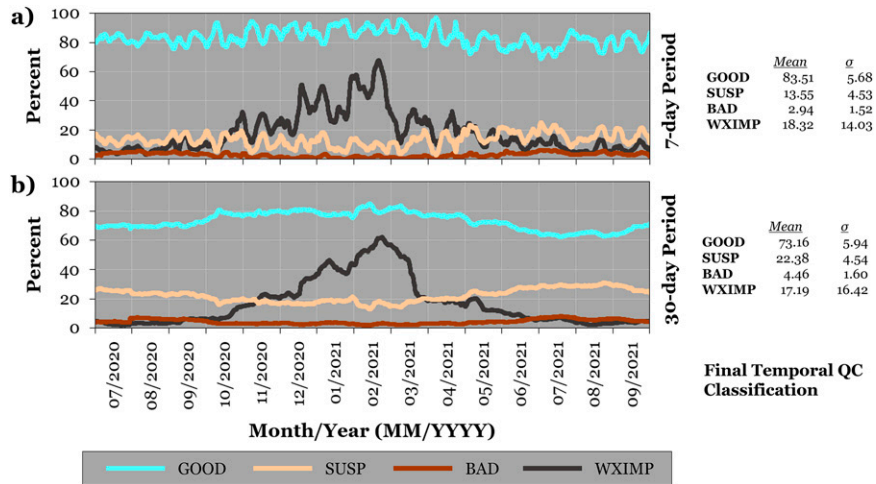
FIG. 6. The percentage distribution of the final temporal QC classification (GOOD, SUSP, and BAD) along with the WXIMP classification for each hour across the entire study period. The plot of each distribution along with the mean hourly percentage and standard deviation of the hourly percentage are presented for (a) the running 7-day analysis and (b) the running 30-day analyses.

c. Data and study limitations

There are limitations to consider with the temporal QC scheme, notably with how some gauge sites get classified based on their observational history, the use of gridded QPE data as a comparative data source, and gridded QPE data biases in the analyses. Gauge systematic errors from environmental factors can potentially create significant differences between the gauge observation and radar-derived QPE that could influence the temporal QC designation. There is limited research related to the thawing of winter precipitation, including the challenge of melting during a precipitation event, which can influence the bias ratio comparison with gridded radar QPE (Martinaitis et al. 2015). Biases and limitation with radar-derived QPEs values can influence the hourly QC and the precipitation-based temporal analysis. Radar-derived

QPE limitations can include miscalibration of radar data, unrepresentative rate relationships, sampling through the melting layer, nonmeteorological echoes, and nonuniform beam filling (e.g., Wilson and Brandes 1979; Smith 1986; Rosenfeld et al. 1992; Young et al. 1999). Biases in model fields used in the various QC methodologies can also influence how a gauge is analyzed within the decision trees.

4. Temporal QC findings

a. Overall temporal QC classifications

Findings from the final temporal QC designations across the 15-month period were characterized by a generally consistent application of the GOOD, SUSP, and BAD classifications for both the running 7- and 30-day analyses (Fig. 6). The distribution of temporal QC classifications for the running 7-day analysis was as follows: 83.5% for GOOD, 13.6% for SUSP, and 2.9% for BAD (Fig. 6a). A similar distribution

TABLE 2. Overall and seasonal mean hourly percentage distribution and percentage standard deviation of the final temporal QC classifications (GOOD, SUSP, and BAD) and the additional WXIMP classification for the 7-day analysis.

Temporal QC classifications	Overall	Seasonal			
		SON	DJF	MAM	JJA
Mean percentage					
GOOD	83.51	82.99	88.59	84.68	80.26
SUSP	13.55	13.90	10.10	13.43	15.36
BAD	2.94	3.11	1.30	1.89	4.38
WXIMP	18.32	13.80	40.22	20.42	7.81
Standard deviation					
GOOD	5.68	5.11	3.96	5.80	4.41
SUSP	4.53	4.17	3.42	5.02	3.86
BAD	1.52	1.13	0.72	0.94	0.86
WXIMP	14.03	7.47	12.79	6.47	3.15

TABLE 3. As in Table 2, but for the 30-day analysis.

Temporal QC classifications	Overall	Seasonal			
		SON	DJF	MAM	JJA
Mean percentage					
GOOD	73.16	73.67	79.58	76.38	67.08
SUSP	22.38	21.87	17.39	20.37	26.91
BAD	4.46	4.46	3.03	3.25	6.01
WXIMP	17.19	9.66	41.27	24.36	4.74
Standard deviation					
GOOD	5.94	4.93	2.02	3.51	2.89
SUSP	4.54	3.80	1.60	2.94	2.33
BAD	1.60	1.25	0.53	0.65	1.22
WXIMP	16.42	6.46	11.99	13.46	2.31

Final Temporal QC Classification
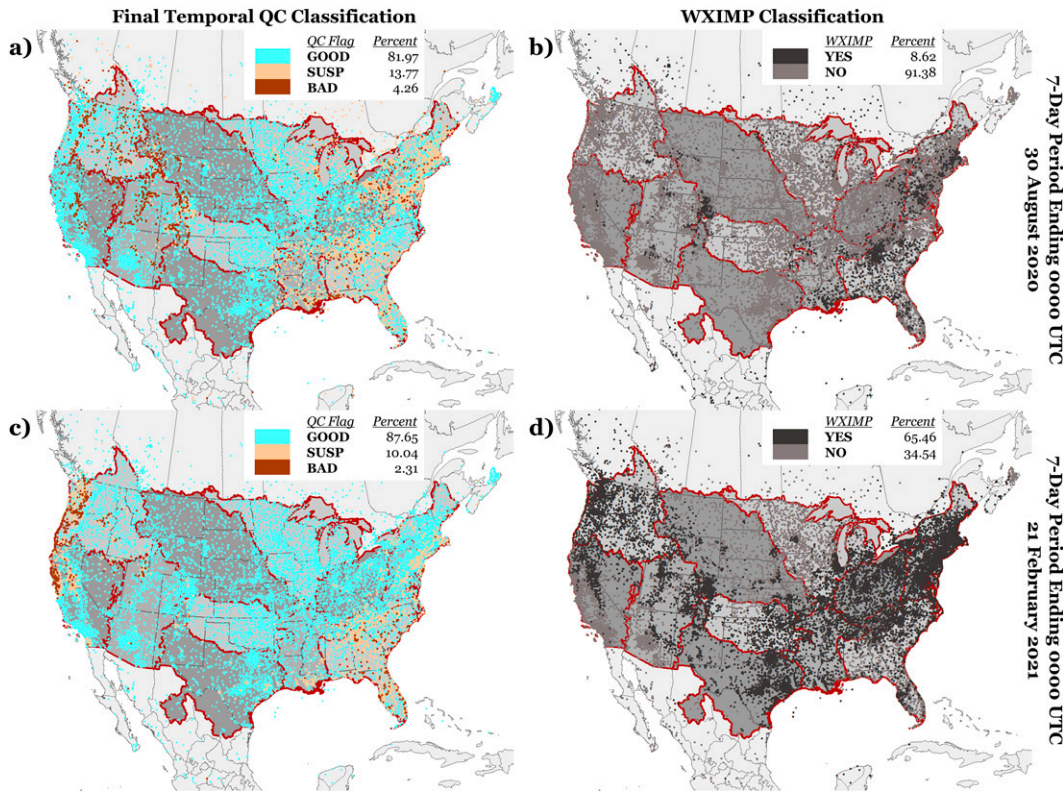
WXIMP Classification

FIG. 7. Map of the (left) final temporal QC classifications and (right) the WXIMP classification for the 7-day analysis ending (top) 0000 UTC 19 Dec 2020 and (bottom) 0000 UTC 30 Jun 2021. Each map is for the MRMS CONUS domain area, which includes southern Canada and northern Mexico. Highlighted in brown contours are the NWS River Forecast Center areas of interest.

was found with the 30-day analysis at 73.2% for GOOD, 22.4% for SUSP, and 4.4% for BAD (Fig. 6b). The significant reduction of GOOD temporal QC classifications and increase in SUSP temporal QC classifications with the running 30-day analysis can be attributed to the ability to capture more observations that

could help highlight more subtle trends that could identify a potential instrument malfunction. This also explains the slight increase in BAD classifications within the longer 30-day analysis.

There were seasonal subtleties across the three-tiered temporal QC designations. The running 7-day temporal QC analysis
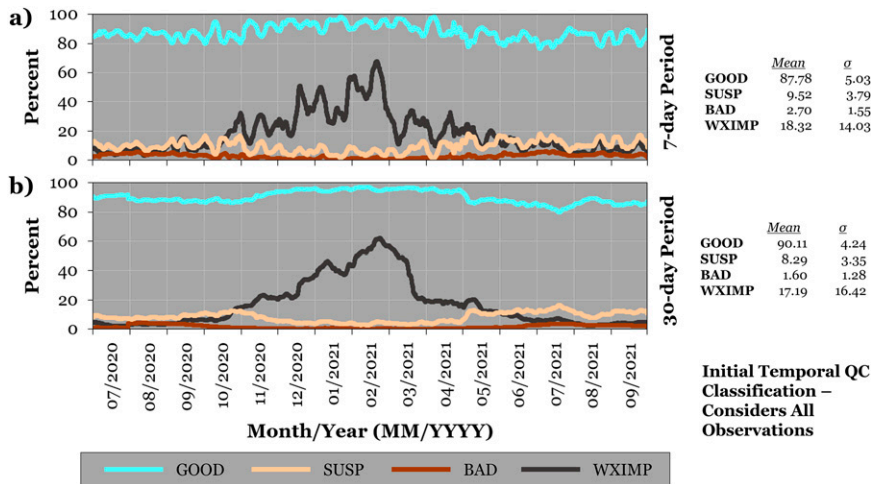

FIG. 8. As in Fig. 6, but for the initial temporal QC classifications and the WXIMP classification for the analysis of all available observations.

TABLE 4. Seasonal mean hourly percentage distribution and percentage standard deviation of the percentage of gauge sites that met the temporal QC precipitation-only analysis criteria.

Temporal QC analysis period	Overall	Seasonal			
		SON	DJF	MAM	JJA
Mean percentage					
7-day analysis	21.00	19.34	21.84	23.67	20.22
30-day analysis	55.65	51.52	54.63	57.55	58.36
Standard deviation					
7-day analysis	7.58	6.28	7.08	8.78	7.54
30-day analysis	5.49	5.91	2.39	4.50	4.78

saw the average percentage of GOOD flags peak during the December–February (DJF) months at 88.6% while decrease to an average of 80.3% during the June–August (JJA) months (Table 2). Contrasting this were the peaks of both SUSP and BAD classifications occurring during the JJA months and the reduction of SUSP and BAD classifications during the DJF period. Greater variations in the final temporal QC classifications occurred during the seasonal transition months of September–November (SON) and March–May (MAM) based on standard deviation values. A similar seasonal pattern was also found in the running 30-day analysis (Table 3).

The greatest variations within the temporal QC algorithm were found in the WXIMP yes/no classification. A significant number of gauges were designated with the additional WXIMP classification from October 2020 to May 2021 (Fig. 6), which coincided with the period of winter precipitation impacts across the domain. The average percentage of WXIMP yes classifications was 18.3% for the 7-day analysis and 17.2% for the 30-day analysis, yet the average percentage of gauges assigned the additional WXIMP flag was >40% for the DJF months but <8% for the JJA months (Tables 2 and 3). It is no coincidence that there was a decrease in the SUSP and BAD temporal QC classifications when more gauge observations

observed weather impacts. Both the hourly FAIL and WXCN classifications utilize the same MRMS gauge QC flags. The scheme looks to delineate between truly failed observations and observations that were likely influenced by a meteorological factor through basic thresholds in this presented methodology, yet it is plausible that some hourly observations that should be declared as FAIL were misclassified as WXCN. Further examination and discussion of the WXIMP classifications and its role with the three-tiered temporal QC classifications are discussed in section 4e.

An examination of two points in time highlights the distribution of the temporal QC classifications and the impacts of associated meteorological conditions on the classifications (Fig. 7). The majority of gauge sites with a SUSP classification were in regions that received the majority of rain events. Gauges classified as BAD were scattered throughout the domain without any spatial patterns, with the exception of the SNOTEL gauges in the western CONUS. Hourly SNOTEL observations are subject to 1) data fluctuations related to temperature oscillations that create variations in the speed of sound with the ultrasonic detector despite a corrective algorithm (Osterhuber et al. 1994; Avanzi et al. 2014) and 2) reporting at a coarser data resolution of 2.54 mm. The lack of SUSP and BAD classifications across most of the domain with the 0000 UTC 21 February 2021 example (Fig. 7b) were a by-product of winter precipitation impacts on gauge sites (Fig. 7d), which limited the ability to properly interrogate those gauges.

b. Reasoning for final temporal QC classifications

Investigating how gauges were assigned a temporal QC classification each hour for the 7- and 30-day analyses can be explored through the all-observations and precipitation-only components of the temporal QC logic described in sections 2c and 2d. The temporal QC component considering all available observations was characterized by a high percentage of GOOD classifications for both the running 7- and 30-day
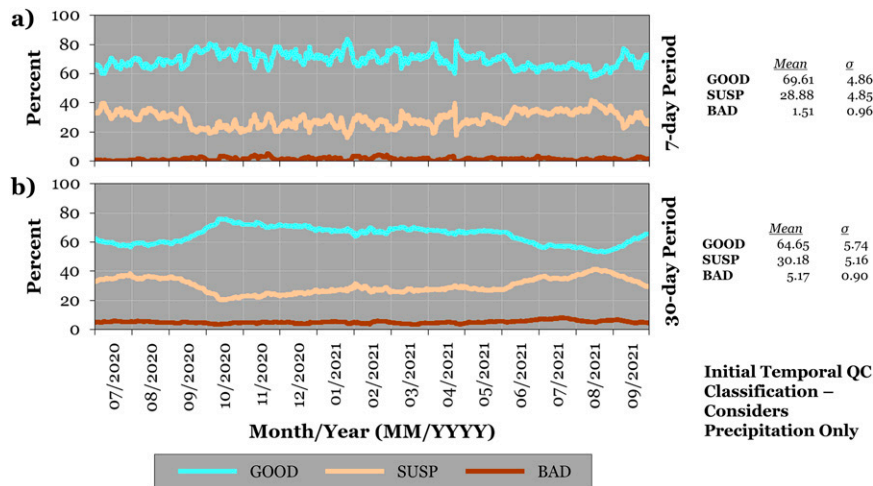

FIG. 9. As in Fig. 6, but for the initial temporal QC classifications for the analysis of precipitation-only observations.

TABLE 5. The average hourly percentage distribution and percentage standard deviation of the 7- and 30-day final temporal QC classification pairings across the study period.

Temporal QC classification			
7-day analysis	30-day analysis	Average percentage	Standard deviation
GOOD	BAD	28.43	13.84
SUSP	BAD	24.05	5.03
BAD	BAD	25.12	11.10
BAD	SUSP	20.11	9.27
BAD	GOOD	2.29	2.07

analyses at 87.8% and 90.1%, respectively (Fig. 8). The average percentage of SUSP classifications was between 8% and 10%, and the average percentage of BAD classifications was under 3%. Some seasonal variations can be seen across the three-tiered temporal QC designation, while there were also additional daily to weekly variations based on the precipitation occurrences and their influences on gauges across the domain. Regardless of these variations, the standard deviation values for the GOOD, SUSP, and BAD classifications were ≤5.0%.

The distribution of temporal QC classifications with the precipitation-only component of the scheme has a number of differences when compared to the all-observations component. The requirements for the precipitation-only component of the temporal QC analysis stated that there must be at least eight hours of precipitation during the analysis period and the gauge must reside in an area of an average RAQI ≥ 0.40 (Fig. 3). These requirements limited the number of gauges that went through this component of the temporal QC analysis. An average of 21.0% of gauge sites met this criterion across the running 7-day analysis and 55.7% for the 30-day analysis with some seasonal variability (Table 4).

The percentage of site classified as GOOD for the precipitation-only analyses was 69.6% and 64.6% for the two different temporal analysis periods, respectively (Fig. 9). This reduction of GOOD classifications compared to the all-observations analyses was in response to the increase of SUSP classifications to 28.9% (7-day analysis) and 30.2% (30-day analysis), respectively. The average percentage of gauges that were classified as BAD during the 7-day was only 1.5% (Fig. 9a), yet the

percentage of gauges with BAD classifications for the 30-day analysis increased to 5.2% (Fig. 9b). This is attributed to more precipitation hours being available for analysis with the longer temporal period. There were some seasonal variations in both the application of the GOOD and SUSP temporal QC classifications; however, the percentage of observations classified as BAD within the precipitation-only component was consistent throughout the study period.

c. BAD classifications between temporal evaluation periods

One notable result from the temporal QC classifications was that a BAD classification via the 7-day analysis does not directly translate to a BAD classification with the 30-day analysis and vice versa. An average of 25.1% of gauge sites per hour had a BAD classification for both the 7- and 30-day analyses with a percentage standard deviation of 11.1% (Table 5). Similar percentage values were shown for when the 7-day analysis was SUSP and GOOD while the 30-day analysis was BAD. This can be attributed to erroneous observations appearing at various frequencies across a running 7-day period but was more noticeable in a 30-day evaluation. Approximately 20.1% of gauge sites were classified as BAD with the 7-day analysis and SUSP with the 30-day analysis, which signifies that impact was localized to that 7-day period but enough to trigger suspicion at a longer temporal analysis period. Only 2.3% of gauge sites were shown as BAD with the 7-day analysis but GOOD on the 30-day analysis.

d. Evaluation of the BAD QC classification

Understanding why gauge sites were assigned with the temporal QC BAD classifications required investigating the hourly performance of gauges. A breakdown of the initial hourly FAIL classifications showed that the false nonzero precipitation values (i.e., MRMS QC flag = 20) were the dominant classification (Fig. 10) and represented an average of 76.0% of hourly FAIL classifications. The associated large standard deviation of observations per hour was attributed to both changes in daily weather patterns and seasonal variability, notably the substantial seasonal decrease during the winter months. All other modes resulting in an hourly FAIL classification were steady throughout the study period with no

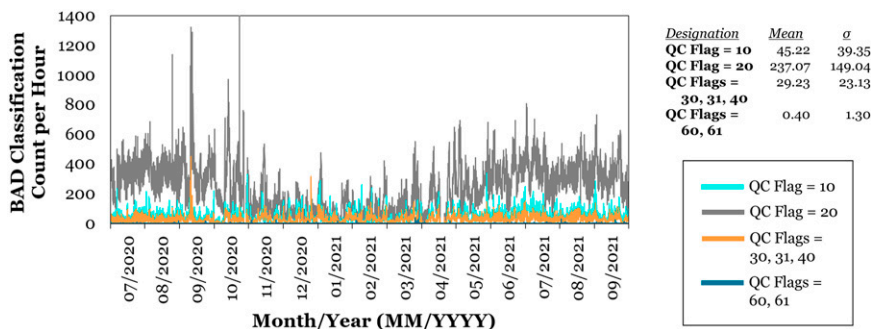| Designation	Mean	σ
QC Flag = 10	45.22	39.35
QC Flag = 20	237.07	149.04
QC Flags = 30, 31, 40	29.23	23.13
QC Flags = 60, 61	0.40	1.30

FIG. 10. Number of observations per hour for each MRMS gauge QC classification that receive the initial hourly FAIL classification based on Fig. 1.

TABLE 6. Average hourly percentage distribution of how gauge sites were assigned the FAIL classification within the precipitation-only analysis.

Temporal QC analysis period	Failed percentage of FAIL observations only	Failed bias ratio check only	Failed both checks
7-day analysis	88.29	0.84	10.87
30-day analysis	86.79	1.66	11.55

notable seasonal variations. The important factor in relating the hourly FAIL classifications to the resulting BAD temporal QC classification is that the two dominant reasonings for a FAIL flag (false zero and false nonzero observations) would be captured through the all-observations component of the temporal QC logic.

The precipitation-only component yielded the greatest percentage of the BAD classifications. Most gauge sites with a BAD classification through the precipitation-only analysis acquired the classification through the percentage of FAIL hours (Table 6). For the 7-day running analysis, an average of 88.3% of declared BAD gauge sites were failed through only the percentage of FAIL hours. The high percentage of BAD gauges via the precipitation-only analysis meeting the FAIL percentage threshold and not having a poor accumulative bias ratio can be attributed to variations between smaller gauge hourly accumulations (e.g., <5 mm) and radar-derived QPE that could flag a gauge enough times during an analysis period while performing adequately with larger hourly accumulations. This could also be a by-product of the combination of minimum number of hours (20) and the percentage of FAIL gauges threshold (12%) needed to result in the BAD classification. About 10.9% failed both the percentage of FAIL hours and the bias ratio check, while only 0.8% of gauge sites failed only the bias ratio check. This means that most gauges that failed the bias ratio check also failed the percentage of FAIL hours criteria. Similar results were found in the 30-day analysis (Table 6), and both temporal period evaluations had no notable seasonal variations.

e. Weather impact identification

The reduction in temporal QC BAD classifications during the cool season can be contributed to the challenge of identifying potential instrumentation malfunctions during instances of winter precipitation, which tend to create poor gauge observations. Hourly gauge observations identified as having winter precipitation impacts (MRMS QC flags = 50, 51) and false nonzero observations (MRMS QC flag = 20) that are potentially related to postwinter precipitation thawing were the dominant reasons for the hourly WXCN classifications across the cool season (Fig. 11). The MRMS QC algorithm can identify of hourly accumulations that were either significantly reduced or were zero during winter precipitation (Martinaitis et al. 2015, 2021); however, there was uncertainty in determining the classification of gauge impacts related to postwinter precipitation thawing.

A distinct diurnal cycle existed in the false nonzero observations (not shown) that was equivalent to the findings of Martinaitis et al. (2015). There was an average of 243 false nonzero observations per hour between 0100 and 1400 UTC, which then increased to an average of 609 false nonzero observations per hour from 1700 to 2100 UTC. An examination of hours with at least 500 false nonzero observations (i.e., hours with large areas of postwinter precipitation thawing) found significant coefficient of determination (R^2) values with the T_{wb} ranges of $-5.0° < T_{wb} \leq 0.0°C$ ($R^2 = 0.45$) and $0.0° < T_{wb} \leq 5.0°C$ ($R^2 = 0.60$), respectively (Fig. 12). Most other T_{wb} ranges had poor data correlation, yet some instances of significant hourly counts of gauge observations within the $5.0° < T_{wb} \leq 10.0°C$ range signified that there might be some hourly gauges that observed winter precipitation thawing but were misidentified as FAIL in the scheme.

All other hourly observations having a WXCN classification were from false zero observations (MRMS QC flag = 10) that were likely a result of precipitation detected aloft by radar but evaporated before reaching the surface (Fig. 11). No seasonal variability existed in this dataset, yet a large standard deviation existed based on daily weather patterns. Approximately 88.04% of observations declared as a false zero were in regions where hourly radar precipitation values were ≤1.27 mm (Fig. 13). This represented an average of 283 gauge observations per hour. Approximately 69.5% of all observations occurred in regions where the radar precipitation was ≤0.51 mm. An average of 38 observations per hour were classified as FAIL with having collocated precipitation values > 1.27 mm.
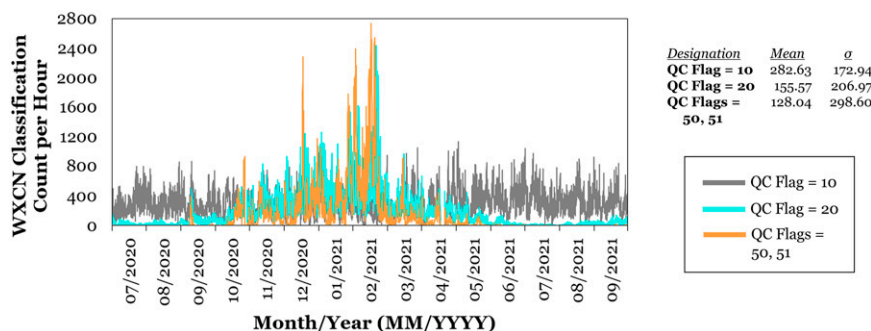

Designation	Mean	σ
QC Flag = 10	282.63	172.94
QC Flag = 20	155.57	206.97
QC Flags = 50, 51	128.04	298.60

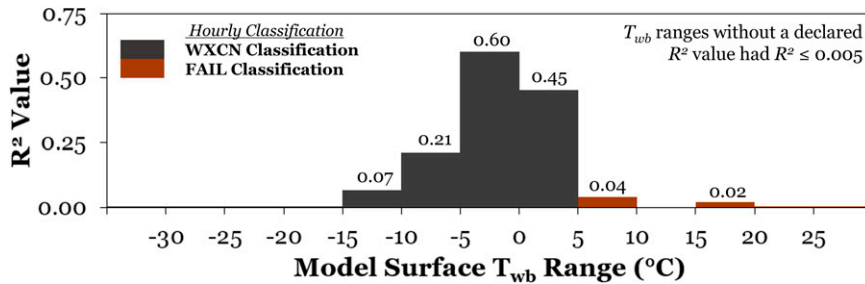FIG. 11. As in Fig. 10, but for the initial hourly WXCN classification.

FIG. 12. Histogram of the coefficient of determination (R^2) values comparing the overall number of observations per hour with the MRMS QC flag = 20 and the number of observations per hour within the specific model surface T_{wb} range. The histogram colors are based on the model surface T_{wb} values that lead to an hourly WXCN (gray) or FAIL (brown) classification.

Having a significant number of gauges with the additional WXIMP flag during the cool season months can potentially masked instrumental errors with gauges. The blanket application of potential weather impacts on hourly observations that could be influenced by winter precipitation, postwinter precipitation thawing, and potential virga not fully removed by the MRMS system can inaccurately mark hourly observations as WXCN that were truly erroneous observations. It can also be possible that having a greater frequency of winter precipitation impacts across the cool season minimizes the opportunities to detect potential instrumentation malfunctions with gauges.

5. Verification

The most common gauge instrumentation malfunction in the study was a clogged orifice or cylinder. These examples of a gauge clog showed a period of precipitation followed by multiple hours of either a constant or periodic recording of hourly accumulations that were recorded as false nonzero observations (Fig. 14). The resulting residual accumulations would vary based on the amount of clogging and how long it occurred after the precipitation event. Typical residual accumulations occurring immediately after a precipitation event

would range from 0.52 to 2.03 mm while more sporadic accumulations toward the end of the period of false observations were 0.25 mm. Each clogged gauge shown in Fig. 14 was characterized by different properties based on the degree of clogging. The frequency in the reporting of residual rainwater with the clogging of PHIW2 (Fig. 14a) increased based on the amount of estimated precipitation that occurred at the gauge site (Table 7). The gauge TDGL1 was characterized by a more substantial clog and tapering of residual rainfall collection after two significant precipitation events (Fig. 14b) where less than 40% of the estimated precipitation was recorded during the event followed by over 140 h of residual accumulations (Table 7). The evolution of a clogging with STKN4 influenced both the amount of precipitation accumulated during the precipitation event and the duration of residual accumulations postprecipitation (Fig. 14c). Bias ratio between the gauge and MRMS radar-derived QPE for the first two precipitation events were >0.50 and then later followed by a significant underestimation bias ratio < 0.11 (Table 7). The period of residual accumulations after the precipitation events continued to increase as the clogging became more significant until the clog was cleared between 1600 and 1700 UTC 24 August 2021.
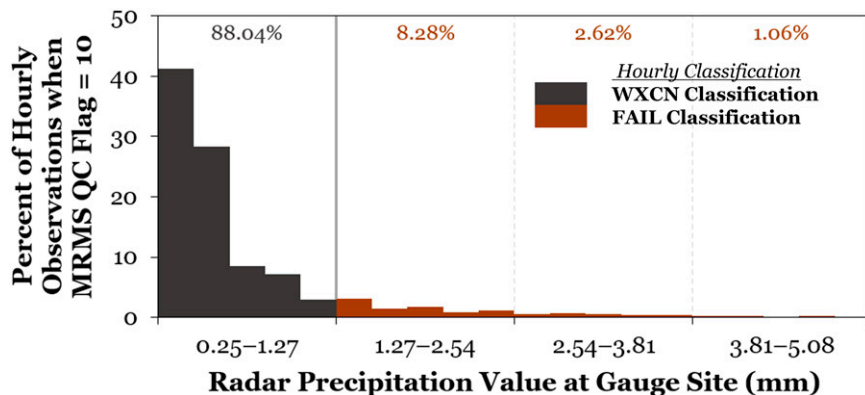

FIG. 13. Histogram of the percentage of hourly observations with the MRMS QC flag = 10 and the MRMS gridded radar QPE values collocated at the gauge site. The histogram colors are based on the MRMS gridded radar QPE values that lead to an hourly WXCN (gray) or FAIL (brown) classification.
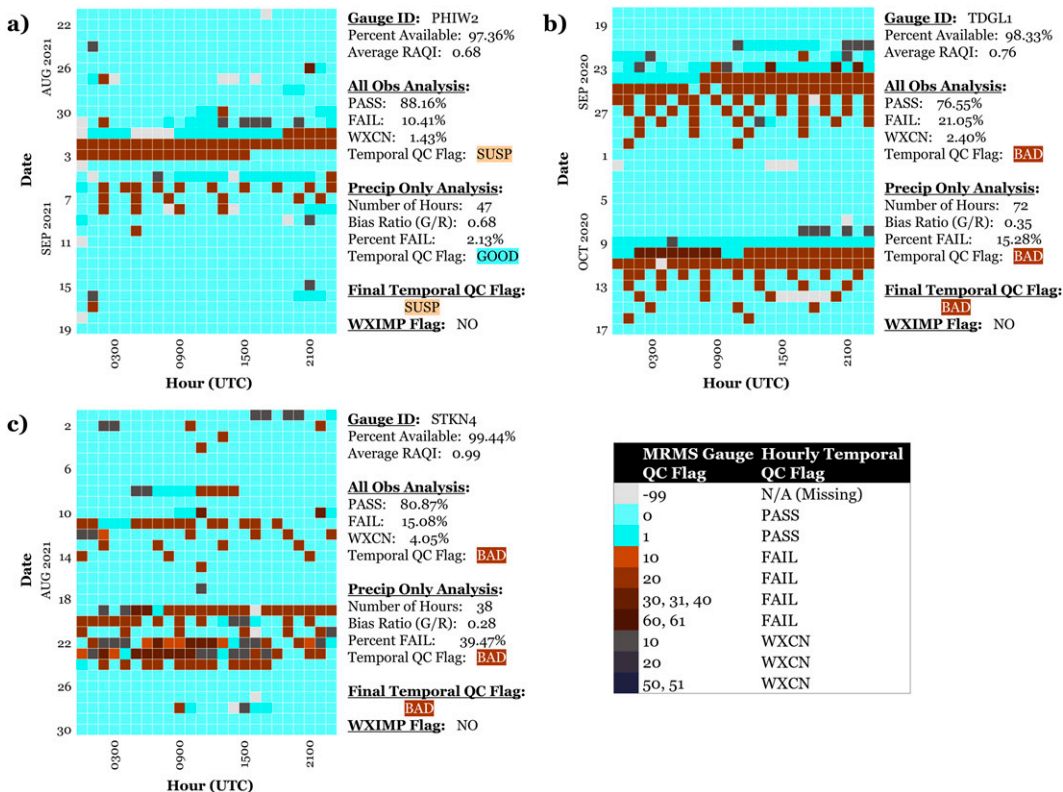
FIG. 14. A 30-day × 24-h matrix of hourly QC classifications for gauges that were impacted by the clogging of the gauge orifice or gauge cylinder. The hourly classifications shown in each matrix are based on the MRMS gauge QC scheme and its resulting hourly temporal QC classification. The general shading of each hour was based on the temporal QC classifications of PASS (cyan), FAIL (brown), WXCN (dark gray), and N/A (or missing; light gray). The gauges shown in the analysis are (a) PHIW2 from 0000 UTC 21 Aug to 2300 UTC 19 Sep 2021, (b) TDGL1 from 0000 UTC 18 Sep to 2300 UTC 17 Oct 2020, and (c) STKN4 from 0000 UTC 1 Aug to 2300 UTC 30 Aug 2021. Presented with each gauge are a series of statistical measures and temporal QC classifications, including the percentage of gauge observations available, average RAQI value for each analysis period, analysis from the all observations and precipitation-only components with their resulting temporal QC classifications, the final temporal QC classification, and the additional WXIMP designation.

Another example of an obstruction that influenced gauge observations were events that generally impacted the precipitation hours only and left no residual drip or accumulation postevent. The clogging of LBJK2 still had 92% of all available hourly observations set as PASS with only 3.5% of all observations set as FAIL (Fig. 15). Most of the FAIL observations were assigned a MRMS QC flag = 40 (outlier low) during the precipitation events, which resulted in a precipitation-only fail rate of 61.3%; moreover, the bias ratio of 0.15 showed a significant underreporting by the gauge. There were also numerous hourly observations assigned a MRMS QC flag = 10 that were designated as BAD or WXCN. The frequency of false zero observations regardless of hourly classification combined with the high quality of radar data (RAQI = 1.00; radar beam elevation of approximately 160 m) further exemplified the obstruction-type issue with LBJK2.

Additional types of instrumentation malfunctions were also observed and verified. The site TDGL1 was located on a highway bridge and had a new rain gauge installed on 13 August 2021;

however, instability issues with the gauge installation platform yielded a mix of erroneous and missing observation values (Fig. 16a). Most missing and false nonzero observations occurred between 1300 and 0300 UTC daily, corresponding to increase motor vehicle traffic on the bridge. The abrupt change in performance for the site HBCN1 in August 2020 was due to a firmware update that resulted in a software issue with the data collection platform setup (Fig. 16b). Data reporting after the firmware update featured a mix of missing observations and erroneous reporting of false hourly gauge values of 5–160 mm until the gauge was reprogrammed on 26 August 2020. A likely hardware issue or associated wiring issue with the datalogger and transmitter resulted in erroneous values during precipitation events for the site LNRI2 (Fig. 16c). Erroneous hourly gauge observations during precipitation events were typically exceeding 300 mm and peaked at 1600 mm, while MRMS hourly radar-derived QPE accumulations were 3–18 mm. This resulted in both a percentage classification of FAIL during precipitation hours at 95.5% and a bias ratio of 56.49.

TABLE 7. Analysis of precipitation events and the time of postprecipitation drip into the gauge from clogging for the gauges PHIW2, TDGL1, and STKN4 shown in Fig. 14. Listed in the table are the start and end times for analyzed precipitation periods that were influenced by a clogging with the gauge, the accumulation of precipitation from the gauge (G) and MRMS radar-derived QPE (R) during the period of precipitation only (i.e., not accounting for the collection of water by the gauge after the precipitation period), and the duration of time (in hours) after the precipitation period that observed constant and/or periodic dripping of residual rainfall into the gauge. The asterisk denotes that the last three precipitation and residual dripping events listed for gauge STKN4 were overlapping, and the residual accumulation period accounts for the combination of all three events until the clog was removed by a USGS technician.

	Precipitation period with associated blockage		Accumulation during precipitation period			
Gauge ID	Start time	End time	Gauge total (mm)	Radar total (mm)	Bias ratio (G/R)	Duration of residual accumulation (h)
PHIW2	2300 UTC 31 Aug 2021	1700 UTC 1 Sep 2021	23.87	36.20	0.659	45
	0900 UTC 5 Sep 2021	0000 UTC 6 Sep 2021	6.07	8.60	0.706	63
TDGL1	1200 UTC 23 Sep 2020	0700 UTC 24 Sep 2020	21.35	77.00	0.277	148
	0000 UTC 9 Oct 2020	1100 UTC 10 Oct 2020	36.59	102.50	0.357	145
STKN4	0700 UTC 8 Aug 2021	1100 UTC 8 Aug 2021	4.07	7.70	0.529	4
	2200 UTC 10 Aug 2021	0400 UTC 11 Aug 2021	8.64	12.60	0.686	103
	0500 UTC 19 Aug 2021	0700 UTC 19 Aug 2021	3.81	39.80	0.096	134*
	0600 UTC 22 Aug 2021	1600 UTC 22 Aug 2021	1.25	41.80	0.030	
	2100 UTC 22 Aug 2021	1000 UTC 23 Aug 2021	2.76	25.30	0.109	

The investigations into various gauge sites that were flagged within the temporal QC algorithm demonstrated various modes of failure and various degrees of impact on the history of observations. The examples for PHIW2 (clogging; Fig. 14a) and TDGL1 (platform stability; Fig. 16a) depict how a further investigation of gauge sites classified as SUSP could identify a potential instrumental factor that resulted in errors. And while a gauge might be classified as SUSP across a 30-day analysis period, events that highlight the potential instrumental error with the gauge could possibly be seen in the 7-day analysis. The 7-day period ending 2300 UTC 3 September 2021 that followed the first residual dripping event from the

clogging had 29.6% of available hourly observations that were marked as FAIL, which resulted in a final BAD temporal QC classification. A 30-day analysis matrix can also display patterns in data that could characterize the potential instrumentation malfunction, such as the residual precipitation drips from clogging (e.g., Fig. 14).

6. Summary

This paper presented a temporal gauge QC methodology to identify potential gauge instrumentation malfunctions based on the history of the gauge observational and QC
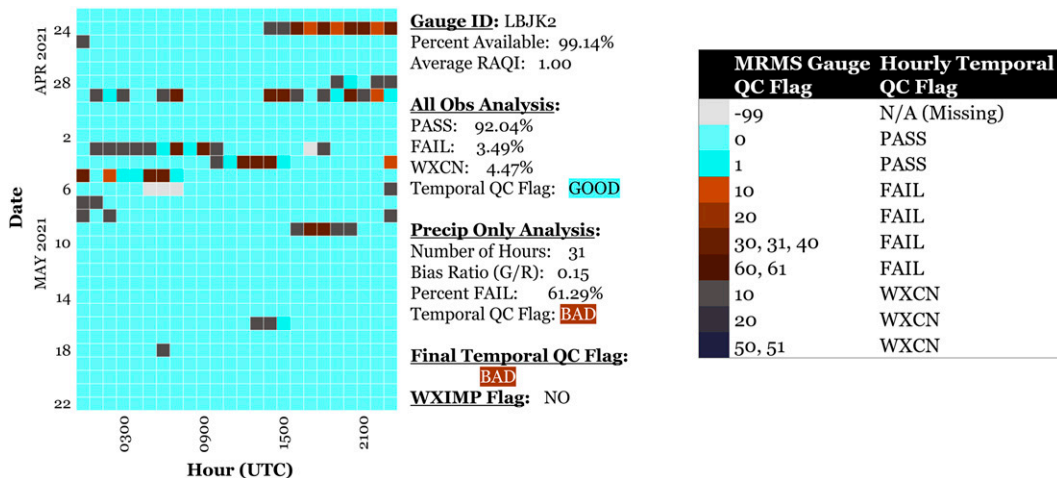

FIG. 15. As in Fig. 14, but for the gauge LBJK2 that experienced a clogging issue that impacted the precipitation-only periods. The matrix period for LBJK2 is from 0000 UTC 23 Apr to 2300 UTC 22 May 2021.
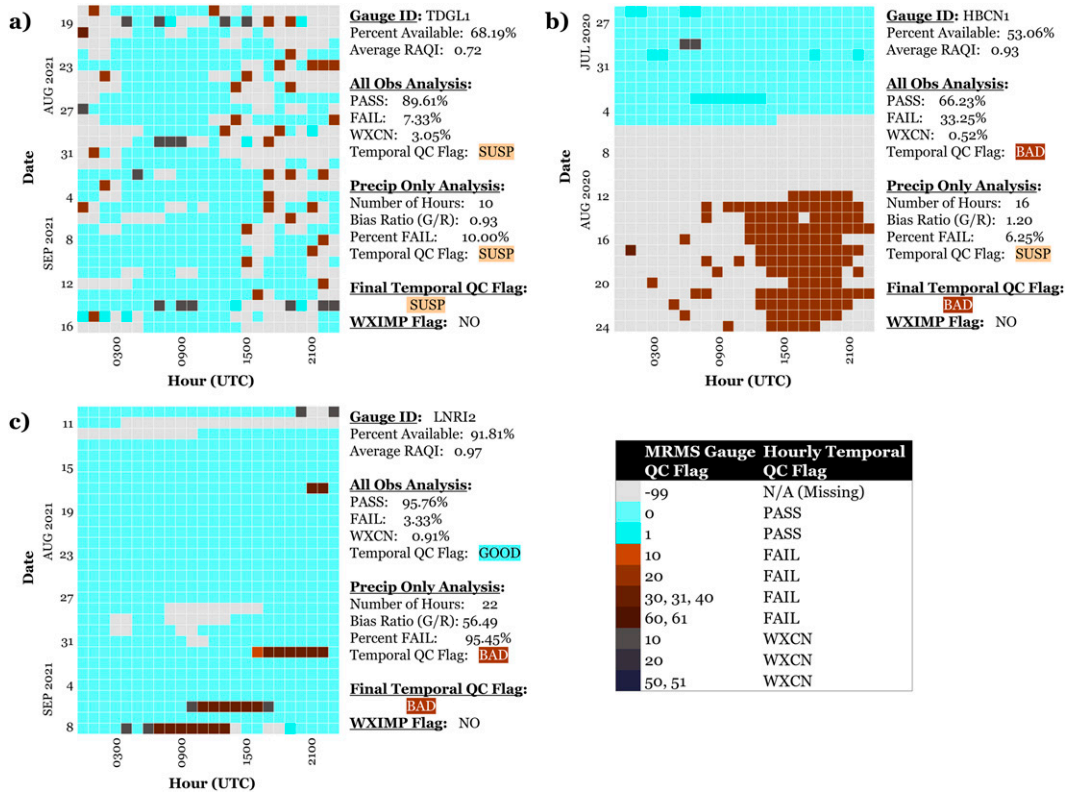
FIG. 16. As in Fig. 14, but for gauges that were impacted by other instrumental factors not related to the clogging of a gauge. The gauges shown in the analysis are (a) TDGL1 from 0000 UTC 18 Aug to 2300 UTC 16 Sep 2021, (b) HBCN1 from 0000 UTC 26 Jul to 2300 UTC 24 Aug 2020, and (c) LNRI2 from 0000 UTC 10 Aug to 2300 UTC 8 Sep 2021.

performance. This temporal QC scheme utilized the hourly gauge QC within the MRMS system to create initial hourly classifications. Those hourly classifications were analyzed over running 7- and 30-day periods using two groups of data: 1) all available observations and 2) observations where both the gauge and collocated gridded radar QPE accumulations were nonzero. The importance of having two different temporal analysis periods and two groups of gauge data allowed for a detailed examination from various perspectives.

The temporal QC logic can provide the ability to diagnose gauge performance history through quantified measures. The gauge analyses were characterized by a generally consistent BAD temporal QC classification with an average of 2.9% h^{-1} for the running 7-day analysis period and 4.4% for the running 30-day analysis period. A temporal QC scheme can be beneficial to network operators for the monitoring of gauge performance and identifying performance degradation prior to routine maintenance; moreover, these benefits allow for more prompt gauge maintenance and for improved product generation and verification through more high-quality observations.

While the automated temporal QC scheme presented here provides a baseline analysis for identifying the potential for instrumentation malfunctions through the gauge performance, the authors acknowledge that some aspects can benefit from additional research. Future sensitivity studies could explore a more optimum temporal analysis period that could best detect gauges with degraded performance from an instrumentation malfunction. Some gauge sites and gauge observations that were classified as BAD might not have an instrumentation malfunction. This could be contributed to biases and limitations with the collocated gridded radar QPE, the hourly MRMS gauge QC scheme, or with the temporal QC scheme presented in this study. The increase in the additional WXIMP classifications during the cool season was unsurprising and provided a useful insight that most gauges cannot handle winter precipitation, which is more a function of the design limitations of the gauge. How a gauge was classified with weather-related impacts to the observations highlighted the challenges in distinguishing between nominal functioning gauges and those with an instrumentation malfunction. This was especially true with instances of postwinter precipitation thawing. Future studies will focus on delineating between false nonzero observations that were a result of winter precipitation thawing or of another factor. Additional work can also investigate instances when truly erroneous observations were masked by the weather impacts classification (e.g., false zero observations during light rain) and how to mitigate those misclassifications.

Data availability statement. Hourly gauge observations from the MADIS (https://madis.noaa.gov/index.shtml) and HADS (https://hads.ncep.noaa.gov/) networks were obtained through NCEP Central Operations with permission. Hourly gauge observations from SNOTEL (https://www.wcc.nrcs.usda.gov/snow/), FCDMC (https://www.fcd.maricopa.gov/5308/Flood-Control-District), and HCFCD (https://www.hcfcd.org/) were obtained directly by NSSL from their respective sources. The observation ingest from the MADIS data feeds includes files from distribution category 2 (distribution to government, research, and educational organizations) and distribution category 4 (distribution to NOAA only). MADIS data restrictions prohibit the redistribution of data from these categories due to proprietary data. The ability to access the MADIS observational data requires permission from NOAA, per the MADIS data restriction web page (https://madis.noaa.gov/madis_restrictions.shtml). MRMS system products utilized in this study were developed and archived internally at NSSL. MRMS datasets stored internally at NSSL are available upon request.

REFERENCES

Allerup, P., and H. Madsen, 1980: Accuracy of point precipitation measurements. *Hydrol. Res.*, **11**, 57–70, https://doi.org/10.2166/nh.1980.0005.

Avanzi, F., C. De Michele, A. Ghezzi, C. Jommi, and M. Pepe, 2014: A processing-modeling routine to use SNOTEL hourly data in snowpack dynamic models. *Adv. Water Resour.*, **73**, 16–29, https://doi.org/10.1016/j.advwatres.2014.06.011.

Båserud, L., C. Lussana, T. N. Nipen, I. A. Seierstad, L. Oram, and T. Aspelien, 2020: TITAN automatic spatial quality control of meteorological in-situ observations. *Adv. Sci. Res.*, **17**, 153–163, https://doi.org/10.5194/asr-17-153-2020.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Boudala, F. S., G. A. Isaac, P. Filman, R. Crawford, D. Hudak, and M. Anderson, 2017: Performance of emerging technologies for measuring solid and liquid precipitation in cold climate as compared to the traditional manual gauges. *J. Atmos. Oceanic Technol.*, **34**, 167–185, https://doi.org/10.1175/JTECH-D-16-0088.1.

Brandes, E. A., 1975: Optimizing rainfall estimates with the aid of radar. *J. Appl. Meteor.*, **14**, 1339–1345, https://doi.org/10.1175/1520-0450(1975)014<1339:OREWTA>2.0.CO;2.

Carvalho, L. M. V., 2020: Assessing precipitation trends in the Americas with historical data: A review. *Wiley Interdiscip. Rev.: Climate Change*, **11**, e627, https://doi.org/10.1002/wcc.627.

Chen, M., P. Xie, J. E. Janowiak, and P. A. Arkin, 2002: Global land precipitation: A 50-yr monthly analysis based on gauge observations. *J. Hydrometeor.*, **3**, 249–266, https://doi.org/10.1175/1525-7541(2002)003<0249:GLPAYM>2.0.CO;2.

Chubb, T., M. J. Manton, S. T. Siems, A. D. Peace, and S. P. Bilish, 2015: Estimation of wind-induced losses from a precipitation gauge network in the Australian snowy mountains. *J. Hydrometeor.*, **16**, 2619–2638, https://doi.org/10.1175/JHM-D-14-0216.1.

Cocks, S. B., and Coauthors, 2019: A prototype quantitative precipitation estimation algorithm for operational S-band polarimetric radar utilizing specific attenuation and specific differential phase. Part II: Performance verification and case study analysis. *J. Hydrometeor.*, **20**, 999–1014, https://doi.org/10.1175/JHM-D-18-0070.1.

Derin, Y., and Coauthors, 2016: Multiregional satellite precipitation products evaluations over complex terrain. *J. Hydrometeor.*, **17**, 1817–1836, https://doi.org/10.1175/JHM-D-15-0197.1.

Duchon, C. E., 2008: Using vibrating-wire technology for precipitation measurements. *Precipitation: Advances in Measurement, Estimation and Prediction*, S. C. Michaelides, Ed., Springer, 33–58.

——, and G. R. Essenberg, 2001: Comparative rainfall observations from pit and aboveground gauges with and without wind shields. *Water Resour. Res.*, **37**, 3253–3263, https://doi.org/10.1029/2001WR000541.

——, and C. J. Biddle, 2010: Undercatch of tipping-bucket gauges in high rain rate events. *Adv. Geosci.*, **25**, 11–15, https://doi.org/10.5194/adgeo-25-11-2010.

——, C. Fiebrich, and D. Grimsley, 2014: Using high-speed photography to study undercatch in tipping-bucket rain gauges. *J. Atmos. Oceanic Technol.*, **31**, 1330–1336, https://doi.org/10.1175/JTECH-D-13-00169.1.

Førland, E. J., and Coauthors, 1996: Manual for operational correction of Nordic precipitation data. Norwegian Meteorological Institute Rep. 24/96, 66 pp.

Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation measurement intercomparison. WMO Instruments and Observing Methods Rep. 67, 212 pp., https://library.wmo.int/doc_num.php?explnum_id=9694.

Gourley, J. J., D. P. Jorgensen, S. Y. Matrosov, and Z. L. Flamig, 2009: Evaluation of incremental improvements to quantitative precipitation estimates in complex terrain. *J. Hydrometeor.*, **10**, 1507–1520, https://doi.org/10.1175/2009JHM1125.1.

Gowan, T. A., and J. D. Horel, 2020: Evaluation of IMERG-E precipitation estimates for fire weather applications in Alaska. *Wea. Forecasting*, **35**, 1831–1843, https://doi.org/10.1175/WAF-D-20-0023.1.

Groisman, P. Y., and D. R. Legates, 1994: The accuracy of United States precipitation data. *Bull. Amer. Meteor. Soc.*, **75**, 215–228, https://doi.org/10.1175/1520-0477(1994)075<0215:TAOUSP>2.0.CO;2.

Habib, E., W. F. Krajewski, and A. Kruger, 2001: Sampling errors of tipping-bucket rain gauge measurements. *J. Hydrol. Eng.*, **6**, 159–166, https://doi.org/10.1061/(ASCE)1084-0699(2001)6:2(159).

Helms, D., P. A. Miller, M. F. Barth, D. Starosta, B. Gordon, S. Schofield, F. Kelly, and S. Koch, 2009: Status update of the transition from research to operations of the Meteorological Assimilation Data Ingest System (MADIS). *25th Conf. on Int. Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Phoenix, AZ, Amer. Meteor. Soc., 5A.3, https://ams.confex.com/ams/89annual/techprogram/paper_149883.htm.

Hulme, M., and M. New, 1997: Dependence of large-scale precipitation climatologies on temporal and spatial sampling. *J. Climate*, **10**, 1099–1113, https://doi.org/10.1175/1520-0442(1997)010<1099:DOLSPC>2.0.CO;2.

Kim, D., B. Nelson, and D.-J. Seo, 2009: Characteristics of reprocessed Hydrometeorological Automated Data System (HADS) hourly precipitation data. *Wea. Forecasting*, **24**, 1287–1296, https://doi.org/10.1175/2009WAF2222227.1.

Kochendorfer, J., and Coauthors, 2017: The quantification and correction of wind-induced precipitation measurement errors. *Hydrol. Earth Syst. Sci.*, **21**, 1973–1989, https://doi.org/10.5194/hess-21-1973-2017.

Kondragunta, C. R., and K. Shrestha, 2006: Automated real-time operational rain gauge quality-control tools in NWS hydrologic operations. *20th Conf. on Hydrology*, Boston, MA, Amer. Meteor. Soc., P2.4, https://ams.confex.com/ams/Annual2006/techprogram/paper_102834.htm.

Laber, J. A., 2020: An overview of the 9 January 2018 extreme flash flood and debris flow event in Montecito, California. *34th Conf. on Hydrology*, Boston, MA, Amer. Meteor. Soc., 6A.3, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/362056.

Lanza, L. G., and A. Cauteruccio, 2022: Accuracy assessment and intercomparison of precipitation instruments. *Precipitation Science—Measurement, Remote Sensing, Microphysics and Modeling*, S. Michaelides, Ed., Elsevier, 3–35.

Larson, L. W., 1993: ASOS heated tipping bucket precipitation gage (FRIEZ) evaluation at WSFO, Bismarck, North Dakota, March 1992–March 1993. NWS Central Region Final Rep., 54 pp.

Leeper, R. D., M. A. Palecki, and E. Davis, 2015: Methods to calculate precipitation from weighing-bucket gauges with redundant depth measurements. *J. Atmos. Oceanic Technol.*, **32**, 1179–1190, https://doi.org/10.1175/JTECH-D-14-00185.1.

Liao, M., J. Liu, A. Liao, Z. Cai, Y. Huang, P. Zhuo, and X. Li, 2020: Investigation of tipping-bucket rain gauges using digital photographic technology. *J. Atmos. Oceanic Technol.*, **37**, 327–339, https://doi.org/10.1175/JTECH-D-19-0064.1.

Lincoln, W. S., R. F. L. Thomason, M. Stackhouse, and D. S. Schlotzhauer, 2017: Utilizing crowd-sourced rainfall and flood impact information to improve the analysis of the north central Gulf Coast flood event of April 2014. *J. Oper. Meteor.*, **5**, 26–41, https://doi.org/10.15191/nwajom.2017.0503.

Maksimović, Č., L. Bužek, and J. Petrović, 1991: Corrections of rainfall data obtained by tipping bucket rain gauge. *Atmos. Res.*, **27**, 45–53, https://doi.org/10.1016/0169-8095(91)90005-H.

Manz, B., S. Páez-Bimos, N. Horna, W. Buytaert, B. Ochoa-Tocachi, W. Lavado-Casimiro, and B. Willems, 2017: Comparative ground validation of IMERG and TMPA at variable spatiotemporal scales in the tropical Andes. *J. Hydrometeor.*, **18**, 2469–2489, https://doi.org/10.1175/JHM-D-16-0277.1.

Martinaitis, S. M., S. B. Cocks, Y. Qi, B. T. Kaney, J. Zhang, and K. Howard, 2015: Understanding winter precipitation impacts on automated gauge observations within a real-time system. *J. Hydrometeor.*, **16**, 2345–2363, https://doi.org/10.1175/JHM-D-15-0020.1.

——, H. M. Grams, C. Langston, J. Zhang, and K. Howard, 2018: A real-time evaporation correction scheme for radar-derived mosaicked precipitation estimations. *J. Hydrometeor.*, **19**, 87–111, https://doi.org/10.1175/JHM-D-17-0093.1.

——, S. B. Cocks, M. J. Simpson, A. P. Osborne, S. S. Harkema, H. M. Grams, J. Zhang, and K. W. Howard, 2021: Advancements and characteristics of gauge ingest and quality control

within the Multi-Radar Multi-Sensor system. *J. Hydrometeor.*, **22**, 2455–2474, https://doi.org/10.1175/JHM-D-20-0234.1.

Mascaro, G., 2017: Multiscale spatial and temporal statistical properties of rainfall in central Arizona. *J. Hydrometeor.*, **18**, 227–245, https://doi.org/10.1175/JHM-D-16-0167.1.

Matsuo, T., and Y. Sasyo, 1981: Non-melting phenomena of snowflakes observed in subsaturated air below freezing level. *J. Meteor. Soc. Japan*, **59**, 26–32, https://doi.org/10.2151/jmsj1965.59.1_26.

Metcalfe, J. R., and B. E. Goodison, 1992: Automation of winter precipitation measurements: The Canadian experience. *Proc. WMO Technical Conf. on Instruments and Methods of Observation*, Vienna, Austria, WMO, WMO/TD-462, 81–85.

——, B. Routledge, and K. Devine, 1997: Rainfall measurement in Canada: Changing observational methods and archive adjustment procedures. *J. Climate*, **10**, 92–101, https://doi.org/10.1175/1520-0442(1997)010<0092:RMICCO>2.0.CO;2.

Molini, A., L. G. Lanza, and P. La Barbera, 2005: Improving the accuracy of tipping-bucket rain records using disaggregation techniques. *Atmos. Res.*, **77**, 203–217, https://doi.org/10.1016/j.atmosres.2004.12.013.

Nešpor, V., and B. Sevruk, 1999: Estimation of wind-induced error of rainfall gauge measurements using a numerical simulation. *J. Atmos. Oceanic Technol.*, **16**, 450–464, https://doi.org/10.1175/1520-0426(1999)016<0450:EOWIEO>2.0.CO;2.

NOAA, 2019: Weather Forecast Office water resources products specification. National Weather Service Manual 10-922, 97 pp., https://www.nws.noaa.gov/directives/sym/pd01009022curr.pdf.

Nystuen, J. A., 1998: Temporal sampling requirements for automatic rain gauges. *J. Atmos. Oceanic Technol.*, **15**, 1253–1260, https://doi.org/10.1175/1520-0426(1998)015<1253:TSRFAR>2.0.CO;2.

——, J. R. Proni, P. G. Black, and J. C. Wilkerson, 1996: A comparison of automatic rain gauges. *J. Atmos. Oceanic Technol.*, **13**, 62–73, https://doi.org/10.1175/1520-0426(1996)013<0062:ACOARG>2.0.CO;2.

Ośródka, K., I. Otop, and J. Szturc, 2022: Automatic quality control of telemetric rain gauge data providing quantitative quality information (RainGaugeQC). *Atmos. Meas. Tech.*, **15**, 5581–5597, https://doi.org/10.5194/amt-15-5581-2022.

Osterhuber, R. S., T. Edens, and B. J. McGurk, 1994: Snow depth measurement using ultrasonic sensors and temperature correction. *Proc. 62nd Annual Western Snow Conf.*, Santa Fe, NM, Western Snow Conference, 159–162, https://westernsnowconference.org/sites/westernsnowconference.org/PDFs/1994Osterhuber.pdf.

Parsons, D. A., 1941: Calibration of a weather bureau tipping-bucket raingage. *Mon. Wea. Rev.*, **69**, 205, https://doi.org/10.1175/1520-0493(1941)069<0205:COAWBT>2.0.CO;2.

Perica, S., and Coauthors, 2013: *Southeastern States (Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi)*. Vol. 9, version 2.0, *Precipitation-Frequency Atlas of the United States*, NOAA Atlas 14, 163 pp.

Pollock, M. D., and Coauthors, 2018: Quantifying and mitigating wind-induced undercatch in rainfall measurements. *Water Resour. Res.*, **54**, 3863–3875, https://doi.org/10.1029/2017WR022421.

Qi, Y., J. Zhang, and P. Zhang, 2013a: A real-time automated convective and stratiform precipitation segregation algorithm in native radar coordinates. *Quart. J. Roy. Meteor. Soc.*, **139**, 2233–2240, https://doi.org/10.1002/qj.2095.

——, ——, ——, and Q. Cao, 2013b: VPR correction of bright band effects in radar QPEs using polarimetric radar observations. *J. Geophys. Res. Atmos.*, **118**, 3627–3633, https://doi.org/10.1002/jgrd.50364.

Rabiei, E., and U. Haberlandt, 2015: Applying bias correction for merging rain gauge and radar data. *J. Hydrol.*, **522**, 544–557, https://doi.org/10.1016/j.jhydrol.2015.01.020.

Rasmussen, R., and Coauthors, 2012: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc.*, **93**, 811–829, https://doi.org/10.1175/BAMS-D-11-00052.1.

Rosenfeld, D., D. Atlas, D. B. Wolff, and E. Amitai, 1992: Beamwidth effects on *Z–R* relations and area-integrated rainfall. *J. Appl. Meteor. Climatol.*, **31**, 454–464, https://doi.org/10.1175/1520-0450(1992)031<0454:BEOZRR>2.0.CO;2.

Ryzhkov, A., M. Diederich, P. Zhang, and C. Simmer, 2014: Potential utilization of specific attenuation for rainfall estimation, mitigation of partial beam blockage, and radar networking. *J. Atmos. Oceanic Technol.*, **31**, 599–619, https://doi.org/10.1175/JTECH-D-13-00038.1.

Savina, M., B. Schäppi, P. Molnar, P. Burlando, and B. Sevruk, 2012: Comparison of a tipping-bucket and electronic weighing precipitation gauge for snowfall. *Atmos. Res.*, **103**, 45–51, https://doi.org/10.1016/j.atmosres.2011.06.010.

Seo, D.-J., and J. P. Breidenbach, 2002: Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *J. Hydrometeor.*, **3**, 93–111, https://doi.org/10.1175/1525-7541(2002)003<0093:RTCOSN>2.0.CO;2.

Serreze, M. C., M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. S. Pulwarty, 1999: Characteristics of the western United States snowpack from Snowpack Telemetry (SNOTEL) data. *Water Resour. Res.*, **35**, 2145–2160, https://doi.org/10.1029/1999WR900090.

Sevruk, B., 1982: Method of correction for systematic error in point precipitation measurement for operational use. WMO Tech. Note 589, 91 pp., https://library.wmo.int/doc_num.php?explnum_id=1238.

——, 1989: Wind-induced measurement error for high intensity rains. *Proc. Int. Workshop on Precipitation Measurements*, Saint Moritz, Switzerland, ETH Zurich, 199–204, http://www.wmo.ch.

——, 1996: Adjustment of tipping-bucket precipitation gauge measurements. *Atmos. Res.*, **42**, 237–246, https://doi.org/10.1016/0169-8095(95)00066-6.

Sieck, L. C., S. J. Burges, and M. Steiner, 2007: Challenges in obtaining reliable measurements of point rainfall. *Water Resour. Res.*, **43**, W01420, https://doi.org/10.1029/2005WR004519.

Smith, C. J., 1986: The reduction of errors caused by bright bands in quantitative rainfall measurements made using radar. *J. Atmos. Oceanic Technol.*, **3**, 129–141, https://doi.org/10.1175/1520-0426(1986)003<0129:TROECB>2.0.CO;2.

Steiner, M., J. A. Smith, S. J. Burges, C. V. Alonso, and R. W. Darden, 1999: Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation. *Water Resour. Res.*, **35**, 2487–2503, https://doi.org/10.1029/1999WR900142.

Sypka, P., 2019: Dynamic real-time volumetric correction for tipping-bucket rain gauges. *Agric. For. Meteor.*, **271**, 158–167, https://doi.org/10.1016/j.agrformet.2019.02.044.

Tabary, P., J. Desplats, K. Do Khac, F. Eideliman, C. Gueguen, and J.-C. Heinrich, 2007: The new French operational radar rainfall product. Part II: Validation. *Wea. Forecasting*, **22**, 409–427, https://doi.org/10.1175/WAF1005.1.

Tang, L., J. Zhang, C. Langston, J. Krause, K. Howard, and V. Lakshmanan, 2014: A physically based precipitation–nonprecipitation radar echo classifier using polarimetric and environmental data in a real-time national system. *Wea. Forecasting*, **29**, 1106–1119, https://doi.org/10.1175/WAF-D-13-00072.1.

——, ——, M. Simpson, A. Arthur, H. Grams, Y. Wang, and C. Langston, 2020: Updates on the radar data quality control in the MRMS quantitative precipitation estimation system. *J. Atmos. Oceanic Technol.*, **37**, 1521–1537, https://doi.org/10.1175/JTECH-D-19-0165.1.

Upton, G. J. G., and A. R. Rahimi, 2003: On-line detection of errors in tipping-bucket raingauges. *J. Hydrol.*, **278**, 197–212, https://doi.org/10.1016/S0022-1694(03)00142-2.

Ushio, T., T. Matsuda, T. Tashima, T. Kubota, M. Kachi, and S. Yoshida, 2013: Gauge adjusted Global Satellite Mapping of Precipitation (GSMaP_Gauge). *Proc. 29th Int. Symp. on Space Technology and Science*, Nagoya, Japan, JAXA, 2013-n-48.

Wang, Y., S. Cocks, L. Tang, A. Ryzhkov, P. Zhang, J. Zhang, and K. Howard, 2019: A prototype quantitative precipitation estimation algorithm for operational S-band polarimetric radar utilizing specific attenuation and specific differential phase. Part I: Algorithm description. *J. Hydrometeor.*, **20**, 985–997, https://doi.org/10.1175/JHM-D-18-0071.1.

Wijayarathne, D., S. Boodoo, P. Coulibaly, and D. Sills, 2020: Evaluation of radar quantitative precipitation estimates (QPEs) as an input of hydrological models for hydrometeorological applications. *J. Hydrometeor.*, **21**, 1847–1864, https://doi.org/10.1175/JHM-D-20-0033.1.

Wilson, J. W., and E. A. Brandes, 1979: Radar measurement of rainfall—A summary. *Bull. Amer. Meteor. Soc.*, **60**, 1048–1060, https://doi.org/10.1175/1520-0477(1979)060<1048:RMORS>2.0.CO;2.

Wolff, D. B., W. A. Petersen, A. Tokay, D. A. Marks, and J. L. Pippitt, 2019: Assessing dual-polarization radar estimates of extreme rainfall during Hurricane Harvey. *J. Atmos. Oceanic Technol.*, **36**, 2501–2520, https://doi.org/10.1175/JTECH-D-19-0081.1.

Yang, D., B. E. Goodison, J. R. Metcalfe, V. S. Golubev, R. Bataes, T. Pangburn, and C. L. Hanson, 1998: Accuracy of NWS 8″ standard nonrecording precipitation gauge: Results and application of WMO intercomparison. *J. Atmos. Oceanic Technol.*, **15**, 54–68, https://doi.org/10.1175/1520-0426(1998)015<0054:AONSNP>2.0.CO;2.

Ye, B., D. Yang, Y. Ding, T. Han, and T. Koike, 2004: A bias-corrected precipitation climatology for China. *J. Hydrometeor.*, **5**, 1147–1160, https://doi.org/10.1175/JHM-366.1.

Yeung, H. Y., C. Man, S. T. Chan, and A. Seed, 2014: Development of an operational rainfall data quality-control scheme based on radar-raingauge co-kriging analysis. *Hydrol. Sci. J.*, **59**, 1293–1307, https://doi.org/10.1080/02626667.2013.839873.

Young, C. B., B. R. Nelson, A. A. Bradley, J. A. Smith, C. D. Peters-Lidard, A. Kruger, and M. L. Baeck, 1999: An evaluation of NEXRAD precipitation estimates in complex terrain. *J. Geophys. Res.*, **104**, 19 691–19 703, https://doi.org/10.1029/1999JD900123.

Zhang, J., and Y. Qi, 2010: A real-time algorithm for the correction of brightband effects in radar-derived QPE. *J. Hydrometeor.*, **11**, 1157–1171, https://doi.org/10.1175/2010JHM1201.1.

——, ——, D. Kingsmill, and K. Howard, 2012: Radar-based quantitative precipitation estimation for the cool season in complex terrain: Case studies from the NOAA Hydrometeorology Testbed. *J. Hydrometeor.*, **13**, 1836–1854, https://doi.org/10.1175/JHM-D-11-0145.1.

——, and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

——, L. Tang, S. Cocks, P. Zhang, A. Ryzhkov, K. Howard, C. Langston, and B. Kaney, 2020a: A dual-polarization radar synthetic QPE for operations. *J. Hydrometeor.*, **21**, 2507–2521, https://doi.org/10.1175/JHM-D-19-0194.1.

——, and Coauthors, 2020b: High-resolution QPE products from the experimental MRMS system. *36th Conf. on Environmental Information Processing Technologies*, Boston, MA, Amer. Meteor. Soc., 3A.3, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/368414.