Multiparameter Regression Modeling for Improving Quality of Measured Rainfall and Runoff Data in Densely Instrumented Watersheds

Menberu Meles Bitew, Ph.D., P.E., M.ASCE[1]; David C. Goodrich, Ph.D., P.E., M.ASCE[2]; Eleonora Demaria, Ph.D.[3]; Philip Heilman, Ph.D.[4]; Mary Nichols, Ph.D., P.E.[5]; Lainie Levick[6]; Carl L. Unkrich[7]; and Mark Kautz[8]

Abstract: The Walnut Gulch Experimental Watershed is a semi-arid experimental watershed and long-term agro-ecosystem research (LTAR) site managed by the USDA-Agricultural Research Services (ARS) Southwest Watershed Research Center for which high-resolution, long-term hydroclimatic data are available across its 149-km^2 drainage area. Quality control and quality assurance of the massive data set are a major challenge. We present the analysis of 50 years of data sets to develop a strategy to identify errors and inconsistencies in historical rainfall and runoff databases. A multiple regression model was developed to relate rainfall, watershed properties, and the antecedent conditions to runoff characteristics in 12 subwatersheds ranging in area from 0.002–94 km^2. A regression model was developed based on 18 predictor variables, which produced predicted runoff with correlation coefficients ranging from 0.4–0.94 and Nash efficiency coefficients up to 0.76. The model predicted 92% of runoff events and 86% of no-runoff events. The modeling approach is a complement to existing quality assurance and quality control (QAQC) procedures and provides a specific method for ensuring that rainfall and runoff data in the USDA-ARS Walnut Gulch Experimental Watershed database are consistent and contain minimal error. The model has the potential for making runoff predictions in similar hydroclimatic environments with available high-resolution observations. **DOI: 10.1061/(ASCE)HE.1943-5584.0001825.** *© 2019 American Society of Civil Engineers.*

Author keywords: Walnut gulch; Rainfall-runoff; Multiparameter regression model; Quality assurance and quality control (QAQC); Database; Uncertainty; Semiarid.

Introduction

Watershed and natural resource research observatories such as the USDA, Agricultural Research Service (ARS), and Forest Service Experimental Watersheds, Critical Zone Observatories (CZO—Brantley et al. 2017), German Terrestrial Environmental Observatories (TERENO—Zacharias et al. 2011), and National Ecological Observatory Network (NEON) sites with ongoing or planned long-term operations are on the rise. Quality control and quality assurance of the data collected at these observatories are critical for research and analyses relying on them. The densely instrumented USDA-ARS Walnut Gulch Experimental Watershed (WGEW) long-term agroecosystem research (LTAR) site has been in operation since the mid-1950s, collecting an array of hydroclimatic observations. Quality Assurance and Quality Control (QAQC) procedures for the WGEW and USDA-ARS Watershed Experimental Network have evolved as observations transition from those made by analog (Brakensiek et al. 1979) to digital instrumentation (Moran et al. 2008a). The Data Access Project (DAP) for the WGEW was initiated shortly after the transition from analog to digital instrumentation in 2000 (Nichols and Anson 2008).

The DAP revisited prior QAQC procedures and developed new QAQC procedures applicable to the digital collection and transmission of hydroclimatic observations. The database consists of data available for public use that have passed through specific quality checks. Historically, hydrographs with problematic recession curves and erroneous time stamps associated with analog clocks were typically corrected before entering the database. However, errors are not uncommon in DAP because several versions of equipment were

[1]Research Civil Engineer, US Dept. of Agriculture–Agricultural Research Service (USDA–ARS), Southwestern Watershed Research Center, Tucson, AZ 85719 (corresponding author). ORCID: https://orcid.org/0000-0002-1980-9243. Email: Menberu.Bitew@ars.usda.gov

[2]Research Hydraulic Engineer, US Dept. of Agriculture–Agricultural Research Service (USDA–ARS), Southwestern Watershed Research Center, Tucson, AZ 85719. Email: Dave.Goodrich@ars.usda.gov

[3]Research Hydrologist-Meteorologist, US Dept. of Agriculture–Agricultural Research Service (USDA–ARS), Southwestern Watershed Research Center, Tucson, AZ 85719. Email: eleonora.demaria@ars.usda.gov

[4]Research Leader, US Dept. of Agriculture–Agricultural Research Service (USDA–ARS), Southwestern Watershed Research Center, Tucson, AZ 85719. Email: Phil.Heilman@ars.usda.gov

[5]Research Hydraulic Engineer, US Dept. of Agriculture–Agricultural Research Service (USDA–ARS), Southwestern Watershed Research Center, Tucson, AZ 85719. Email: Mary.Nichols@ars.usda.gov

[6]Principal Research Specialist, School of Natural Resources and the Environment, Univ. of Arizona, Tucson, AZ 85721. ORCID: https://orcid.org/0000-0003-4727-0836. Email: llevick@email.arizona.edu

[7]Hydrologist, US Dept. of Agriculture–Agricultural Research Service (USDA–ARS), Southwestern Watershed Research Center, Tucson, AZ 85719. Email: Carl.Unkrich@ars.usda.gov

[8]Hydrologist, US Dept. of Agriculture–Agricultural Research Service (USDA–ARS), Southwestern Watershed Research Center, Tucson, AZ 85719. Email: Mark.Kautz@ars.usda.gov

Note. This manuscript was submitted on September 19, 2018; approved on April 17, 2019; published online on July 25, 2019. Discussion period open until December 25, 2019; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699.

used and as personnel changed over time, human errors were inevitable at various levels of data reading, storage, and digitizing paper charts. The current challenge is to find new ways to improve data quality by identifying and correcting errors within the large number of rainfall and runoff events since the 1950s.

The fundamental role that the spatiotemporal distribution of rainfall plays in semiarid runoff generation (Osborn and Hickok 1968; Osborn et al. 1980) is exploited to further improve QAQC for both rainfall and runoff observations. Precipitation properties such as the total depth, intensity, duration, and precipitation scale with respect to the watershed contributing area and its spatial structure across the basin are important for quantifying watershed response such as runoff rate and the timing of peak flow. The dense recording rain gauge network in WGEW provides high temporal and spatial resolution data sets (e.g., Goodrich et al. 2008a, b; Keefer et al. 2008) to draw upon. Syed et al. (2003) analyzed the correlation between spatially interpolated rainfall data and basin runoff response from the WGEW and found that the size of the storm core, the area with intensities greater than 25 $\mathrm{mm\,h^{-1}}$, and the position of the storm core with respect to the watershed outlet were good predictors of runoff volume and rate.

WGEW provides long-term rainfall and runoff data and more than half a century of investigations and experiments conducted across the nested watersheds in hydrology, ecology, and rangeland studies, which provides a broad set of data and proxies related to rainfall-runoff relations. The database has been used extensively by the hydroecological modeling community for watershed and process model validation (Costa et al. 2012; Duan et al. 2006; Goodrich et al. 2004, 2012; Niu et al. 2014; Scott and Biederman 2017) and by the remote sensing community (ground and satellite) as a key ground validation site for retrieval algorithm validation (Amitai et al. 2012; Das et al. 2008; Houser et al. 1998; Knipper et al. 2017; Kolassa et al. 2018; Moran et al. 2008b; Morin et al. 2003). Therefore, it is critical that its observational database be as free of errors as possible.

Observations using 88 weighing rain gauges and runoff observations from 16 flumes across a range of nested watershed sizes in WGEW provided a unique opportunity to derive potential predictors for quantifying runoff and identifying observational inconsistencies (see Fig. 1 for the location of the rain gauges and flumes). Here, we show how long-term, high-resolution, continuous rainfall and runoff data at hourly time steps was used to develop a regression-based predictive model for WGEW that can be used to identify observations that are inconsistent or erroneous with the observed rainfall-runoff responses of the watershed. The systematic removal or correction of erroneous data will improve analyses that usually rely on these data such as increasing consistencies in rainfall-runoff model parameterization and minimizing error in water balance–related analysis. This work includes descriptions of 22 parameters related to rainfall, watershed, and antecedent conditions obtained from the WGEW database and other national databases. These data sets allowed the development of multiple predictors of runoff related to precipitation and watershed properties, antecedent conditions, and the temporal information of events such as the diurnal and seasonal properties of the rainfall fields. Because dominant, high-intensity, runoff-producing precipitation occurs in the summer at the WGEW, the focus of the study was limited to summer data (June–October).

Regression-based hydrologic predictions are one of the oldest tools in predictive hydrology (e.g., Bridges 1982; Chow 1964; Cochran et al. 1979; McCain and Jarrett 1976). Polyakov et al. (2010) used the detailed observations of rainfall and runoff in Santa Rita Experimental Range watersheds to develop watershed-scale sediment yield predictive models. Using data from WGEW, Osborn and Lane (1969) developed a regression model using 10 rainfall gauges, antecedent conditions, and watershed parameters to show how rainfall properties explain runoff volume, duration of flow, and lag time. Lane et al. (1971) also applied a regression model to describe routing and channel transmission losses using input and output hydrograph parameters derived from WGEW data sets. Regression models using rainfall characteristics, watershed properties, and antecedent conditions have also been applied in many regions; for example, watersheds in India and Oman (e.g., McIntyre et al. 2007; Sharma and Murthy 1996, 1998) used regression-based models to predict hydrologic transmission losses, runoff volume, peak flow, and time to peak. In recent years, different techniques such as machine learning, adaptive data analysis methodology, artificial neural networks, and computational intelligence (e.g., Fotovatikhah et al. 2018; Haupt and Kosovic 2015; Taormina et al. 2015; Wu and Chau 2011; ASCE 2000a, b) have been successfully applied to a wide range of hydrologic problems and may also show functionality in QAQC of hydrologic databases. Despite the computational efficiency and ease of implementation of regression models, their use for QAQC has not been explored. An additional goal is to provide a well-tested QAQC approach that could benefit other experimental watershed observatories with similar data sets, for example, H.J. Andrews, Coweeta, Casper Creek, Little Washita, Little River, Reynolds Creek, and so on, for curating long-term hydroclimatic data.

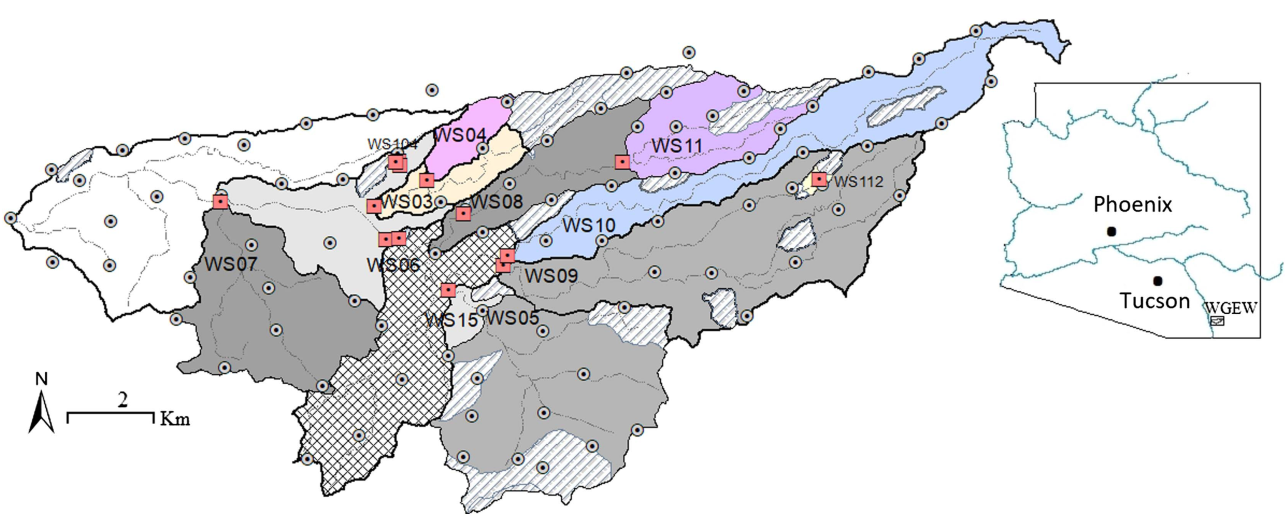

Fig. 1. Walnut Gulch Experimental Watershed (WGEW). Colors and textures represent the subwatershed data sources used for the development of the regression model. Watersheds on the map are referred to as WS06 instead of 63.006, and so on.

The goal of the study was to develop multiple regression models using long-term high-resolution data sets with the primary objectives of using the regression models (1) to identify erroneous precipitation and runoff events in the WGEW database, and (2) to improve the quality of legacy data sets as a preparatory step to process-based simulation studies by ensuring consistent rainfall and runoff relations.

Methods

Study Site

The study was conducted in the WGEW (Fig. 1), which is also a long-term agro-ecosystem research site in the southwest United States managed by the USDA-ARS Southwest Watershed Research Center, for which high-resolution, long-term hydroclimatic data are available. The watershed has a total drainage area of 149 km^2 with elevations ranging between 1,190 and 2,150 masl. Mean annual temperature in the city of Tombstone, located within the watershed, is 17.6°C, with a mean annual rainfall of 324 mm. Most rainfall occurs during the summer monsoon months from July through October. Currently, vegetative cover on the watershed is generally composed of two main vegetation communities, shrub dominated (about 20%) and grass dominated (15%), with the remaining consisting of a mixture of grass, shrubs, woodlands, trees, and bare ground (e.g., King et al. 2008; Skirvin et al. 2008). The spatial distribution of vegetation is closely linked to soil type and variations in annual rainfall. The soils are dominantly sandy, gravely loam that vary from deep, relatively mature, and well-drained soil to thin, immature soils. The underlying geology is a thick alluvial fan that drains to the San Pedro River (Osterkamp 2008; Renard et al. 1993). Because of the thickness and extent of the alluvial fill, the groundwater reserves are substantial and can be found at depths ranging from 50 to 145 m (Goodrich et al. 2004).

The WGEW has been instrumented since 1953 to quantify hydroclimatic variables [see 2008 WRR special issue; Moran et al. (2008a)—introductory paper]. Rainfall and runoff data for this study were collected from a selected 12 WGEW subwatersheds ranging from 0.002 to 94 km^2. A dense network of sensors captures the spatial structure of both rainfall and watershed runoff responses. Currently, runoff is measured with one v-notch weir, two H flumes, 11 large supercritical flumes (Smith et al. 1982), and five small supercritical flumes. Runoff instrumentation is located at the outlets of 16 nested subwatersheds. A total of 88 weighing-type recording rain gauges with a precision of 0.25 mm and 1-min time step (for digital data) are distributed within the watershed at a density of 1.7 rain gauges per square kilometer (Goodrich et al. 2008a; Stone et al. 2008). The largest subwatershed contains 56 rain gauges, and the smallest subwatershed contains 1 rain gauge. Data are curated and accessed through the Southwest Watershed Research Center—Data Access Project (SWRC-DAP) database (USDA-ARS n.d., Nichols and Anson 2008).

Data Sets

Hourly rainfall and runoff data from 1960 to 2016 were used for the analysis. Event data are available as breakpoint-formatted rainfall hyetographs and runoff hydrographs that include time and accumulated depth at slope breaks on analog strip charts (Goodrich et al. 2008a). Prior to January 1, 2000, the data were collected using analog instruments, which were upgraded to a digital electronic system in 2000. Details of the database can be found in Goodrich et al. (2008a), Keefer et al. (2008), and Nichols and Anson (2008).

A total of 22 parameters describing rainfall and spatially distributed subwatershed characteristics were logically grouped to create four subsets of properties that potentially influence runoff (Table 1). The four subsets were grouped as rainfall properties, watershed properties, antecedent conditions, and temporal properties, as described in the subsequent paragraphs. Most of those variables have long been understood to affect rainfall-runoff relations, but some of them, such as season, month, and time of the storm occurrences; stream density; and stream order, are introduced here as proxies for runoff quantity. Watershed physical properties such as area, average slope, flow length, and shape were derived from a 1×1-m-resolution grid that was aggregated from a 0.5×0.5-m light detection and ranging (LiDAR)-derived digital elevation model or DEM (Kuxhausen 2015). Other spatial data representing average properties of soil, ephemeral stream networks, sizes of stock pond contributing areas, and area of the alluvial stream channel were assembled from observations, SWRC-DAP, and national databases such as SSURGO (USDA-NRCS) and NHD Plus (McKay et al. 2012). The list in Table 1 is not exhaustive because one could come up with more proxies.

Rainfall event properties: The subwatershed rainfall properties were represented by (1) conditional mean of hourly rainfall (average rainfall for observations greater than zero), (2) the maximum 15 min intensity, (3) conditional mean of rainfall duration (average duration for observations greater than zero), (4) location of the center of the storm with respect to the subwatershed outlet, and (5) the storm size as a fraction of the total watershed area. The time components of the storm event properties were computed for each of the events. We assigned an event time in one of two ways: (1) for a rainfall associated with runoff, the event rainfall assumed the time at which maximum runoff was observed, and (2) in the absence of runoff, the rainfall time was represented as the hour at which the maximum depth of rain occurred. Each of the event rainfall properties (Table 1) was calculated from the continuous hourly data to associate rainfall properties and values leading up to the generation of runoff. Based on the general estimates of travel time in the channels and duration of intense rain, we defined event runoff as the total runoff depth in mm from the 3 h preceding and following the maximum observed runoff rate. The event rainfall depth and duration were calculated as the sum of hourly values from the 3 h preceding and 1 h following the hour of the event time.

Watershed properties: We quantified several physiographic variables describing each of the subwatersheds in the WGEW. These are area, shape, slope, flow length, stream density, stream order, size of stock ponds, channel bed area, saturated hydraulic conductivity, hydrologic soil group, and land cover properties. Refer to Table 1 for details of each of the watershed properties and the source of information. The area of the channel bed included the area of the swales, although this was a minor contribution, based on a geographic information systems (GIS) layer created by Miller et al. (1996).

Antecedent conditions: The antecedent condition represents the condition of the watershed due to the previous rainfall and runoff events and has a significant influence on runoff generation, as well as on the quantity and timing of the runoff. We defined two parameters to incorporate antecedent conditions: antecedent moisture conditions (AMC) and antecedent runoff conditions (ARC). AMC is a watershed scale categorical wetness index indicating wet, average, and dry state based on the amount of rainfall received based on Soil Conservation Service (SCS) threshold in the 5 days prior to a given rainfall event (SCS 1972), which was represented as AMC 5d. However, because the arid conditions in WGEW result in rapid evaporation and drying out of the surface soil, we also classified the wetness condition based on 2 days accumulated rainfall prior to the time in question, that is, AMC 2d, using the same thresholds given in SCS (1972). The ARC differentiates between the contributing area and the channel bed. Based on the amount of flow in the

J. Hydrol. Eng., 2019, 24(10): 04019036

Table 1. Types and list of potential parameters related to rainfall runoff response in Walnut Gulch nested watersheds

Group/class	Variable type	Unit	Definition
Precipitation properties	**Event rainfall depth**	mm/h	Conditional hourly mean rainfall accumulation for 3 h preceding and 1 h following event time of the watershed runoff.
	Maximum rainfall intensity	mm/min	Average of a conditional maximum 15-min intensity of 3 h preceding and 1 h following an event.
	Duration	min	Total duration of rainfall 3 h preceding and 1 h following an event.
	Storm size	m^2/m^2	Areal extent of rainfall event using Theissen polygons normalized by the watershed area.
	Storm distance	m/m	The ratio of flow path from the storm center to the maximum watershed flow path length.
Antecedent condition	5-day antecedent moisture	—	Prior moisture condition in the contributing area based on accumulation of rainfall over 5 days.
	2-day antecedent moisture	—	Dry (<1.27 mm), wet (>5.3 mm), or else average (SCS 1972).
	5-day antecedent runoff	—	Prior moisture condition in the channels based on accumulation of runoff over 5/3 days before the
	3-day antecedent runoff	—	event. Dry (<0.001 mm), wet (> 1 mm), or else average.
Watershed properties	**Area**	km^2	Watershed contributing area.
	Shape	m/m	The ratio of watershed width (in the direction of main channel flow) to length of watershed.
	Slope	%	Average slope in percent.
	Length	Km	The longest flow path based on D8 algorithm.
	Stock pond area	km^2/km^2	Contributing areas of the detention stock ponds in some of the subwatersheds.
	Area of channel bottom	km^2	Miller et al. (1996) measured the channel bottom area of Walnut Gulch channels to estimate transmission loss.
	Stream density	m/m^2	The ratio of total length of NHD high-resolution stream networks to watershed area.
	Stream order ratio	m/m	The ratio of length of first-order stream network to the total length of stream orders 2 and above.
	Hydraulic conductivity	mm/h	Average watershed scale surface layer property for soil water movement from SSURGO database.
	Hydrologic soil group[a]	—	Average soil group showing infiltration ability of the watershed.
	Average land productivity	—	Normal year rangeland production in lbs/acre/yr normalized by average production in Walnut Gulch.
Temporal properties	**Event month**	1–12	Rainfall distribution varies significantly within the summer months.
	Rainfall hours	0–23	Hours representing the rainfall event time showing the diurnal effects.

Note: A total of 22 parameters were identified to be used as predictors for runoff estimation. The significant parameters of the optimal regression equation are indicated in bold.
[a]Hydrologic soil groups A, B, C, and D were assigned numeric values 1 to 4, respectively.

channel, 5 days (ARC 5d) and 3 days (ARC 3d) prior, we classified the ARC into wet "3," average "2," and dry "1." The threshold was defined based on the distribution of runoff depths recorded in WGEW since 1967. Wet conditions were set for prior accumulated runoff with a depth more than 1 mm, and for a dry condition, we chose a prior accumulation of 0.1 mm or less. Wet ARC values describe the condition of relatively wet channel bed sediment, thus limiting transmission losses and encouraging subsequent flow to reach the outlets.

Temporal properties: The time of rainfall occurrence was characterized at three time scales: season of occurrence (summer or nonsummer), month of occurrence, and hour of occurrence. The period of rainfall occurrence helps identify the type of storm received in the area, which could be indicative of its seasonality. Rainfall events in the WGEW are characterized seasonally by different frequency and intensity, and the majority of runoff in the WGEW occurs from summer convective storms associated with the North American Monsoon. These events are episodic and are relatively intense as compared to precipitation outside the summer months (Renard et al. 1993). Observations show that early summer rains (July and August) are typically intense, short-duration, and localized rainfall events. The high-intensity summer rain rates often exceed soil infiltration rates and generate runoff via infiltration excess. Long-duration winter rains, typically originating from frontal storm systems, rarely have sufficiently high intensities to generate runoff (Renard et al. 1993).

Model Selection

Regression modeling requires the identification of the strength of predictors that effectively explain the relationship between the rainfall and runoff. The rainfall event data and associated predictors from WGEW range from very small to large values, allowing us to assess different forms of regression models. We transformed runoff data using exponential functions for better linear relationships with the predictors and developed a regression equation of the form given in the following equation:

$$y_i^k = \sum_{j=0}^{M} \beta_j x_{ji} + \varepsilon_i \tag{1}$$

where y = dependent variable (runoff) at the ith event with a transformation function k; β = regression coefficient of the jth variable; and x = predictor value at the ith observation event. Values of x vary between events (for time-variant properties) and watersheds (for watershed properties); $i = 1 - N$ for the number of observations, $j = 0 - M$ for the number of predictors, and ε_i is an error term of the ith event.

We conducted a multiregression analysis systematically. First, the data were divided into two parts: 85% were used to develop the regression equations, and the remaining 15% were reserved to validate the performance of the models. The validation data included data collected with both the analog (1991–1996) and digital (2011–2016) instruments. Second, runoff was modeled based on combinations of the individual explanatory variables in each of the four groups of variables (Table 1) using p-values and F-test statistics of the multiple regression model based on transformed runoff values. After individual explanatory variables were identified, interaction among variables were evaluated to further improve the performance of the regression models using p-values and F-test statistics. Seventeen interaction terms were evaluated (Table 2.).

Based on combinations of the individual predictors and interaction terms, three regression models were developed: Model I, Model II, and Model III. Model I included the 22 individual predictors, which consist of all predictors in Table 1. The second group (Model II) included individual predictors plus the interactions within precipitation properties, that is, the 11 additional parameters

Table 2. List of known interaction terms among the precipitation and watershed properties

Group/class	Variable type	Interaction within precipitation											Watershed and precipitation interaction					
		1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6
Precipitation properties	Event depth	X	—	X	—	X	—	X	X	X	X	X	—	X	—	X	X	X
	Max. intensity	X	X	—	X	—	—	X	X	—	—	X	X	—	—	—	—	—
	Storm size	—	X	X	—	—	X	X	X	X	X	X	—	—	—	—	—	—
	Duration	—	—	—	X	X	X	—	—	X	X	X	—	—	X	—	—	—
Watershed properties	Hydraulic conductivity	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—
	Channel area	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
	Slope	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—
	Length	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—
	AMC	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—
	Stock pond area	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X

Note: The numbers 1–11 and 1–6 show the different interaction scenarios. Including the individual predictors listed in Table 1, the number of total predictors becomes 39.

in column 1 of Table 2. The third model (Model III) included all the predictors in Tables 1 and 2, increasing the total number of predictors to 39. Each of the models was based on approximately 30,000 summer rainfall and corresponding runoff events measured based on available information from the 12 subwatersheds.

Each of the three regression models was further evaluated to identify a single optimal regression model. This was a two-step process: First, F-tests were conducted to identify the model that best fit the population from which the data were sampled using least squares. All significant parameters and their interaction terms with p-values <0.05 were used as regression predictors. In this step, the objective functions were to minimize the deviation of individual values from the distribution through the sum of squared deviates as a standard error and the residual of the sum of squares (RSS) to maximize the adjusted R-squared. The RSS is the measure of how much the predicted runoff value varies from the observed value for each data point. Adjusted R-squared shows the fraction of variance of the error distribution compared to the variance of the dependent sample variable (runoff volume). We also used the Durbin-Watson test (Durbin and Watson 1950) to assess the normality of the residuals. The F-test regression runs were also checked for normality of the error distribution of the signs. Second, we compared the model predictions of each of the three models to observations in the training data set to evaluate the accuracy of the predictions. These prediction accuracy evaluations of the three models were conducted to shed light on overfitting, limitations, and strengths of the models in QAQC applications that were not apparent in regression analysis. Here, the model with the best predictive skills was selected for further analysis.

Finally, the predictor variables of the model identified as the best of the three models (Model I) were further evaluated using a multimodel inference approach that used the Akaike Information Criteria (AIC; Akaike 1973). Here, the objective was to assess if those predictors could lead to overfitting, which the F-test does not usually reveal. The multimodel inference approach compares the relative quality of models through estimation of information that would be lost if a particular model consisting of the subset of the predictors was used. The multimodel was created from the predictor variables of the best model regardless of their p-values (from the previous F-test), except for two of the antecedent conditions, to avoid redundancies of the same type of parameters in the model. We then applied the automated exhaustive searching algorithm (e.g., Saft et al. 2016). We used only one of each of the AMCs (AMC 2d) and ARCs (ARC 5d) based on the p-values obtained in the previous analysis. The exhaustive searching algorithm is known to be robust (Barto 2017), but also inefficient to generate and fit sets of models through the repeated evaluation of a subset of the model parameters. The objective function was to minimize the AIC and change in AIC (delta) values, which are the measure of the goodness of fit that favors smaller residual error. The analysis then penalizes the inclusion of more predictor variables if the additional regression terms contribute little or no additional information to the regression model (e.g., Marshall et al. 2005).

Model Validation

After the final model was determined by selecting the model with the lowest AIC value, the model was validated using the 15% of the observed events withheld for validation. Here, the objective was to show the performance of the regression model using independent data not included in the regression training. We used the final regression equation to predict runoff, and its performance was evaluated using both categorical (probability of runoff or no runoff prediction) and numerical statistics such as bias, correlation coefficient, and Nash Sutcliffe efficiency coefficient (NSE) applied to each subwatershed and all the combined subwatershed data. These statistical metrics were computed using the untransformed observed and predicted data.

Results and Discussion

Rainfall Characteristics and Watershed Responses

Runoff-generating rainfall events on the WGEW are dominantly summer rains. About 12.5% of the summer rainfall events generated runoff, whereas only 0.5% of the winter rainfall events generated runoff. Comparison of rainfall events at subwatershed scales showed runoff-generating storms in WGEW were characterized by more rainfall depth, higher intensity, larger spatial coverage of the subwatershed, and slightly longer duration than events that did not generate runoff. An average of 8.7% of all measured rainfall events resulted in a runoff at the outlets of the watersheds considered in this study. The mean rainfall depth and maximum intensity for runoff-generating events were higher than the overall mean by about 270% and 140%, respectively (Table 3, rows 6 and 7). The conditional mean of rainfall duration (duration >0 min) and storm size for runoff-generating events were 22% and 23% higher than the overall mean of event duration and storm sizes, respectively (see Table 3, rows 8–10 and 11). Table 3 shows increasing runoff depth per event with decreasing area, consistent with the findings

Table 3. Summary rainfall and runoff event characteristics for selected subwatersheds in Walnut Gulch

Number	Description	Subwatershed							
		63.006	63.015	63.010	63.003	63.011	63.004	63.104	63.112
1	Watershed area (km sq)	93.34	23.58	15.57	9.362	7.85	2.26	0.048	0.019
2	Ratio of runoff to rainfall events (%)	8.18	6.39	6.62	7.48	8.47	6.69	13.9	6.98
	Summer events (%)	11.99	9.67	9.91	11.39	13	9.56	19.26	11.48
	Nonsummer events (%)	0.42	0.19	0.235	0.13	0.32	0.3	1.35	0.26
3	Average runoff ($Q_f > 0$)	0.414	0.688	0.48	0.46	0.91	1.16	3.23	3.1
4	Event mean rainfall ($d > 0$)	4.75	3.94	3.74	4.05	3.95	4.22	4.78	4.6
5	Event mean rainfall ($Q_f > 0$)	13.3	15.46	13.26	14.58	14.82	17.61	14.44	17.6
6	Mean max. rainfall intensity ($d > 0$)	46.8	32.7	36.78	71.4	30.26	62	21.27	16.97
7	Mean max. rainfall intensity ($Q_f > 0$)	164.5	142.1	148.5	135.8	130.9	121.4	77.7	78.9
8	Mean rainfall duration ($d > 0$)	112.5	113.2	114.4	114.6	113.4	115.3	101.13	102.1
9	Mean rainfall duration ($Q_f > 0$)	153.2	153.5	157.2	145.1	138.1	142	124.7	123
10	Average storm size ($d > 0$)	0.426	0.797	0.71	0.562	0.83	0.744	0.995	1
11	Average storm size ($Q_f > 0$)	0.807	0.936	0.83	0.911	0.96	0.968	1	1
12	Stock pond detention area (%)	—	27.91	2.93	42.1	4.5	0	0	0
13	Number of rain gauges (gauge)	56	18	14	15	9	5	2	1

Note: Duration in minutes, storm size as a fraction of the watershed area, runoff in mm, rainfall depth in mm. The analysis included in the table was the conditional average of the measured rainfall ($d > 0$) and runoff ($Q_f > 0$).

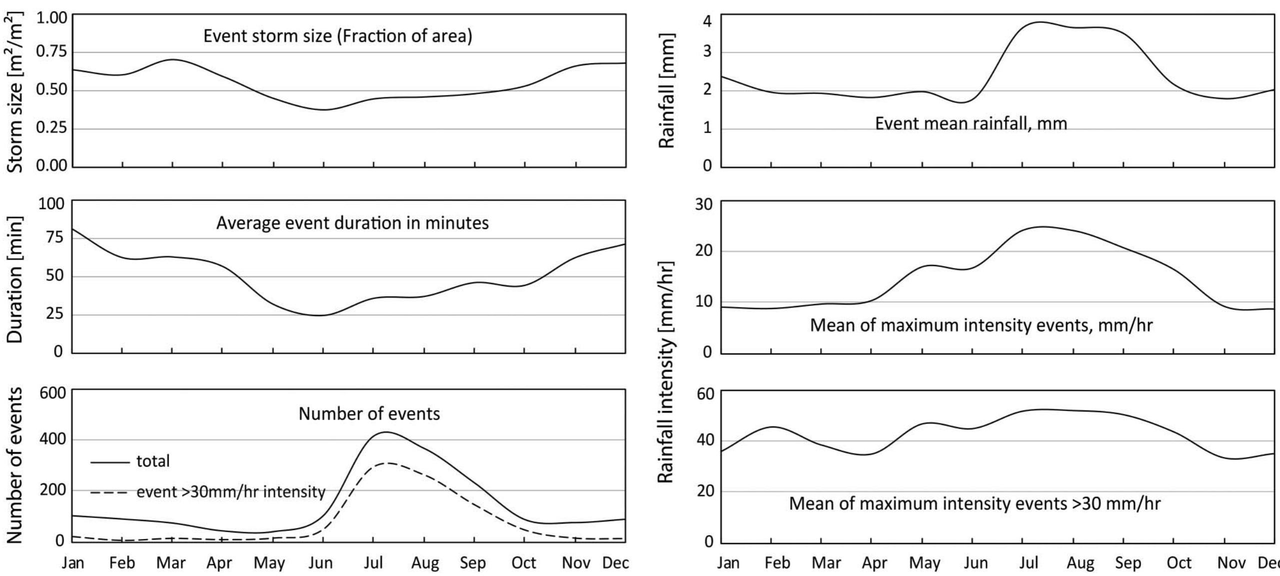

Fig. 2. Long-term monthly average rainfall properties such as storm sizes, event durations, number of events, mean event rainfall depth, and maximum intensity summarized on a monthly scale for the entire Walnut Gulch watershed (149 km^2).

of Goodrich et al. (1997). Average runoff depths per event range from a minimum of 0.4 mm in watershed 63.006 (94 km^2) to a maximum of 3.23 mm in watershed 63.104 (0.048 km^2).

The monthly summary of precipitation over the WGEW (Fig. 2) reveals the typical characteristics of precipitation over the semiarid southern Arizona environment. Summer rain in the WGEW is typically more frequent, shorter duration, highly localized, and more intense (Fig. 2) than that during the winter period. Fig. 3 shows the distribution of historical antecedent conditions and the diurnal and seasonal rainfall properties in the WGEW. In the summer, the likelihood of dry antecedent conditions in both the watershed uplands and the channel beds is significantly higher [Fig. 3(a)], which is typical of the arid environment of southwestern US watersheds. The AMC 5d showed about 21% average and 24% wet conditions during the summer season [Fig. 3(a)]. The distribution of the 2-day AMC (AMC 2d) closely resembled the distribution of both ARC 5d and ARC 3d that describe the wetness conditions of channel bed sediment [Fig. 3(a)], which was consistent with the smaller number of runoff-generating events. This analysis shows a high probability of dry AMC and dry ARC [Fig. 3(a)], indicating dry conditions before most of the events, resulting in large abstractions of runoff water in the channels and contributing areas resulting in low runoff-to-rainfall ratios (e.g., Goodrich et al. 2004). Seasonal and diurnal precipitation patterns were important factors in the quantity and frequency of runoff in Walnut Gulch. The average diurnal cycle [Fig. 3(b)] of the WGEW shows a large peak around 21:00 MST, with a significant portion of the rain occurring after midday. Events with longer duration and larger spatial coverage with small rainfall intensity are typical, characteristic of winter storms. The winter storms are usually less frequent, and did not generally produce runoff [Figs. 2, 3(c and d)].

Fig. 4 shows runoff responses of the interaction of two precipitation properties, rainfall depth versus intensity or rainfall duration or storm size. Most runoff was generated from high-intensity

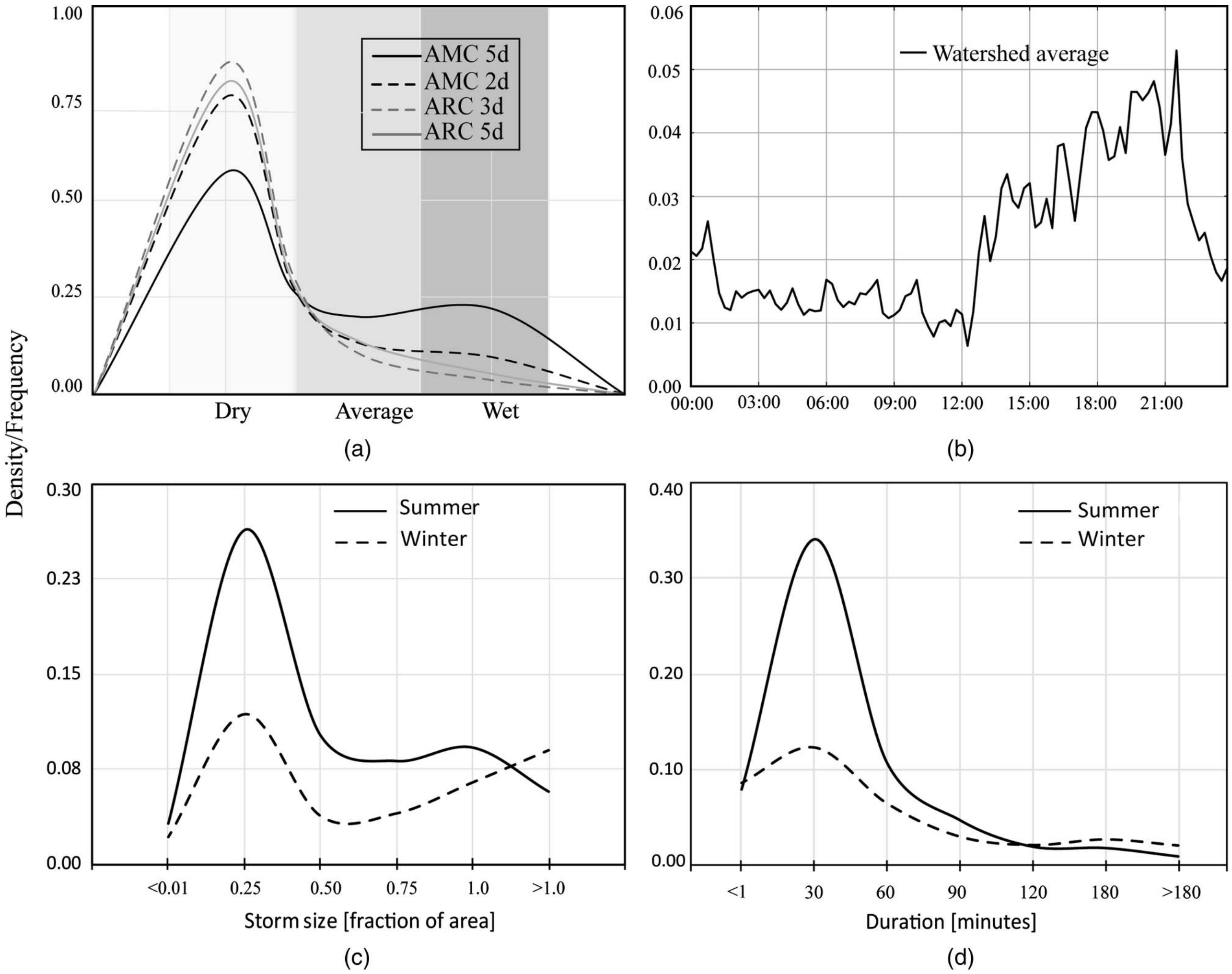

Fig. 3. Summary of rainfall properties and their responses based on rainfall and runoff events recorded across Walnut Gulch: (a) summer period antecedent conditions over the watershed based on prior 5-day (AMC 5d) or 2-day (AMC 2d) accumulated rainfall, and channel floor (ARC 5d and ARC 2d) based on the magnitude of prior accumulated runoff at the outlet; (b) average diurnal cycle of rainfall observations of 83 rain gauges in Walnut Gulch; (c) seasonal distribution of storm size as a fraction of summation of areas of Theissen polygons of 83 rain gauges to total watershed area; and (d) seasonal distribution of storm duration of events.

rainfall. Intensities less than 30 $\mathrm{mm\,h^{-1}}$, regardless of the rainfall depth, resulted in little chance of runoff [Fig. 4(a)], consistent with the findings of Syed et al. (2003). Small, localized rainfall events (covering less than 40% of the watershed) and very short-duration events (<25 min) generated little runoff [Figs. 4(b and c)]. However, as the duration and storm size with respect to watershed size increased, runoff generation increased. The typical rainfall of the southwest, which is a localized, high-intensity rainfall with very short duration, usually resulted in smaller runoff depth per event at the watershed outlet in larger watersheds (Table 3). This is due to more runoff leaving the system through channel transmission losses as the ephemeral channel area increases with increasing watershed size (Miller et al. 1996). In addition, the percentage of watershed area covered by high-intensity rainfall generally decreases as watershed drainage area increases (Goodrich et al. 2004).

Predictive Model

The application in data quality assessment of the WGEW showed Model I as the best model. Based on the relation between predicted and observed runoff, Model I correctly predicted runoff and no-runoff events more than 85% of the time. Models II and III erroneously predicted runoff for a large number of events when the observed runoff was zero. Both Models II and III showed signs of overfitting based on evaluation of the predicted runoff against the observations. Model II incorrectly predicted runoff more than 65% of the time when the rainfall did not generate runoff; that is, a large number of false positives occurred, disqualifying the use of Models II and III for improving data quality in the WGEW. The scatter plots of Model II (not shown) and Model III versus the observed data [Fig. 5(d)] showed better agreement for large runoff events than the same comparison for Model I [Fig. 5(c)]. However, Models II and III showed a large overestimation for low-runoff events compared to Model I. Although the coefficient of determination and standard errors were better for Models II and III, with adjusted R^2 of 0.68 and 0.7, respectively, Model I showed superior performance for data quality assessment across all ranges of runoff values based on the probability of runoff detection and false alarms. It is also important to note that the comparison of the predicted and observed runoff in Model I [Fig. 5(c)] showed significant spread, indicating inaccuracy in predicted runoff magnitudes, which is also inherent to semiarid modeling difficulties.

J. Hydrol. Eng., 2019, 24(10): 04019036

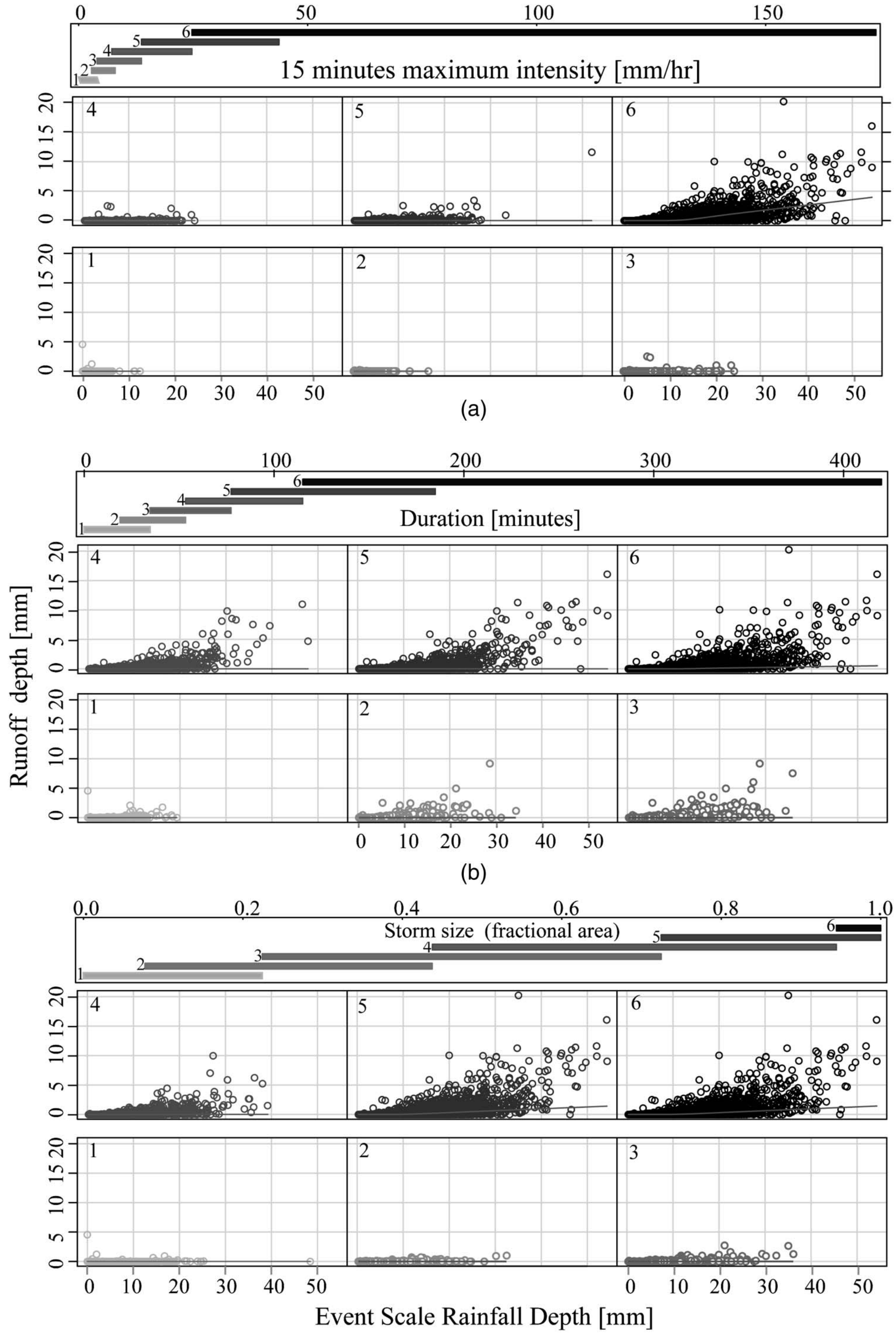

Fig. 4. Description of combined effects of rainfall properties on runoff: (a) 15-min maximum intensity versus rainfall depth relation to runoff; (b) duration-rainfall depth-runoff relation; and (c) storm size-rainfall depth-runoff. Bar graphs in (a–c) show maximum 15-min intensity, duration, and storm sizes subdivided into six categories, where the subdivisions are labeled 1–6 and gray to black color. Corresponding scatter plots show relation between event rainfall depth to runoff as a watershed response for each of the categories in the bar chart from lower left (light gray starting with label number 1) to top right (black with label number 6).

The multimodel inference approach resulted in an optimal model that includes the 18 model parameters listed in bold in Table 1 as the predictors. AIC analysis was applied to all Model I predictors except AMC 5d and ARC 3d. We removed AMC 5d and ARC 3d based on the results of the F-tests to avoid the redundancy of antecedent conditions, and used AMC 2d and ARC 5d as the regression parameters for AIC analysis. The F-test showed, out of four antecedent conditions, that AMC 5d was the only nonsignificant (with p-value >0.05) parameter in Model I. Both antecedent runoff conditions were found to be significant parameters in

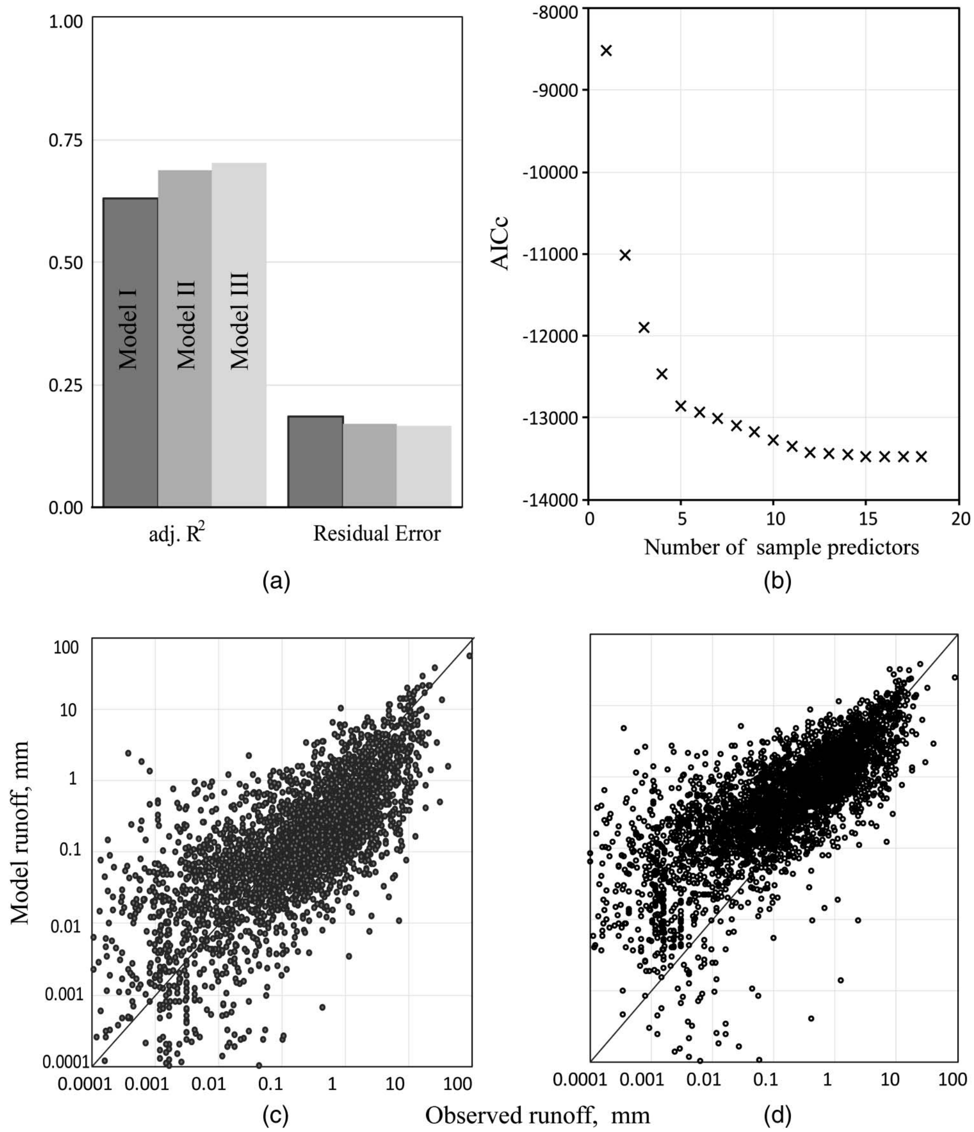

Fig. 5. Comparison of the regression model performances based on (a) adjusted R^2 and residual errors (mm) from series of F-tests of models I, II, and III; (b) corrected AIC values of models through the automated exhaustive search approach; (c) comparison of Model I predicted runoff versus observed runoff; and (d) comparison of Model III predicted runoff versus observed runoff. Watershed runoff depth was computed as the total observed flow volume divided by the watershed area. Computed watershed depths will be quite small in association with small runoff events.

Model I, but included ARC 5d ($p < 0.05$), reducing the number of parameters to 20. The combination of 18 of the 20 parameters showed a minimum calculated AIC value of −13,478 and a delta AIC value of 0. The second best-performing model showed a delta AIC value greater than 4. Fig. 5(b) shows the AIC values for the various combinations of predictors. The 18 parameters identified included all precipitation, watershed, and time properties except the area of channel bottom and hour of the occurrence of the rainfall events. The coefficients of the regression equation that relates runoff to rainfall and other watershed parameters are given in Table 4.

In general, the optimum model overestimated smaller events and underestimated larger events [Figs. 5(c), 6(a and b)]. The Durbin-Watson statistical test (1.84) of the residuals shows a distribution around zero on the y-axis [Fig. 6(a)]. The large residual [Fig. 6(b)] for events around 22,000 shows a large underestimation in watershed 104, which is a relatively flashier watershed, and the significant overestimation for events was around 23,500 for 112. The deviation of the quantile plot (Fig. 6) at both ends from the theoretical fit shows the model performs well over the range of

Table 4. Regression equation coefficients for runoff prediction ($k = 0.25$)

Variable	Coefficient
Event rainfall depth	0.035
Maximum rainfall intensity	0.001356
Duration	−0.000377
Storm size	−0.051
Storm distance	−0.039
2-day antecedent moisture	0.005331
5-day antecedent runoff	0.051
Event month	0.012
Area	0.011
Shape	−0.000043
Slope	−0.058
Length	−0.012
Stock pond area	−0.008015
Area of channel bottom	−0.003208
Stream density	0.218
Stream order ratio	−0.281
Hydraulic conductivity	0.053
Average land productivity	0.165

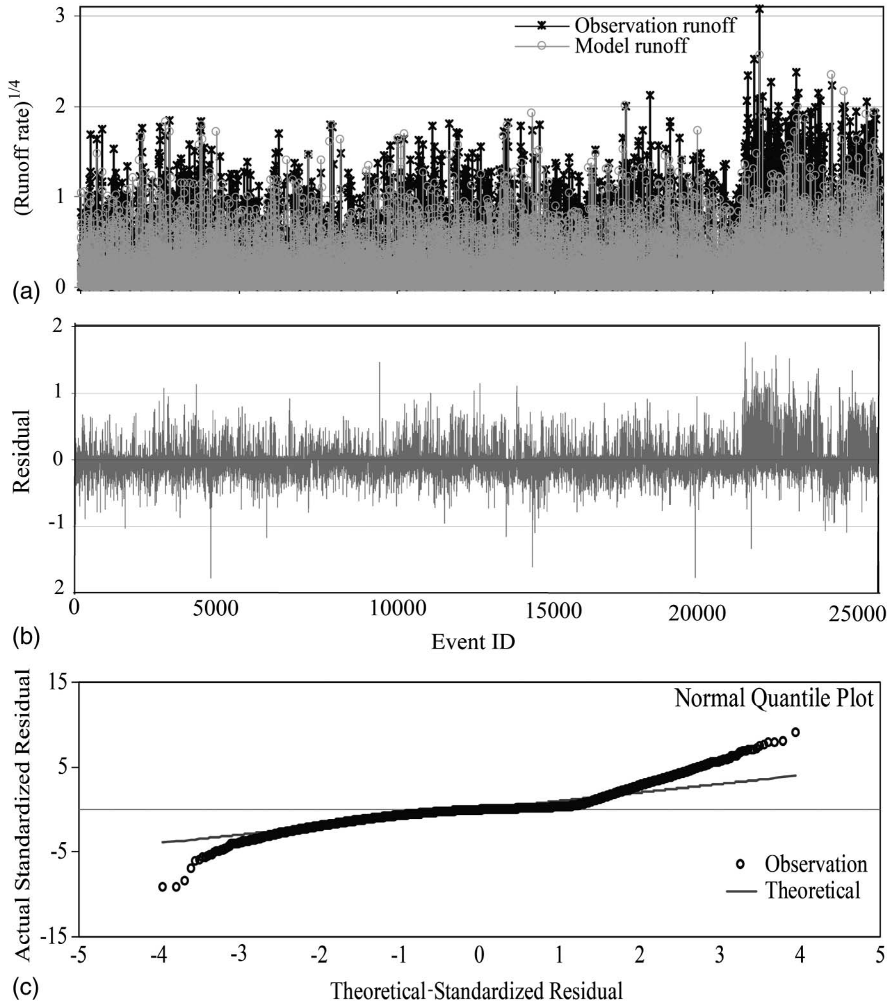

Fig. 6. (a) Comparison of the optimal model-predicted runoff and transformed observed runoff time series; (b) the residual from the 12 subwatersheds; and (c) the normal quantile plot of the optimal multiple regression model. The events are organized based on the watershed IDs from 03 to 121. The large and medium-sized watersheds that are designated with small ID numbers such as 03, 04, 06, and so on are on the right side, and watersheds with the smallest area (104, 112, and 121) are on the very right of (a and b).

runoff values except for very small and large runoff values. Regressions showed the highest correlation when event data from each of the subwatersheds were regressed separately, where only precipitation properties and AMC are relevant predictors as all watershed properties remain constant. Small, flashy watersheds such as 63.104 and 63.121 showed the highest correlation, explaining more than 75% of the variances (not shown). With the increase in watershed sizes, the partial area response became more dominant (Stone and Paige 2003), reducing the runoff rainfall ratio at the outlets (Goodrich et al. 1997).

Model Validation and Evaluation

We considered all predicted runoff using the coefficients in Table 4 with values less than 0.0001 mm as no runoff. The performance of Model 1 using the validation data subset is shown in Table 5 and Fig. 7. The models showed correlation coefficients ranging from 0.48 in watershed 63.010 up to 0.9 in watershed 63.003 at the individual subwatershed scale. The NSE was as high as 0.76 for watershed 63.121 and 0.66 for watershed 63.011. Watershed 63.010 showed the lowest NSE (−10.36), showing that the average

Table 5. Comparison of the predictive skill of the validation data at subwatershed scale and all watersheds combined

	Subwatersheds								
Statistics	WG	63.003	63.006	63.004	63.121	63.112	63.104	63.010	63.011
Correlation coefficient	0.65	0.48	0.81	0.859	0.94	0.723	0.53	0.485	0.812
Bias	0.055	−0.012	−0.007	−0.025	−0.084	−0.044	−0.68	0.019	−0.052
NSE	0.425	0.135	0.61	0.74	0.76	0.47	0.162	−10.36	0.66
P (flow hits)	0.9	0.897	0.889	1	0.98	0.96	0.93	0.92	0.81
P (no flow hits)	0.86	0.879	0.895	0.87	0.87	0.76	0.82	0.87	0.87
Number of observed flow	486	35	56	31	38	32	48	77	61
Amount of validation data	4,801	459	609	239	296	469	468	796	459

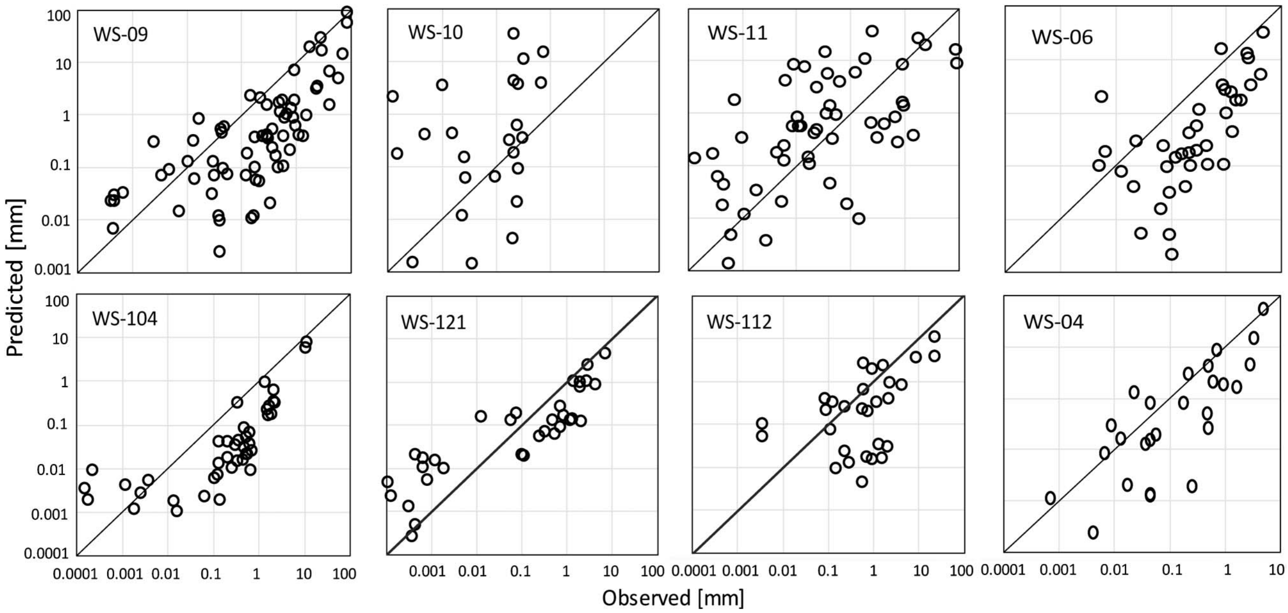

Fig. 7. Comparison of observed and modeled runoff for the 15% validation data set to evaluate the predictive skill of the model. The performance in watershed 63.010 (WS-10) is the poorest.

of the observed runoff is a better predictor than the model due to the large negative value. The regression equation consistently underestimated the runoff in all watersheds with a bias less than 0.03. The small watersheds (63.104, 63.112, and 63.121) showed large underestimation where the calculated bias was nearly double that of the medium sized watersheds. When all subwatershed data were combined, the model performance showed a modest accuracy, with a correlation coefficient of 0.46 and NSE of 0.425. Regardless of the accuracy of runoff values, the regression model predicted the occurrence of runoff 93% of the time (see probability of flow hits in Table 5) and no-runoff event occurrences 86% of the time. The model performed poorly in watershed 63.010, which has an elongated shape compared to the rest of the subwatersheds. Shape was not a significant predictor in the general regression model. However, the distinctively elongated watershed has a very long main channel and a relatively large channel bottom area with alluvial deposits that may increase the potential for channel transmission losses, resulting in the difference in the rainfall-runoff relation of watershed 63.010.

Application of the Regression Model for Assessing Measurement Quality

In this exercise, we illustrate how the regression model was applied to identify potential errors previously unidentified in the database. We applied the regression model to predict runoff for watersheds 63.004 (small to medium size) and 63.006 (relatively large) to demonstrate the quality assurance potential of the regression model. Using the predicted and observed runoff with the depth of precipitation events, we identified events with questionable rainfall-runoff relations based on the following three situations.

The first condition was the identification of substantial rainfall events (>20 mm depth) that did not produce runoff. Watershed 63.004 produced runoff in 93% of the rainfall events with precipitation depth greater than 20 mm and intensity higher than 30 mm/h.

Table 6. Examples of flagged problematic rainfall and runoff events based on the inconsistencies observed in the application of the regression model for watersheds 63.004 and 63.006

Watershed	Condition	Event date	Runoff (mm)	Predicted (mm)	Residual (mm)	Rainfall (mm)
63.004	1	6/29/1996 19:00	0	0.987	−0.987	31.657
		7/29/1987 14:00	0	0.637	−0.637	27.527
		7/14/1985 18:00	0	0.405	−0.405	23.778
		10/19/1972 14:00	0	0.183	−0.183	21.242
		8/6/1966 0:00	0	0.241	−0.241	22.543
	2	9/5/1977 15:00	1.233	0	1.233	1.973
		7/12/1981 23:00	1.096	0	1.095	3.494
63.006	1	7/24/1969 13:00	0	0.250	0.250	21.758
		3/19/1973 13:00	0	0.168	0.168	26.643
		7/31/1982 15:00	0	0.210	0.210	22.510
		8/17/1986 23:00	0	0.222	0.222	24.779
		10/14/1988 19:00	0	0.266	0.266	24.760
		7/9/1993 2:00	0	0.619	0.619	27.849
	2	7/31/1981 8:00	0.612976	0	−0.613	6.910

Note: Condition 1 shows a list of events with substantial rainfall events for which regression predicted runoff but there was no associated runoff observation; and Condition 2 lists events with substantial runoff, but no runoff model prediction.

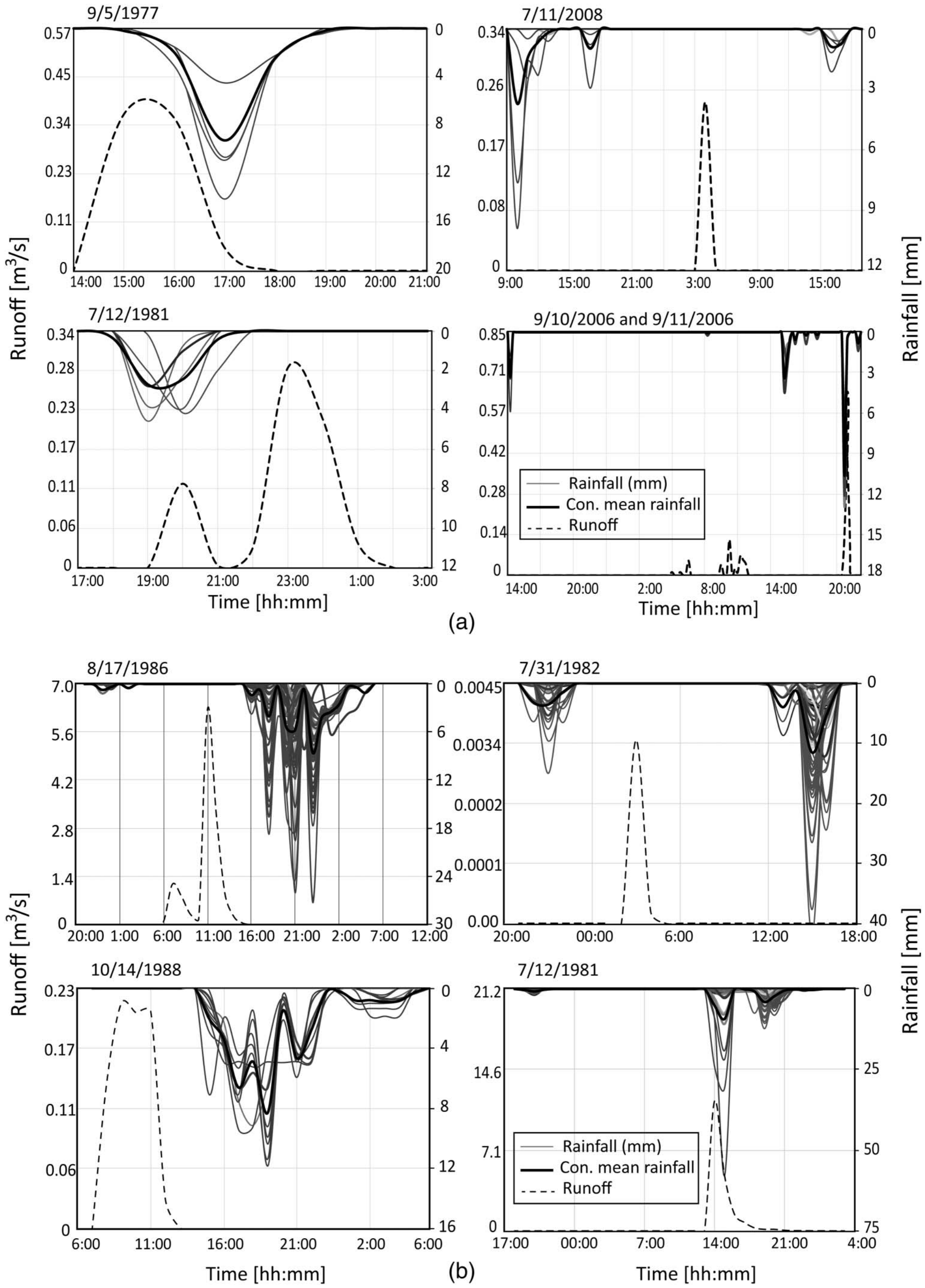

Fig. 8. Selected runoff and rainfall from multiple gauges within the watershed of interest on inverted y-axis observations from the lists in Table 6 of (a) watershed 63.004; and (b) watershed 63.006 that explains the reason those events were picked up by the regression.

Watershed 63.006 produced runoff in 85% of rainfall events under similar conditions. Based on the 20-mm threshold, we identified five rainfall events in watershed 63.004 and six rainfall events in watershed 63.006 that did not generate runoff when the regression model predicted significant runoff (Table 6).

The second condition identified events with substantial observed runoff (>0.5 mm depth), but zero predicted runoff. Under this condition, we identified two runoff events for watershed 63.004 on 9/5/1977 and 7/12/1981 and one runoff event for watershed 63.006 on 7/31/1981 (Table 6).

The third condition identified those events with observed runoff (>0 mm depth) but zero rainfall accumulation and predicted flow. Under this condition, we identified 11 events in watershed 63.004 and 5 events in watershed 63.006 (not shown in Table 6). Those data points with observed runoff but no precipitation associated with the runoff within the defined time window could be related to timestamp error, instrument malfunction, or human error during instrument maintenance and calibration. Fixing those error types requires further investigation and combing through the available site notebooks or the analog charts for indications of measurement error.

In Table 6, we show only about 56% of the identified data points for which we found a good reason to believe that there were errors either in rainfall or runoff recordings. With regard to those data points identified as problematic that we did not include in Table 6, even though there are some possible reasons related to the rainfall data such as low-intensity rain with long duration, which favors

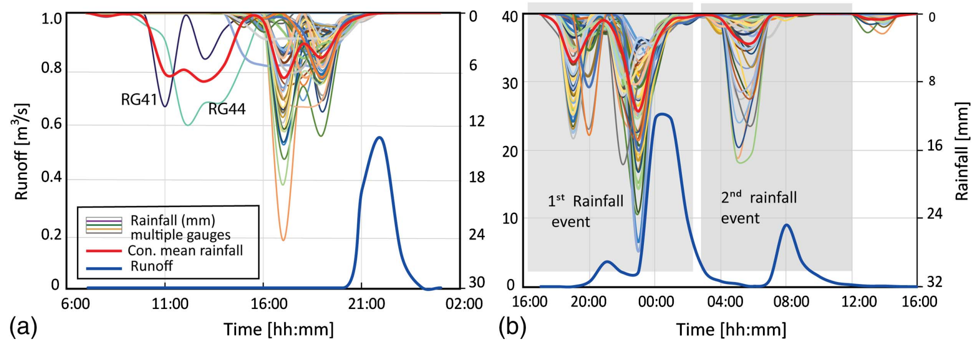

Fig. 9. Two events in watershed 63.006 identified as problematic due to (a) possible recording error in rain gauges 41 and 43 on 3/19/1973 that extended the rainfall event duration outside the bounds of the model assumptions; and (b) a false alarm on 7/31/1981 identified due to a small second rainfall event that fell on a very wet (saturated) surface.

more infiltration than runoff generation, or very wet antecedent conditions, which could result in runoff generation even for small rainfall depth, and so on, we could not find any reason to believe that there were problems with those data points.

Closer examination of the database as well as inspection of analog strip charts for the identified questionable data points in Table 6 revealed different kinds of problems related to the timing of the data recording and storage. Nine of the events in Table 6 (five in watershed 63.004 and four in watershed 63.006) are shown in Fig. 8 in terms of the timing of the recordings of rainfall and runoff. In some of the events, such as the events on 9/5/1977 in 63.004 and events 8/17/1986, 10/14/1984, and 7/12/1981 in 63.006, the source data show that most of the rainfall with the potential to cause runoff in the watershed appeared to have been recorded while the runoff was receding. In most of the events shown in Fig. 8, events with substantial runoff observation and zero model predictions are possibly related to either an incorrect timestamp of the precipitation or the runoff, or an error in the digitization process of the analog chart. In watershed 63.004 on 7/12/1981, the structure of the rainfall and its timing do not seem to explain the dynamics recorded in the runoff.

The precipitation event on 3/19/1973 in Table 6 showed a conditional mean of 26.6 mm rain that does not have associated runoff. In the source database, a very long-duration rainfall that extended over 15 h [see the watershed conditional average rain depth (red line) in Fig. 9(a)] was split into multiple events when a window of the 3 h preceding and 3 h following maximum runoff was applied to summarize the hourly data into event scale. A detailed look into all the precipitation recorded in the watershed from multiple rain gauges revealed that the problem in this data point could be a wrong timestamp on some of the gauge recordings. At any given location, the precipitation duration was not as long as it appeared on the watershed scale. Two rain gauges, 63.041 and 63.044 [Fig. 9(a)], recorded 8 h earlier and stopped recording right before the neighboring gauges started recording. A cross-check on the rainfall and runoff analog charts on 3/19/1973 also confirmed that gauges 63.041 and 63.044 recorded the rainfall at the same time with the neighboring gauges and the shift in the digitized data within the database could be related to data entry error made during reading and digitizing the charts for the two rain gauges.

On 7/31/1981 [Fig. 9(b)], two rainfall events occurred across the watershed but were summarized as a single event of 16 h. The first event was large enough that the regression model predicted runoff for it, but the second event was so small that the regression did not predict runoff. Considering a very wet surface because of rain a few hours earlier, it is very likely that that the second rain also generated runoff that the regression model was not able to predict.

Application of the regression model to identify possible data errors in the WGEW showed that the method should be applied with care because of accuracy limitations in predicting the magnitude of runoff events. Without a doubt, this type of improvement in the data would directly affect the quality of our resulting hydrologic analysis; data interpretation; and findings such as forecasts, warnings, and decision support guidance. Most applications of data are the responsibility of the user, and their accuracy and consistency are usually ignored. However, it is a known fact that errors are common in long-term databases like the WGEW. There are several reasons, including human mistakes and instrument malfunctions, for errors. The most common approach before any data usage would be to check if there are recordings that were off by orders of magnitude or outside an acceptable range of values, big jumps in value for no reason, and so on. Tools such as this regression-based model that are able to identify, flag, and correct errors that are not as obvious as outliers have paramount importance in improving the quality of historical hydroclimatic observations.

Conclusions

A regression approach used in this study to develop a method for flagging erroneous data based on the concept of a causal relationship among watershed properties, antecedent conditions, rainfall, and runoff was applied in the WGEW. The development of the multiparameter regression model required careful evaluation of its application, in this case its use for QAQC of historical data. In addition to the combination of F-tests and the exhaustive search approach, it needed to select an optimal model based on the accuracy of the predicted runoff. The precipitation properties explained more of the variance in the rainfall-runoff relation than the watershed and antecedent properties. AMC 3d explained the rainfall runoff relations better than AMC 5d, the standard prior 5 days' accumulation (SCS 1972), which could be related to the semiarid environment in the Southwest that facilitates rapid drying of soils.

The regression model was used to predict runoff given measured rainfall characteristics. It showed significant improvement in runoff prediction with the inclusion of known interaction terms, indicating the model with more variables had better performance. However, the model with the least number of predictors, consisting of only individual precipitation and watershed properties (18 predictors), showed superior performance for quality assurance assessments because it showed the smallest false alarm runoff. The optimal model predicted runoff with a fairly modest accuracy (adjusted R^2 of 0.63 and standard error equal to 0.218) in terms of magnitude estimation, with a general underestimation of the observed runoff. However, the model showed good skill in detecting runoff-generating rainfall events, with the lowest false alarm runoff.

The use of predicted runoff in combination with the observed rainfall and runoff illustrated good potential for implementing additional quality assurance procedures in the DAP database. Using this regression model, we identified a total of 25 data points that had some type of error. Most of the problems identified in this exercise were recorded during the analog period, that is, prior to the year 2000, and were usually due to human errors in assigning events at the wrong time when reading the analog charts and during the digitization process. It is also important to note that flagging events with questionable rainfall or runoff points requires caution. Once the questionable data points are identified, it requires validation of the errors using all available information, such as field notebooks, a copy of the analog charts, and observations in the neighboring gauges. This exercise demonstrated that the regression model was able to identify additional inconsistent observations beyond those identified by QAQC procedures on precipitation or runoff independently. It thus provides a useful addition to complement existing QAQC procedures to ensure that rainfall and runoff data in the Walnut Gulch database are consistent and contain minimal errors. The model also has the potential for making runoff predictions in similar hydroclimatic environments where high-resolution ground-based radar-rainfall estimates are available.

In addition to the regression approach, we are currently working on the development of other potential tools for flagging inconsistent and erroneous data points based on the causality between rainfall and runoff. The causal relations include (1) rainfall is spatially correlated in the watershed, which can be applied to the recorded rainfall fields only; (2) there is a temporal link between rainfall and runoff (showing the presence of reasonable lag time between rainfall and runoff); and (3) there is an upper limit to the amount of runoff generated from a rainfall event (a threshold of runoff coefficient). Unlike the regression method, the application of the methods that apply to the latter two causal relations is only limited to events that resulted in runoff.

Acknowledgments

Funding and support were provided by the US Department of Agriculture, ARS. Thanks to the many dedicated USDA-ARS Southwest Watershed Research Center staff in Tombstone and Tucson, Arizona, who made possible the collection of high-quality rainfall and runoff records in the WGEW. The vision and commitment of all ARS scientists and administrators to construct, manage, and operate the experimental networks in Walnut Gulch for the long term are to be commended.

References

Akaike, H. 1973. “Information theory and an extension of the maximum likelihood principle.” In *Proc., 2nd lnt. Symp. on Information Theory*, edited by B. N. Petrov, and F. Csaki, 267–281. Budapest, Hungary: Akademia Kiado.

Amitai, E., C. L. Unkrich, D. C. Goodrich, E. Habib, and B. Thill. 2012. “Assessing satellite-based rainfall estimates in semiarid watersheds using the USDA-ARS Walnut Gulch gauge network and TRMM PR.” *J. Hydrometeorol.* 13 (5): 1579–1588. https://doi.org/10.1175/JHM-D-12-016.1.

ASCE. 2000a. “Task committee on application of artificial neural networks in hydrology. I: Preliminary concepts.” *J. Hydrol. Eng.* 5 (2): 115–123. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(115).

ASCE. 2000b. “Task committee on application of artificial neural networks in hydrology. II: Hydrologic applications.” *J. Hydrol. Eng.* 5 (2): 124–137. https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124).

Barto, K. 2017. “MuMIn v1.7.” Accessed April 1, 2017. https://www.rdocumentation.org/packages/MuMIn.

Brakensiek, D. L., H. B. Osborn, and W. J. Rawls. 1979. *Field manual for research in agricultural hydrology*. USDA Agriculture Handbook 224. Washington, DC: USDA.

Brantley, S. L., et al. 2017. “Designing a network of critical zone observatories to explore the living skin of the terrestrial Earth.” *Earth Surf. Dyn.* 5 (4): 841–860. https://doi.org/10.5194/esurf-5-841-2017.

Bridges, W. C. 1982. *Technique for estimating magnitude and frequency of floods on natural-flow streams in Florida*. USGS Water-Resources Investigations 82-4012. Washington, DC: USGS.

Chow, V. T. 1964. *Handbook of applied hydrology*. New York: McGraw-Hill.

Cochran, B. J., H. E. Hodges, R. K. Livingston, and R. D. Jarrett. 1979. *Rainfall-runoff data from small watersheds in Colorado*. USGS Open File Rep. No. 79-1261. Washington, DC: USGS.

Costa, A. C., A. Bronstert, and J. C. de Araujo. 2012. “A channel transmission losses model for different dryland rivers.” *Hydrol. Earth Syst. Sci.* 16 (4): 1111–1135. https://doi.org/10.5194/hess-16-1111-2012.

Das, N. N., B. P. Mohanty, M. H. Cosh, and T. J. Jackson. 2008. “Modeling and assimilation of root zone soil moisture using remote sensing observations in Walnut Gulch Watershed during SMEX04.” *Remote Sens. Environ.* 112 (2): 415–429. https://doi.org/10.1016/j.rse.2006.10.027.

Duan, Y., P. Heilman, and D. P. Guertin. 2006. “A web-based tool for economic analysis of sediment control on rangeland watersheds.” In *Proc., IEMSs 3rd Biennial Meeting: “Summit on Environmental Modelling and Software”*. Burlington, VT: International Environmental Modelling and Software Society.

Durbin, J., and G. S. Watson. 1950. “Testing for serial correlation in least squares regression: I.” *Biometrika* 37 (3–4): 409–428. https://doi.org/10.1093/biomet/37.3-4.409.

Fotovatikhah, F., M. Herrera, S. Shamshirband, K. Chau, S. F. Ardabili, and M. J. Piran. 2018. “Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work.” *Eng. Appl. Comput. Fluid Mech.* 12 (1): 411–437. https://doi.org/10.1080/19942060.2018.1448896.

Goodrich, D. C., I. S. Burns, C. L. Unkrich, D. Semmens, D. P. Guertin, M. Hernandez, S. Yatheendradas, J. R. Kennedy, and L. Levick. 2012. “KINEROS2/AGWA: Model use, calibration, and validation.” *Trans. ASABE* 55 (4): 1561–1574. https://doi.org/10.13031/2013.42264.

Goodrich, D. C., T. O. Keefer, C. L. Unkrich, M. H. Nichols, H. B. Osborn, J. J. Stone, and J. R. Smith. 2008a. “Long-term precipitation database, Walnut Gulch Experimental Watershed, Arizona, United States.” *Water Resour. Res.* 44 (5): W05S04. https://doi.org/10.1029/2006WR005782.

Goodrich, D. C., L. J. Lane, R. M. Shillito, S. N. Miller, K. H. Syed, and D. A. Woolhiser. 1997. “Linearity of basin response as a function of scale in a semiarid watershed.” *Water Resour. Res.* 33 (12): 2951–2965. https://doi.org/10.1029/97WR01422.

Goodrich, D. C., C. L. Unkrich, T. O. Keefer, M. H. Nichols, J. J. Stone, L. Levick, and R. L. Scott. 2008b. “Event to multidecadal persistence in rainfall and runoff in southeast Arizona.” *Water Resour. Res.* 44 (5): W05S14. https://doi.org/10.1029/2007WR006222.

Goodrich, D.C., D. G. Williams, C. L. Unkrich, J. F. Hogan, R. L. Scott, K. R. Hultine, D. Pool, A. L. Coes, and S. N. Miller. 2004. “Comparison of methods to estimate ephemeral channel recharge, Walnut Gulch, San Pedro River Basin, Arizona.” In *Recharge and Vadose Zone processes: Alluvial basins of the Southwestern United States*. Water Science and Application No. 9, edited by F. M. Phillips, J. F. Hogan, and B. Scanlon, 77–99. Washington, DC: American Geophysical Union.

Haupt, S. E., and B. Kosovic, 2015. “Big data and machine learning for applied weather forecasts: Forecasting solar power for utility operations.” In *Proc., IEEE Symp. Series on Computational Intelligence*, 496–501. New York: IEEE.

Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich. 1998. “Integration of soil moisture remote sensing and hydrologicmodeling using data assimilation.” *Water Resour. Res.* 34 (12): 3405–3420. https://doi.org/10.1029/1998WR900001.

Keefer, T. O., C. L. Unkrich, J. R. Smith, D. C. Goodrich, M. S. Moran, and J. R. Simanton. 2008. "An event-based comparison of two types of automated-recording, weighing bucket rain gauges." *Water Resour. Res.* 44 (5): W05S12. https://doi.org/10.1029/2006WR005841.

King, D. M., S. M. Skirvin, C. D. Holifield Collins, M. S. Moran, S. H. Biedenbender, M. R. Kidwell, M. A. Weltz, and A. Diaz-Gutierrez. 2008. "Assessing vegetation change temporally and spatially in southeastern Arizona." *Water Resour. Res.* 44 (5): W05S15. https://doi.org/10.1029/2006WR005850.

Knipper, K. R., T. S. Hogue, K. J. Franz, and R. L. Scott. 2017. "Downscaling SMAP and SMOS soil moisture with moderate-resolution imaging spectroradiometer visible and infrared products over southern Arizona." *J. Appl. Remote Sens.* 11 (2): 026021. https://doi.org/10.1117/1.JRS.11.026021.

Kolassa, J., et al. 2018. "Estimating surface soil moisture from SMAP observations using a neural network technique." *Remote Sens. Environ.* 204 (Jan): 43–59. https://doi.org/10.1016/j.rse.2017.10.045.

Kuxhausen, D. 2015. *Ground control survey report, United States geological survey FY15 ARS-USDA AZ Walnut Gulch QL1 LIDAR*. Woolpert Project No. 75861. Washington, DC: USGS.

Lane, L. J., M. H. Diskin, and K. G. Renard. 1971. "Input-output relationships for an ephemeral stream channel system." *J. Hydrol.* 13: 22–40. https://doi.org/10.1016/0022-1694(71)90198-3.

Marshall, L., D. Nott, and A. Sharma. 2005. "Hydrological model selection: A Bayesian alternative." *Water Resour. Res.* 41 (10): W10422. https://doi.org/10.1029/2004WR003719.

McCain, J. F., and R. D. Jarrett. 1976. *Manual for estimating flood characteristics of natural-flow streams in Colorado*. Denver: Colorado Water Conservation Board.

McIntyre, N., A. Al-Qurashi, and H. Wheater. 2007. "Regression analysis of rainfall–runoff data from an arid catchment in Oman." *Hydrol. Sci. J.* 52 (6): 1103–1118. https://doi.org/10.1623/hysj.52.6.1103.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea. 2012. *NHDPlus Version 2: User guide*. Washington, DC: National Operational Hydrologic Remote Sensing Center.

Miller, S. N., D. P. Guertin, and D. C. Goodrich. 1996. "Linking GIS and geomorphologic field research at Walnut Gulch Experimental Watershed." In *Paper Presented at 32nd Annual Conf. and Symp.: GIS and Water Resources*. Ft. Lauderdale, FL: American Water Resource Association.

Moran, M. S., et al. 2008a. "Preface to special section on Fifty Years of Research and Data Collection: US Department of Agriculture Walnut Gulch Experimental Watershed." *Water Resour. Res.* 44 (5): W05S01. https://doi.org/10.1029/2007WR006083.

Moran, M. S., C. D. Holifield Collins, D. C. Goodrich, J. Qi, D. T. Shannon, and A. Olsson. 2008b. "Long-term remote sensing database, Walnut Gulch Experimental Watershed, Arizona, United States." *Water Resour. Res.* 44 (5): W05S10. https://doi.org/10.1029/2006WR005689.

Morin, E., W. F. Krajewski, D. C. Goodrich, X. Gao, and S. Sorooshian. 2003. "Estimating rainfall intensities from weather radar data: The scale-dependency problem." *J. Hydrometeorol.* 4 (5): 782–797. https://doi.org/10.1175/1525-7541(2003)004<0782:ERIFWR>2.0.CO;2.

Nichols, M. H., and E. Anson. 2008. "Southwest watershed research center data access project." *Water Resour. Res.* 44 (5): W05S03. https://doi.org/10.1029/2006WR005665.

Niu, G., P. Troch, C. Paniconi, R. L. Scott, M. Durcik, X. Zeng, T. Huxman, D. C. Goodrich, and J. Pelletier. 2014. "An integrated modelling framework of catchment-scale ecohydrological processes. 2: The role of water subsidy by overland flow on vegetation dynamics in a semi-arid catchment." *Ecohydrology* 7 (2): 815–827. https://doi.org/10.1002/eco.1405.

Osborn, H. B., and R. B. Hickok. 1968. "Variability of rainfall affecting runoff from a semiarid rangeland watershed." *Water Resour. Res.* 4 (1): 199–203. https://doi.org/10.1029/WR004i001p00199.

Osborn, H. B., and L. J. Lane. 1969. "Precipitation–runoff relations for very small semiarid rangeland watersheds." *Water Resour. Res.* 5 (2): 419–425. https://doi.org/10.1029/WR005i002p00419.

Osborn, H. B., L. J. Lane, and V. A. Myers. 1980. "Rainfall/watershed relationships for southwestern thunderstorms." *Trans. ASAE* 23 (1): 82–87. https://doi.org/10.13031/2013.34529.

Osterkamp, W. R. 2008. "Geology, soils, and geomorphology of the Walnut Gulch experimental watershed, Tombstone, Arizona." *J. Ariz.-Nevada Acad. Sci.* 40 (2): 136–154. https://doi.org/10.2181/1533-6085-40.2.136.

Polyakov, V. O., M. A. Nearing, M. H. Nichols, R. L. Scott, J. J. Stone, and M. P. McClaran. 2010. "Long-term runoff and sediment yields from small semiarid watersheds in southern Arizona." *Water Resour. Res.* 46 (9): W09512. https://doi.org/10.1029/2009WR009001.

Renard, K. G., LJ. Lane, J. R. Simanton, W. E. Emmerich, J. J. Stone, M. A. Weltz, D. C. Goodrich, and D. S. Yakowitz. 1993. "Agricultural impacts in an arid environment: Walnut Gulch studies." *Hydrol. Sci. Tech.* 9 (1–4): 145–190.

Saft, M., M. C. Peel, A. W. Western, and L. Zhang. 2016. "Predicting shifts in rainfall-runoff partitioning during multiyear drought: Roles of dry period and catchment characteristics." *Water Resour. Res.* 52 (12): 9290–9305. https://doi.org/10.1002/2016WR019525.

Scott, R. L., and J. A. Biederman. 2017. "Partitioning evapotranspiration using long-term carbon dioxide and water vapor fluxes: New approach to ET partitioning." *Geophys. Res. Lett.* 44 (13): 6833–6840. https://doi.org/10.1002/2017GL074324.

SCS (Soil Conservation Service). 1972. *Hydrology national engineering handbook*. Washington, DC: SCS.

Sharma, K. D., and J. S. R. Murthy. 1996. "Ephemeral flow modelling in arid regions." *J. Arid. Environ.* 33 (2): 161–178. https://doi.org/10.1006/jare.1996.0054.

Sharma, K. D., and J. S. R. Murthy. 1998. "A practical approach to rainfall–runoff modelling in arid zone drainage basins." *Hydrol. Sci. J.* 43 (3): 331–348. https://doi.org/10.1080/02626669809492130.

Skirvin, S., M. Kidwell, S. Biedenbender, J. P. Henley, D. King, C. H. Collins, S. Moran, and M. Weltz. 2008. "Vegetation data, Walnut Gulch Experimental Watershed, Arizona, United States." *Water Resour. Res.* 44 (5): W05S08. https://doi.org/10.1029/2006WR005724.

Smith, R. E., D. L. Chery, K. G. Renard, and W. R. Gwinn. 1982. *Supercritical flow flumes for measuring sediment-laden flow.* USDA ARS Technical Bulletins 1655. Washington, DC: US Dept. of Agriculture, Agricultural Research Service.

Stone, J. J., M. H. Nichols, D. C. Goodrich, and J. Buono. 2008. "Long-term runoff database, Walnut Gulch Experimental Watershed, Arizona, United States." *Water Resour. Res.* 44 (5): W05S05. https://doi.org/10.1029/2006WR005733.

Stone, J. J., and G. B. Paige. 2003. "Variable rainfall intensity rainfall simulator experiments on semi-arid rangelands." In *Proc., 1st Interagency Conf. on Research in the Watersheds*, 83–88. Benson, AZ: US Dept. of Agriculture, Agricultural Research Service.

Syed, K. H., D. C. Goodrich, D. E. Myers, and S. Sorooshian. 2003. "Spatial characteristics of thunderstorm rainfall fields and their relation to runoff." *J. Hydrol.* 271 (1–4): 1–21. https://doi.org/10.1016/S0022-1694(02)00311-6.

Taormina, R., K. Kwok-Wing Chau, and B. Sivakumar. 2015. "Neural network river forecasting through baseflow separation and binary-coded swarm optimization." *J. Hydrol.* 529 (3): 1788–1797. https://doi.org/10.1016/j.jhydrol.2015.08.008.

USDA-ARS (Agricultural Research Service). n.d. "The long-term agroecosystem research network: USDA- 935 ARS." Accessed December 1, 2016. https://ltar.ars.usda.gov/.

Wu, C. L., and K. W. Chau. 2011. "Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis." *J. Hydrol.* 399 (3–4): 394–409. https://doi.org/10.1016/j.jhydrol.2011.01.017.

Zacharias, S., et al. 2011. "A network of terrestrial environmental observatories in Germany." *Vadose Zone J.* 10 (3): 955–973. https://doi.org/10.2136/vzj2010.0139.