Rainfall Thresholds for Flow Generation in Desert Ephemeral Streams

Stephanie K. Kampf, Joshua Faulconer, Jeremy R. Shaw, Michael Lefsky, and David J. Cooper

1Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, Colorado, USA; 2Department of Forest and Rangeland Stewardship, Colorado State University, Fort Collins, Colorado, USA; 3USDA Forest Service, Pacific Southwest Research Station, Arcata, California, USA

Abstract

Rainfall thresholds for streamflow generation are commonly mentioned in the literature, but studies rarely include methods for quantifying and comparing thresholds. This paper quantifies thresholds in ephemeral streams and evaluates how they are affected by rainfall and watershed properties. The study sites are in southern Arizona, USA; one is hyperarid and the other is semiarid. At both sites rainfall and streamflow were monitored in watersheds ranging from $10^{-3}$ to $10^{2} \text{ km}^2$. Streams flowed an average of 0–5 times per year in hyperarid watersheds and 3–11 times per year in semiarid watersheds. Although hyperarid sites had fewer flow events, their flow frequency (fraction of rain events causing flow) was higher than in semiarid sites for small (<1 km$^2$) watersheds. At both locations flow frequency decreased with drainage area, but the decrease was steeper in hyperarid watersheds. Watershed mean 60-min intensity thresholds ranged from 3–13 mm/hr in hyperarid watersheds and 7–16 mm/hr in semiarid watersheds. Higher runoff thresholds and lower flow frequencies in small semiarid watersheds likely relate to greater ground cover and soil development compared to the desert pavement and bedrock surfaces in hyperarid sites. The choice of rain data strongly influenced threshold values; single rain gauges were only adequate for threshold prediction in <1-km$^2$ watersheds, and incomplete rainfall data led to increases in thresholds with drainage area. We recommend using mean rainfall intensity over the drainage area for threshold analysis because this reduces apparent scale dependence in thresholds caused by incomplete rainfall information.

Plain Language Summary

Ephemeral streams in deserts are usually dry, flowing only after heavy rains. Our goal was to determine how much rain is needed for these streams to flow. We studied streams in dry southwestern and in wetter southeastern Arizona, USA. The dry site has mostly bare rock across the landscape, whereas the wetter site has more vegetation, including shrubs, grasses, and oak. In the dry area, small streams flowed 3–5 times per year, and larger streams flowed 0–2 times per year. Streams required 3–13 mm of rain over 1 hr to trigger flow. The small streams flowed more frequently because rain falling on bare rock could rapidly reach the stream. Water was lost into the channel bed before reaching larger streams. In the wetter area, both small and large streams flowed 3–11 times per year. These streams required 7–16 mm of rain over 1 hr to trigger flow. This range of rainfall is slightly higher than for streams in the dry study area, likely because vegetation cover and soil development allow more rainfall to infiltrate into soil before reaching streams. Information on the amount of rainfall needed to trigger streamflow can help with issuing flash flood warnings for ephemeral streams.

1. Introduction

Rainfall thresholds for streamflow generation are important components of streamflow prediction models and flood frequency distributions (Gioia et al., 2008; Kirkby et al., 2002, 2005). These thresholds indicate the lowest magnitude or intensity of rainfall that will cause streamflow. For infiltration excess overland flow at the point scale, these thresholds should be greater than or equal to the infiltration capacity of the soil, as overland flow develops when rainfall intensity exceeds this capacity plus detention storage. Over larger domains, however, threshold values may change with spatial scale, varying with flow path length (Yair & Raz-Yassif, 2004). Such scale-dependent patterns can be simulated using distributed models that represent spatially variable land cover, soil properties, and/or hydrologic connectivity (Kirkby et al., 2005; Plerini et al., 2014), but most study areas lack empirical data to verify the accuracy of simulated thresholds and their scale dependence.
One type of stream system where threshold-type responses are common is desert ephemeral streams (Osborn & Lane, 1969). Because these streams rarely flow, threshold-exceeding storms can be critical for seed dispersal, riparian vegetation establishment, and species diversity (Friedman & Lee, 2002; Shaw & Cooper, 2008; Stromberg et al., 2009). Several studies in arid and semiarid environments have identified rainfall depth or intensity thresholds for the occurrence of overland flow on hillslopes or for streamflow generation (Cammeraat, 2004; De Boer, 1992; Martínez-Mena et al., 1998; Mayor et al., 2011; Osborn & Lane, 1969). Of these, some have found that rainfall threshold values increase with drainage area (Cammeraat, 2004; Mayor et al., 2011). This increase has been attributed both to transmission losses along channels (Kirkby et al., 2005; Simanton & Osborn, 1983) and to spatial patterns of high-intensity storms, which often cover only small portions of larger watersheds (Goodrich et al., 1997; Morin et al., 2006; Syed et al., 2003). The relative contributions of rainfall spatial patterns and transmission losses to threshold scaling have not been quantified. Further while multiple studies mention rainfall thresholds for streamflow generation, methods for quantifying threshold values are rarely presented (Cammeraat, 2004; Yair & Klein, 1973), making it difficult to compare thresholds between study areas.

This research analyzes ephemeral streams draining small catchment to watershed-scale areas ($10^{-3}$ to $10^3$ km$^2$) in southern Arizona, USA, to identify rainfall thresholds for streamflow generation. The goal is to examine how rainfall patterns and drainage area size interact to affect threshold values. We compare thresholds for initiating streamflow in a hyperarid part of the Sonoran desert, USA, to those of the semiarid Walnut Gulch Experimental Watershed. We also examine how watershed vegetation cover and slope alter the thresholds. This information on spatial scaling of rainfall thresholds provides an empirical foundation for selecting parameters in hydrologic models, establishes lower bounds for flood warning systems, and can guide riparian management strategies for ephemeral stream systems.

2. Study Site

The hyperarid study area is located in the U.S. Army’s Yuma Proving Ground in southwestern Arizona, in the Sonoran desert. Annual precipitation averages 90 mm with most rain falling either during July–September or November–March. Summer storms are short-duration, high-intensity convective thunderstorms associated with the North American monsoon, and winter storms are from frontal systems that bring lower intensity, longer-duration rains (Hallack-Alegria & Watkins, 2007). The site is one of the hottest regions of the Sonoran desert with a mean annual minimum temperature of 16 °C and a mean annual maximum temperature of 31 °C. Maximum daily temperatures in July and August frequently exceed 40–42 °C.

The two study watersheds are Mohave Wash and Yuma Wash (Figure 1), which both drain to the Colorado River. Mohave Wash is 37 km long with a 323-km$^2$ drainage area, and Yuma Wash is 28 km long with a 188-km$^2$ drainage area. Elevation in Mohave Wash ranges from 210 to 720 m, and elevation in Yuma Wash ranges from 120 to 740 m. Mountainous portions of both watersheds consist primarily of exposed dacite and granodiorite, and lowland valleys are filled with Pliocene to Holocene alluvium. Gently sloping alluvial fans (bajadas) separate mountain blocks from lowland valleys, which contain active channels (Bacon et al., 2010). The alluvial fans are also known as piedmont surfaces, where desert pavement has formed on the surface of partially consolidated alluvial deposits. Desert pavement has a one- or two-particle thick layer of closely packed, darkly varnished gravel (>2 mm) and cobble (>60 mm) overlaying a vesicular layer of eolian fines with low permeability (McFadden et al., 1987; Springer, 1958). Documented infiltration rates in this area are less than 10 mm/hr (McDonald et al., 2004), which is substantially lower than saturated hydraulic conductivities measured on other surfaces in the region (Caldwell et al., 2012). Desert pavement surfaces have sparse vegetation consisting primarily of Larrea tridentata (creosote bush) and Ambrosia dumosa (white bursage) (Shreve & Wiggins, 1964). Vegetation is more abundant and diverse in riparian areas, including large woody plants such as Parkinsonia spp. (palo verde), Olneya tesota (iron wood), and Prosopis spp. (mesquite).

Streams in Mohave and Yuma Washes have been classified into five geomorphic types: piedmont headwater (PH), bedrock (BK), bedrock with alluvium (BA), incised alluvium (IA), and braided (BR) (Sutfin et al., 2014). Piedmont headwater channels are incised into desert pavement, and bedrock channels are headwater streams incised into bedrock. Bedrock with alluvium channels are partially confined by bedrock but have enough persistent alluvium in the channel to create bedforms. Incised alluvium channels are cut into the partially consolidated alluvial material of the piedmont, and they also contain modern alluvial bedforms. Braided
channels are the furthest downstream in the network, and they are large multithreaded channels underlain by deep modern alluvium. All channels have ephemeral flow, and those with modern alluvial fill have high permeability that allows rapid infiltration of channel flow (Kampf et al., 2016).

The semiarid study area is the Walnut Gulch experimental watershed (Figure 1), where mean annual precipitation is 312 mm. Walnut Gulch is 31 km long, with a drainage area of 149 km$^2$. Elevations are higher than in the Yuma area, ranging from 1,220 to 1,930 m. Bedrock geology is diverse, including sedimentary, plutonic, and volcanic rocks with some modern alluvium along channels. Soils are predominantly sand and gravel loams formed on erosion surfaces of the bedrock (Osterkamp, 2008). Land cover in Walnut Gulch is a mixture of shrubs, grass, and oak woodland, and the watershed also contains the small town of Tombstone, AZ (Skirvin et al., 2008). Channels in Walnut Gulch are all ephemeral, and they have width:depth ratios ranging from 2 to 100 (Miller, 1995). This range is consistent with the range of width:depth ratios for Mohave and Yuma Washes, with the exception of the braided channels, which have width:depth ratios $>100$ (Sutfin et al., 2014).

3. Methods

3.1. Instrumentation

Stream stage was monitored at 18 locations in Mohave and Yuma Washes (Table 1) using In-Situ Inc. Rugged TROLL 100 pressure transducers (In-Situ, Fort Collins, CO, USA); monitoring locations were clustered near the watershed outlets due to accessibility. Study sites had drainage areas of 0.002–225 km$^2$, and each watershed had stage measurements at two sites per geomorphic stream type except braided, which had one monitoring location per watershed. Sensors were placed inside vented PVC pipes for protection and bolted onto bedrock channel beds, banks, or trees to keep them in place during flow. The pressure transducers were unvented, so we corrected for barometric pressure using In Situ Inc. Rugged BaroTROLL loggers located at the braided sites. All pressure sensors were recorded at 15-min time steps. Rainfall was monitored with nine tipping bucket rain gauges, which were models RG3-M, TE525, or TB4 (Onset Corporation, Bourne, MA, USA;
## Table 1
Stream Stage Monitoring Locations, Characteristics of Their Contributing Areas and Channel Monitoring Locations, and Streamflow Event Summaries at Mohave and Yuma Washes Grouped by Channel Geomorphic Type

<table>
<thead>
<tr>
<th>Site</th>
<th>Area (km$^2$)</th>
<th>Site Elevation (m)</th>
<th>Dates Recording$^a$</th>
<th>Number Flow Events</th>
<th>Mean Annual Number Flow Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>YPH1</td>
<td>0.021</td>
<td>156</td>
<td>3/2012–3/2015</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>YPH2</td>
<td>0.002</td>
<td>154</td>
<td>3/2013–3/2015</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>MPH1</td>
<td>0.061</td>
<td>226</td>
<td>11/2011–3/2015</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>MPH2</td>
<td>0.049</td>
<td>242</td>
<td>2/2013–3/2015</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>YBK1</td>
<td>0.005</td>
<td>225</td>
<td>11/2011–5/2014</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>YBK2</td>
<td>0.013</td>
<td>166</td>
<td>3/2013–5/2014</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MBK1</td>
<td>0.093</td>
<td>464</td>
<td>11/2011–5/2014</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>MBK2</td>
<td>0.015</td>
<td>384</td>
<td>2/2013–5/2015</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>YBA1</td>
<td>2.2</td>
<td>211</td>
<td>11/2011–3/2015</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>YBA2</td>
<td>1.7</td>
<td>207</td>
<td>3/2013–10/2013</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MBA1</td>
<td>0.9</td>
<td>279</td>
<td>11/2011–3/2015</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>MBA2</td>
<td>0.8</td>
<td>328</td>
<td>3/2013–3/2015</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>YIA2</td>
<td>5.7</td>
<td>162</td>
<td>3/2013–5/2014</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MIA2</td>
<td>3.0</td>
<td>249</td>
<td>2/2013–3/2015</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>MIA1</td>
<td>17.0</td>
<td>270</td>
<td>11/2011–3/2015</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>YBD1</td>
<td>170</td>
<td>116</td>
<td>2/2012–3/2015</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>MBD1</td>
<td>225</td>
<td>201</td>
<td>4/2012–5/2014</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. Site IDs beginning with Y are in Yuma Wash, and those beginning with M are in Mohave Wash. PH are piedmont headwater, BK are bedrock, BA are bedrock with alluvium, IA are incised alluvium, and BD are braided channel locations. $^a$Some stations have data gaps within the period of record. Date formats are MM/YYYY.

Texas Electronics, Dallas, TX, USA; Hydrological Services Pty Ltd, Sydney, Australia) connected to either Campbell Scientific loggers (Campbell Scientific, Logan, UT, USA) or Hobo event loggers (Onset Corporation, Bourne, MA, USA). The full study duration was from November 2011 to March 2015, but due to variability in the timing of sensor installation, instrument failures, and other logistical issues, the period of record varied by location (Table 1).

For Walnut Gulch, we used the 2000–2016 discharge records from 19 flumes (Table 2). These flumes are installed on streams that have drainage areas ranging from 0.002 to 149 km$^2$. Sites with drainage areas >2 km$^2$ have supercritical flumes, and the smaller watersheds have Smith flumes. All flumes have continuous stage measurements from potentiometers attached to stilling well gears (Stone et al., 2008). We used the extensive digital rain gauge network for the watershed, which includes 88 Belfort weighing rain gauges (Belfort Instrument, Baltimore, MD, USA) that were continuously monitoring precipitation since 2000 (Goodrich et al., 2008).

### 3.2. Radar Precipitation
The rain gauge network at Mohave and Yuma Washes was not as extensive as at Walnut Gulch. Therefore, to examine spatial patterns of rainfall across the watersheds, we used NEXRAD radar data from the KYUX station in Yuma, Arizona, available through the NOAA National Center for Environmental Information (https://www.ncei.noaa.gov/). This station is 66–70 km south of the rain gauges in Yuma Wash and 103–108 km south of the rain gauges in Mohave Wash (Figure 1). Starting in May 2012, the KYUX station was upgraded to dual-polarization radar, so we examined radar data from May 2012 through March 2015. We used the level 3 one-hour precipitation NEXRAD product, which gives rainfall totals for a moving hour-long window at irregular time steps averaging 8 min in duration. To evaluate whether radar rainfall values were consistent with rain gauge data, we extracted radar rainfall time series for each pixel containing a rain
gauge and compared 60-min intensities between the radar and the rain gauges (Anagnostou et al., 2010; Xie et al., 2006). We tested for time offsets between radar and rain gauge data (Morin et al., 2003) and determined that no time offset adjustment was needed. However, because time steps did not match for radar and rain gauge data, we used a linear interpolation function in Matlab (version R2013a, MathWorks, Natick, MA, USA) to resample the radar values to the same time steps as the gauge values. We then computed a series of statistics to evaluate the accuracy of the radar values: false positive (FP), the percent of time steps when radar has rainfall but the gauge does not; false negative (FN), the percent of time steps when radar does not record rainfall but the rain gauge does; coefficient of determination ($R^2$); and root-mean-square error. Following this analysis, to map spatial patterns of rainfall across the study area, we used a mean field bias correction (Goudenhoofdt & Delobbe, 2009) and multiplied radar intensities by 0.67, the slope of the linear fit between all radar and rain gauge values (Figure S1).

### 3.3. Rainfall and Streamflow Patterns

We examined spatial and temporal rainfall patterns using both the rain gauge and radar data. We defined discrete rainfall events as any sequence of rainfall separated by at least 7 hr from the next period of rain. This interevent time limit was selected based on the range of interevent times reported elsewhere (Dunkerley, 2008) and the maximum flow duration observed at the watershed outlets in Mohave and Yuma Washes (Faulconer, 2015). To avoid including spurious tips of the rain gauges in event counts, we only considered events that reported at least 1 mm of rainfall and a maximum 60-min rainfall intensity > 1 mm/hr. For each defined event, we computed the total rain depth (P) and the maximum intensities at 15-, 30-, and 60-min time steps ($M_{15}$, $M_{30}$, $M_{60}$) for all rain gauges.

For each rainfall event, we documented whether or not the channels flowed. At Mohave and Yuma Washes, we identified flow events as times where the stage rose above the background noise in the sensor values, or >2 cm above the preevent stage. At Walnut Gulch, we identified flow events using any reported flow from the flume records. For each monitoring location we computed flow frequency (fraction of events causing flow, $P_{obs}$ (flow)) as the number of >1-mm events with streamflow divided by the total number of >1-mm events.

### 3.4. Rainfall Thresholds

We evaluated whether flow occurrence at each monitoring location could be predicted by a rainfall threshold. For a given rainfall event a threshold flow prediction is represented as

\[
\begin{align*}
\text{if } \text{rain} > T, & \text{ flow} \\
\text{if } \text{rain} \leq T, & \text{ no flow}
\end{align*}
\]

where rain is any of the rainfall metrics (P, $M_{15}$, $M_{30}$, $M_{60}$) and T is the threshold. We identified T as the value that optimizes the fraction of events correctly predicted, $P_e$. This was accomplished by iterating through values of T in increments of 0.1 units (mm or mm/hr). If multiple values of T gave the same $P_e$ value, the lowest of the T values was selected as the threshold.

For threshold analysis with rain gauge data at Mohave and Yuma Washes, we used the rain gauge closest to each stream monitoring location and computed thresholds for all rainfall metrics. For threshold analysis with radar data, we used only $M_{60}$ because this is the original unit provided in the radar product, and subhourly time steps were irregular. We computed the radar maximum $M_{60}$ as the maximum intensity for any time step and any pixel in the drainage area. We also computed radar mean $M_{60}$ by first computing the mean $M_{60}$ for each time step over all pixels in the drainage area, then taking the maximum of these drainage area mean $M_{60}$ values. For threshold analysis at Walnut Gulch, we computed thresholds for each flume using $M_{60}$ values from each individual rain gauge in the flume’s contributing area. This allowed us to evaluate how thresholds vary with proximity of rain gauge. We also computed thresholds using the maximum and mean $M_{60}$ for all rain gauges in the contributing area.

We examined how flow frequency and thresholds varied with drainage area size for each study location independently. We then tested for regional differences in the scale dependence of these variables using a generalized linear model of location (region), drainage area, and an interaction term. This analysis was performed using Proc GLM in SAS 9.4.
3.5. Multivariate Analysis

To examine whether other variables besides rainfall affected streamflow occurrence, we used multivariate logistic regression in JMP version 13.0.0 (Klimberg & McCullough, 2012). For independent variables we used rainfall ($M_{60}$), drainage area, percent vegetation cover, mean slope, and antecedent precipitation index (API). For $M_{60}$, we used watershed mean values from rain gauges at Walnut Gulch; at Mohave and Yuma we used individual rain gauge values for small watersheds ($< 0.1 \text{ km}^2$) and radar mean values for all other watersheds.

We computed drainage area and slope from 10-m digital elevation models using the watershed and slope tools in ArcGIS (ESRI, Redlands, CA, USA). For vegetation cover we used LANDFIRE Existing Vegetation Cover (LANDFIRE, 2014), which gives cover in 10% interval bins, and with these values we computed the area-weighted mean vegetation cover for each drainage area. We computed API following the method of Linsley et al. (1982) as

$$API_t = P_t + P_0 k^t$$  \hspace{1cm} (2)

where $t$ is the day, $P_o$ is the daily total precipitation for the most recent prior day with rain, and $k$ is a constant. For any day with rainfall, $P_o$ resets to that day’s total precipitation, and $t$ resets to 0. We set $k$ to 0.8 to represent rapid loss of moisture based on observed water contents in the floodplains near the study area channels (Kampf et al., 2016).

We created logistic regressions to predict flow occurrence for sample subsets of Mohave and Yuma, Walnut Gulch, and all study sites combined. Each regression used log-transformed and standardized values of the independent variables. We tested for collinearity of the independent variables and found that all of the independent variables were significantly correlated with one another. Therefore, we then conducted principal component analysis on the independent variables. $M_{60}$ had the strongest correlation to PC1 (0.86), followed by API (0.82); slope had the highest correlation to PC2 (0.56), cover to PC3 (0.79), and area to PC4 (0.47).

3.6. Performance Evaluation

We evaluated both rainfall threshold and logistic regression representations of flow using four metrics: $p_0$, $\kappa$, FP, and FN. The observed agreement, $p_0$, is the number of event responses ($p_0$(flow) or $p_0$(no flow)) correctly represented by the threshold divided by the total number of events. The kappa statistic of agreement, $\kappa$, is computed as

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$  \hspace{1cm} (3)

$$p_e = p_{obs}(\text{flow}) \times p_T(\text{flow}) + p_{obs}(\text{no flow}) \times p_T(\text{no flow})$$  \hspace{1cm} (4)

where $p_{obs}$ is the fraction of observations of a specified type (flow or no flow) and $p_T$ is the fraction of threshold predictions of the specified type. Higher values of $\kappa$ indicate greater agreement between the threshold predictions and the observations. Following Viera and Garrett (2005), values of $\kappa > 0.4$ indicate moderate or better agreement, values $>0.6$ indicate substantial agreement, and values $>0.8$ indicate almost perfect agreement. We computed the false positives (FP) as the percent of all events for which flow was predicted when no flow was observed and false negatives (FN) as percent of all events for which flow was not predicted when flow was observed.

4. Results

4.1. Rainfall

From November 2011 to March 2015 the highest total rainfall in Mohave and Yuma Washes was in upper Mohave Wash (>500 mm) and the lowest in lower Yuma Wash (<200 mm). Many >1-mm rain events were highly localized; 15% had rain recorded at only one gauge, and only 21% reported rain at all gauges. Of the recorded rain storms, 54% were in summer (May–October; Hallack-Alegría & Watkins, 2007), and these had higher mean precipitation and maximum intensities than winter storms. For pixels containing rain gauges, timing of radar rainfall was similar to gauge rainfall, with FP and FN errors <1%.
Table 3
Performance of Rainfall Metrics for Threshold Predictions of Flow or No Flow

<table>
<thead>
<tr>
<th>Rainfall Variable</th>
<th>Data Source</th>
<th>$p_o$</th>
<th>$\kappa$</th>
<th>FP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>Closest gauge</td>
<td>0.92</td>
<td>0.66</td>
<td>2.5</td>
<td>5.3</td>
</tr>
<tr>
<td>$M_{15}$</td>
<td>Closest gauge</td>
<td>0.91</td>
<td>0.62</td>
<td>4.5</td>
<td>4.8</td>
</tr>
<tr>
<td>$M_{30}$</td>
<td>Closest gauge</td>
<td>0.91</td>
<td>0.61</td>
<td>2.2</td>
<td>6.4</td>
</tr>
<tr>
<td>$M_{60}$</td>
<td>Closest gauge</td>
<td>0.93</td>
<td>0.71</td>
<td>2.8</td>
<td>4.2</td>
</tr>
<tr>
<td>$M_{60} \text{ max}$</td>
<td>Radar</td>
<td>0.94</td>
<td>0.44</td>
<td>1.4</td>
<td>3.6</td>
</tr>
<tr>
<td>$M_{60} \text{ mean}$</td>
<td>Radar</td>
<td>0.95</td>
<td>0.53</td>
<td>0.8</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Walnut Gulch

<table>
<thead>
<tr>
<th>Rainfall Variable</th>
<th>Data Source</th>
<th>$p_o$</th>
<th>$\kappa$</th>
<th>FP</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>All rain gauges</td>
<td>0.87</td>
<td>0.47</td>
<td>2.0</td>
<td>10.9</td>
</tr>
<tr>
<td>$M_{\text{max}}$</td>
<td>All rain gauges</td>
<td>0.88</td>
<td>0.47</td>
<td>3.2</td>
<td>9.0</td>
</tr>
<tr>
<td>$M_{60} \text{ max}$</td>
<td>All rain gauges</td>
<td>0.91</td>
<td>0.61</td>
<td>3.1</td>
<td>5.4</td>
</tr>
<tr>
<td>$M_{60} \text{ mean}$</td>
<td>All rain gauges</td>
<td>0.90</td>
<td>0.62</td>
<td>2.7</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Note: $p_o$ is the fraction of events correctly predicted, $\kappa$ is the kappa agreement statistic, FP is the percent of false positives, and FN is the percent false negatives.

Overall the fit between radar intensities and gauge intensities was good ($R^2 = 0.70$). Radar rainfall totals were highest in upper Mohave and lowest in lower Yuma (Figure S3a). Over the nearly three-year period of record for the radar data, $M_{60}$ values for individual radar pixels were 10 mm/hr or higher for up to seven events in Yuma Wash and up to 12 events in Mohave Wash (Figure S3b). Nearly all of these high-intensity storms were in the summer.

In Walnut Gulch, during the period of measurement at Mohave and Yuma, total precipitation was more than twice that of Mohave and Yuma Washes, ranging from 931 to 1271 mm (Figure S4a). Highest rainfall totals were in the southeast half of the watershed, although the patterns were locally variable. On average, 80% of the total precipitation fell during the summer. During the May 2012–March 2015 time period, rain gauges at Walnut Gulch recorded 67 to 108 events with $M_{60}$ exceeding 10 mm/hr (Figure S4b). This is nearly an order of magnitude higher than the number of high-intensity events in Mohave and Yuma Washes. On average, 99% of these high-intensity events were during the summer season (May–October). Of the >1-mm rainfall events, only 10% were recorded at all rain gauges, and on average, 45% of the gauges reported rainfall during an event.

4.2. Streamflow

At Mohave and Yuma Washes, streamflow was infrequent, ranging from 0 events at YBA2 and YIA2 to 16 flow events at MPH1 (Table 1) over the two- to three-year monitoring period. Seventy-seven percent of recorded flows were in summer. Winter flows were recorded in piedmont headwater sites, Yuma bedrock sites, one bedrock with alluvium site (MBA1), and one incised alluvium site (YIA1). Flows were highly localized, with an average of 30% of the monitoring sites recording flow during runoff-producing storms. Only one storm (13 July 2012) produced flow at all gauges, and this storm had the largest gauge and radar average precipitation at 87 and 69 mm, respectively (Figure S5 and Movie S1).

At Walnut Gulch, all stations recorded streamflow during the monitoring period, and streams flowed an average of 3–11 times per year (Table 2). Flows were heavily concentrated in summer. More than 95% of flow events were during summers for the sites with drainage areas <0.1 km$^2$, and 98–100% of flow events were in summer in all other sites. The flows were usually localized in portions of Walnut Gulch, with 31% of flumes on average recording flow during runoff-producing storms.

4.3. Rainfall Thresholds for Streamflow

All rainfall metrics tested worked well for threshold prediction at Mohave and Yuma Washes (Table 3; $p_o \geq 0.91$, $\kappa \geq 0.44$). Precipitation depth ($P$) thresholds generally increased with drainage area from 5 to 32 mm when using the closest rain gauge, with higher thresholds in Mohave than Yuma Wash (Figure 2a). These thresholds had moderate or better performance ($\kappa > 0.4$) at most of the sites (Figure 2b). $M_{60}$ thresholds performed best overall using the closest rain gauge (Table 3). Threshold values from radar maximum $M_{60}$ ranged from 5 to 14 mm/hr in watersheds <10 km$^2$ and from 18 to 19 mm/hr in the watersheds >100 km$^2$ (Figure 2c), with the exception of three sites that had thresholds >40 mm/hr (YPH2, MBA1, MBD1). The radar threshold performance using the maximum $M_{60}$ was fair or poor at most sites (Figure 2d). Mean $M_{60}$ thresholds from radar ranged from 3 to 13 mm/hr (Figure 2e) and had a weak exception with the log of drainage area. These thresholds performed well at most sites, with the exception of small watersheds <0.1 km$^2$ (Figure 2f) because these are smaller than the ~1-km$^2$ resolution of the radar.

Threshold values for Walnut Gulch varied with the rainfall metric used; overall, $M_{60}$ worked better ($\kappa = 0.61–0.62$ for maximum and mean $M_{60}$, respectively) than $P$ ($\kappa = 0.47$ for both maximum and mean; Table 3). $M_{60}$ threshold values derived from individual rain gauges tended to increase with rain gauge distance from the flume (Figure 3a), and the performance of these threshold predictions declined sharply with distance from the flume (Figure 3b). The $\kappa$ values indicated moderate or better performance (>0.4) for small watersheds (<25 km$^2$) in which the rain gauge was <2.5 km from the flume. When threshold analysis used rain
Figure 2. (a, c, and e) Rainfall thresholds versus drainage area and (b, d, and f) threshold performance $k$ statistic versus drainage area for Mohave (M), Yuma (Y), and Walnut Gulch using the event total precipitation ($P$) from the (a and b) closest rain gauge, (c and d) maximum storm $M_{1,60}$ across the watershed, and (e and f) mean storm $M_{1,60}$ across the watershed. Rainfall values in (e) and (f) are from radar for Mohave and Yuma and all rain gauges in the watershed at Walnut Gulch. Dashed horizontal line indicates the performance value above which threshold prediction is considered “moderate.” For (c), three sites had thresholds higher than the range shown: YPH2 40 mm/hr and MBA1 and MBD1 44 mm/hr.

gauges from across the watershed area, watershed mean $P$ thresholds ranged from 12 to 27 mm, but most of these had only fair or poor agreement with observations (Figures 2a and 2b). Both the maximum and mean $M_{1,60}$ from all watershed rain gauges performed better for threshold prediction (Table 3; $P > 0.90$, $k > 0.61$), with moderate or better performance in the small watersheds (<1 km$^2$) and declining performance in larger watersheds (Figure 2d). The maximum $M_{1,60}$ threshold values increased with drainage area from around 6 mm/hr in the smallest watersheds up to over 26 mm/hr in the largest watersheds (Figure 2c). In contrast, the mean $M_{1,60}$ thresholds were in the 7–14-mm/hr range for all watersheds (Figure 2e). Thresholds for maximum and mean $M_{1,60}$ were the same for watersheds <0.1 km$^2$; the difference between these
Figure 3. (a) $M_{60}$ rainfall thresholds from individual rain gauges and (b) $\bar{x}$ performance statistic versus distance from rain gauge to flume for Walnut Gulch. Small watersheds are grouped together by drainage area, and the three largest watersheds (6, 2, 1) each have their own color and symbol. Dashed horizontal line indicates the performance value above which threshold prediction is considered “moderate.”

Thresholds increased with the log of drainage area for watersheds $> 1 \text{ km}^2$ (Figure 4b). This difference in threshold values with drainage area was also evident in $P$ thresholds (Figure 4a).

Thresholds for streamflow are inversely related to flow frequency ($p_0$ (flow)) because lower thresholds lead to more frequent streamflow. At Mohave and Yuma Washes, the piedmont headwater (PH) streams flowed 8–16 times, with 0.38 average $p_0$ (flow) (Figure 5). Bedrock (BK) headwater streams had lower average $p_0$ (flow) (0.22), with two to eight events recorded at each site. The larger drainage area bedrock with alluvium (BA) sites flowed 0–8 times, with $p_0$ (flow) averaging 0.11. Similarly, incised alluvium (IA) sites flowed 0–6 times, with 0.11 average $p_0$ (flow). Finally, the furthest downstream braided (BD) sites flowed 1–2 times, with mean $p_0$ (flow) of 0.05. Flow frequencies ($p_0$ (flow)) values at the hyperarid Mohave and Yuma sites were higher than those in Walnut Gulch for small watersheds but not for large watersheds. The $< 0.1 \text{ km}^2$ watersheds had $p_0$ (flow) between 0.10 and 0.24, lower than corresponding small watersheds at Mohave and Yuma, whereas the larger watersheds had $p_0$ (flow) between 0.09 and 0.14. For watersheds $< 1 \text{ km}^2$, generalized linear model coefficients confirm that mean $p_0$ (flow) was higher ($p = 0.020$) and declined with drainage area faster ($p = 0.014$) at Mohave and Yuma Washes than at Walnut Gulch (Table S2). Flow frequency in watersheds $> 1 \text{ km}^2$ did not differ among Mohave/Yuma and Walnut Gulch and declined with drainage area at similar rates in both regions.

We also evaluated how thresholds varied with other independent variables describing watershed characteristics. Comparing the two study regions, rainfall thresholds generally increased with watershed vegetation cover and decreased with watershed mean slope (Figure 6), although this pattern was not clearly evident within regions. Cover is higher, and slope is lower at most Walnut Gulch sites compared to Mohave and

Figure 4. Difference in (a) $P$ and (b) $M_{60}$ thresholds between maximum and mean watershed values versus drainage area for Walnut Gulch.
Yuma Washes, and this is associated with higher threshold values. Standardized coefficients from multivariate logistic regression analyses illustrate the relative importance of rainfall and watershed characteristics for flow occurrence (Table 4). Standardized slope coefficients were highest for MI_{60} in both regions (−2.1 to −3.5), and logistic regression with MI_{60} alone performed nearly as well as logistic regression with the other independent variables included. Regressions using rainfall alone were 80% accurate for Mohave and Yuma (radar rainfall) and 92% accurate for Walnut Gulch (rain gauge average). Adding area, slope, cover, and API improved regression performance to 84% accuracy at Mohave and Yuma but did not affect model accuracy at Walnut Gulch; collinearity of the independent variables limited the value of adding the additional information. Comparison of standardized coefficients demonstrates that drainage area (0.8) and mean slope (0.2) were the most influential watershed characteristics in Mohave/Yuma, whereas vegetation cover and API had little effect. For Walnut Gulch, slope, cover, and API were all significant variables in the regression, although accuracy did not improve with the addition of more independent variables. API (standardized coefficient 0.4) was the most important watershed characteristic followed by vegetation cover (−0.3) and mean slope (0.2). The logistic regression using all sites combined was similar to the model for Walnut Gulch because of the larger data set compared to Mohave and Yuma Washes (ρ_{0} = 0.91, κ = 0.56). The majority of errors in the logistic regression predictions were false negatives (Table 4), similar to the errors for rainfall threshold predictions (Table 3).

5. Discussion

5.1. Threshold Performance and Variability

At both study areas, rainfall thresholds are strong predictors of streamflow occurrence, except in some of the >1-km² Walnut Gulch watersheds. Our results show slightly stronger threshold predictions with MI_{60} than with total event precipitation (P) or shorter-duration maximum intensities (MI_{15}, MI_{30}), although the other rainfall metrics work nearly as well (Table 3) because these rainfall variables are highly correlated with one another (R² = 0.95–0.98). Prior studies in this region have focused on shorter-duration maximum intensities (30 min or less), based on the idea that short-duration, high-intensity bursts of rainfall are responsible for overland flow generation (Osborn & Lane, 1969; Syed et al., 2003). However, similar to our findings, Koterba (1986) found that models predicting total and peak flow had similar performance with 10-, 20-, and 30-min intensities, and model performance tended to increase with duration considered. Given this tendency for better predictions using longer durations, we focused on 60 min because of the improved

Figure 5. Flow frequency (p_{0}(flow)) versus drainage area for Mohave Wash (M), Yuma Wash (Y), and Walnut Gulch. Flow frequency is the fraction of >1-mm/hr rainfall events that produced flow, where rainfall values were radar mean MI_{60} for Mohave and Yuma and rain gauge mean MI_{60} for Walnut Gulch. Line fits shown as solid lines with dashed lines for 95% confidence intervals.

Figure 6. MI_{60} rainfall thresholds versus (a) watershed percent vegetation cover and (b) watershed average slope. Threshold values are those derived from radar mean MI_{60} (Mohave, Yuma) and rain gauge mean MI_{60} (Walnut Gulch).
Table 4
Logistic Regression Equations and Performance for Representing Flow or No Flow Response

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Yuma, Mohave</th>
<th>Walnut Gulch</th>
<th>All Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MI60</td>
<td>All</td>
<td>MI60</td>
</tr>
<tr>
<td>$b_1$ (MI60)</td>
<td>$-1.8 (0.25)^*$</td>
<td>$-2.1 (0.34)^*$</td>
<td>$-3.1 (0.07)^*$</td>
</tr>
<tr>
<td>$b_2$ (area)</td>
<td>$0.82 (0.22)^*$</td>
<td>$0.23$</td>
<td>$-0.07 (0.04)$</td>
</tr>
<tr>
<td>$b_3$ (slope)</td>
<td>$0.19 (0.09)^*$</td>
<td>$-2.1 (0.13)$</td>
<td>$-0.28 (0.05)^*$</td>
</tr>
<tr>
<td>$b_4$ (cover)</td>
<td>$-0.12 (0.13)$</td>
<td>$0.04 (0.13)$</td>
<td>$0.37 (0.07)^*$</td>
</tr>
<tr>
<td>$b_5$ (API)</td>
<td>$0.04 (0.13)$</td>
<td>$0.42$</td>
<td>$0.42$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.23</td>
<td>0.34</td>
<td>0.42</td>
</tr>
<tr>
<td>AICc</td>
<td>295</td>
<td>211</td>
<td>6779</td>
</tr>
<tr>
<td>$p_0$</td>
<td>0.80</td>
<td>0.84</td>
<td>0.92</td>
</tr>
<tr>
<td>$x$</td>
<td>0.32</td>
<td>0.50</td>
<td>0.57</td>
</tr>
<tr>
<td>FP</td>
<td>4.0</td>
<td>3.8</td>
<td>2.0</td>
</tr>
<tr>
<td>FN</td>
<td>15.8</td>
<td>12.8</td>
<td>6.5</td>
</tr>
</tbody>
</table>

Note. All independent variables are log-transformed and standardized. MI60 is the watershed mean value, API is the antecedent precipitation index, and $b$s are the regression coefficients for the indicated variables followed by the standard error in parentheses.

*This indicates significance ($p < 0.05$).
using maximum $M_{60}$ at Walnut Gulch (Figure 2c). We therefore recommend using the drainage area mean $M_{60}$ from a dense rain gauge network or from radar for threshold prediction in watersheds >1 km$^2$, as this avoids confounding the partial storm coverage effect on threshold values with the transmission loss effect.

While a transmission loss effect is not clear in mean $M_{60}$ thresholds for Walnut Gulch (Figure 2e), transmission loss is evident for flow frequency, which declines with increasing drainage area (Figure 5). Prior research has also demonstrated that transmission losses affect flow magnitude in these watersheds (Goodrich et al., 1997). Possibly the uncertainties in defining rainfall thresholds made this scaling behavior less evident in threshold values than in flow frequency, particularly because the thresholds did not perform well for most of the >1-km$^2$ watersheds (Figure 2d). The transmission loss effect on threshold values is more apparent for Mohave and Yuma Washes (Figure 2e). The flow frequency also has a steeper decline with drainage area at Mohave and Yuma compared to Walnut Gulch (Figure 5 and Table S2). This could be because transmission losses are even greater in those watersheds, particularly for the extremely wide braided channels; excluding the braided channels, the area dependence of flow frequency (Figure 5) and mean $M_{60}$ thresholds (Figure 2e) would not be as evident.

5.2. Threshold Comparisons to Other Locations

Studies of rainfall thresholds for overland flow generation in other regions have used a wide variety of rainfall metrics, including intensities over varying durations and total precipitation. The only consistent metric across all studies is depth of precipitation ($P$), so we use $P$ to compare thresholds between regions. We identified comparable study sites in Spain, which have mean annual precipitation ranging from 235 to 274 mm (Cammeraat, 2004; Mayor et al., 2011; Rodriguez-Caballero et al., 2014; Figure 7) and Israel, which have mean annual precipitation from 28 to 472 mm (Greenbaum et al., 2006; Ries et al., 2017; Yair & Klein, 1973). For the sites in Spain, where the mean annual precipitation is similar to Walnut Gulch, reported thresholds mostly fall within the range of the precipitation depth thresholds we identified (5–35 mm). These studies found increases in precipitation thresholds with increased drainage area (Cammeraat, 2004; Mayor et al., 2011). Spatial patterns of rainfall were not the focus, but the catchments examined were mostly smaller than the ~1-km$^2$ area at which spatial variability in rainfall changes threshold values (Figure 4a), indicating that transmission losses likely caused the threshold-scale dependence. In Israel, small (<1 km$^2$) watershed studies were in the Negev desert, where the surface cover is mostly bare rock (Greenbaum et al., 2006), resulting in thresholds at the lower end of the range we found in Arizona (<10 mm). Larger watersheds in Israel examined by Ries et al. (2017) have a steep precipitation gradient, ranging from <200 mm at the watershed outlets up to over 600 mm in the headwaters. Their vegetation cover ranges from 21 to 24%, which is a similar cover range to Mohave and Yuma watersheds, yet the documented thresholds for streamflow were near 50 mm, higher than any of the thresholds we found. Ries et al. (2017) had a network of 15 rain gauges, but the spacing between some gauges was more than 20 km. Based on our findings, the thresholds they identified would likely have been lower if they had a denser rain gauge network and used watershed mean values. When all study regions are combined, precipitation thresholds increase significantly with the log of drainage area (Cammeraat, 2004; Mayor et al., 2011).

5.3. Other Factors Affecting Thresholds and Future Applications

In addition to drainage area size, variability in threshold values can also relate to topography and ground surface conditions. Comparing our two study areas, the $M_{60}$ mean thresholds tended to be higher in Walnut Gulch than Mohave and Yuma Washes, particularly for smaller (<1 km$^2$) watersheds. This may be due to differences in ground cover between sites (Figure 6a), as surfaces with greater rock cover can generate runoff at lower rainfall thresholds (Abrahams & Parsons, 1991; Yair & Kossovsky, 2002). Mohave and Yuma headwater watersheds were either bedrock or desert pavement, and lower elevation surfaces had primarily 10–20% shrub cover. Of these land cover types, the piedmont headwater (PH) areas with desert pavement at
Mohave and Yuma had the highest flow frequencies. Low infiltration rates of desert pavement soils have been attributed to the saline-sodic soils below the rocky surface, where deflocculation of soil colloids during rainfall reduces infiltration rates (Musick, 1975). Comparisons of infiltration rates for desert pavements with different ages suggest that pedogenic processes can reduce infiltration rates over time (Meadows et al., 2008). Thresholds we documented on these surfaces are within the range of saturated hydraulic conductivities measured for relatively old (4–100 ka) surfaces (3–6 mm/hr; Young et al., 2004). Consequently, the more frequent runoff on desert pavement is likely sensitive to disturbances that disrupt the pavement or soil below the rocky surface. Compared to desert pavement, watersheds with bedrock channels had less frequent flow because the exposed rhyolite and dacite is highly fractured, and surface topography is irregular enough to cause detention storage.

Pavement and bedrock were not prominent land cover types at the wetter and lower slope Walnut Gulch, where most surfaces had soil that allowed at least some infiltration. Walnut Gulch has a mixture of 30–40% cover of shrub- and grass-dominated vegetation. Saturated hydraulic conductivities measured at the surface of Walnut Gulch soils average 20 mm/hr for crusted soils and 25 mm/hr for uncrusted soils, but both soil types can have saturated hydraulic conductivities as low as 4–5 mm/hr (Becker et al., 2018). Rainfall thresholds from Walnut Gulch range from 7 to 16 mm/hr, which is lower than the mean but within the range of documented hydraulic conductivities. Research in semiarid Spain found that crusted areas were important sources of flow (Marchamalo et al., 2016; Rodríguez-Caballero et al., 2014), and this may explain why the thresholds at Walnut Gulch are less than the average hydraulic conductivities. Vegetation patterns also affect infiltration and overland flow and consequently the rainfall thresholds for response. Infiltration can be more common near shrubs than in bare spaces between, and the distribution of vegetation patches across catchments affects whether flow connects to downstream channels (Lesschen et al., 2009; Puigdefàbregas et al., 1999). Further study of spatial patterns in soil crust and vegetation may help explain some of the variability in thresholds.

Although other studies have suggested that antecedent soil moisture may change the threshold required for overland flow (De Boer, 1992), API was only useful for threshold flow prediction in Walnut Gulch (Table 4). Extremely high evapotranspiration rates cause soils to dry rapidly at Mohave and Yuma Washes (Kampf et al., 2016), which may limit the role of initial wetness in runoff generation. The relatively short period of record may also have caused us to miss conditions in which antecedent conditions were important. While API was a significant predictor of flow occurrence for Walnut Gulch, it did not improve the performance of the logistic regression, potentially indicating that its role is also limited in Walnut Gulch. However, API was significantly correlated to ML_{40} (r = 0.86), so its effects on flow could not be assessed independently. The API values we used were just a proxy for initial wetness, and measurements of soil moisture across the drainage areas could help determine if and how antecedent wetness affects threshold values.

All thresholds we identified here were affected by our method of threshold identification. It is important to use a consistent method when comparing between sites. Our method optimizes prediction accuracy (p₀), and because these streams flow so infrequently, this results in mainly false negative errors. Optimizing to a different metric such as κ would modify the threshold values. The methods we presented only determined presence or absence of flow, not flow magnitude, and they are intended to allow first-order comparison of rainfall-runoff between scales and regions. Further research could relate these findings to information on flow magnitude, more attributes of the storm systems such as size, velocities, and locations (Belachsen et al., 2017; Morin et al., 2006; Morin & Yakir, 2014), and/or attributes of the catchments such as drainage network geometry (Dick et al., 1997; Kirckby et al., 2005).

One benefit of a threshold analysis is that it allows comparison across sites with different gauge densities and precipitation records. A heavily instrumented watershed like Walnut Gulch is expensive to build and maintain. While this level of instrumentation is important for examining hydrologic processes, if we base our hydrologic understanding solely on locations with extensive long-term instrumentation, we will miss important sources of variability in hydrologic response. Rainfall threshold analysis is a low-cost method that does not rely on measuring channel discharge, which is challenging in desert ephemeral streams. Thresholds enable comparisons of scaling behavior between sites (Figure 7) and allow us to examine the relative importance of rainfall and watershed properties on streamflow response (Table 4). Our results highlight how the choice of rainfall data affects both the magnitude of thresholds for an individual watershed and the apparent scale dependence of thresholds between watersheds. Therefore, future studies on the scaling of both rainfall
thresholds and runoff volumes should incorporate dense rain gauge networks and/or radar rainfall data for larger watersheds.

In future applications, rainfall thresholds of flow initiation can be used in flash flood warning systems as lower envelopes of the rainfall conditions likely to cause flooding in dryland ephemeral streams. Although such warning systems focus on flow magnitude (Carpenter et al., 1999; Clark et al., 2014; Gourley et al., 2014; Martina et al., 2005; Morin et al., 2009; Reed et al., 2002, 2007), it can be difficult to forecast ephemeral stream-flow accurately, and the threshold-based predictions developed here can help reduce some of the false positive or false negative errors in existing prediction systems. Field hydrologists can use rainfall thresholds to predict times of erosion and sediment transport and determine when field visits for sample collection are needed. Rainfall thresholds combined with rainfall frequency information are also useful for riparian restoration applications, as they can help predict flow frequency, which affects plant water availability and seed dispersal.

6. Summary and Conclusions

Ephemeral channels in the hyperarid Mohave and Yuma Washes experienced an order of magnitude fewer high-intensity storms than those in semiarid Walnut Gulch, but the two sites had similar flow frequencies and thresholds for streamflow generation. The occurrence of streamflow could be reproduced to >90% accuracy by 60-min rainfall intensity thresholds, except in the larger watersheds of Walnut Gulch. The magnitudes of these thresholds varied with the choice of rain gauge data, whether aggregated across watersheds or using a single gauge location. Single rain gauges were only useful for threshold prediction in watersheds <25 km², where the rain gauge was within 2.5 km of the watershed outlet. For watersheds longer or wider than 2.5 km, multiple rain gauges and/or radar rainfall data are needed for reliable threshold prediction. Scale dependence of threshold values can be an artifact of the choice of rainfall data, as incomplete rainfall information can lead to increases in thresholds with greater drainage area. To provide a consistent metric for comparing thresholds across different watersheds, we recommend using watershed mean rainfall intensities, which dampen the scale dependence of thresholds caused by partial area storm coverage. Thresholds using watershed mean intensities have lower scale dependence in semiarid than in hyperarid watersheds. Thresholds are lowest for small hyperarid watersheds with relatively impermeable desert pavement or bedrock compared to the semiarid watersheds that have higher soil and vegetation cover.

Although many hydrologic applications require information on total flow magnitude, obtaining such data is challenging in ephemeral streams, which have dynamic channel geometry and unstable rating curves. Better understanding of rainfall thresholds that produce streamflow can help fill in gaps in discharge records and improve prediction of when and where flow is likely. The thresholds determined for our study sites are consistent with the ranges of thresholds documented in other arid and semiarid regions and are within the range of saturated hydraulic conductivities documented for the study areas. Future research in new study areas can add to the data sets presented here and help refine methods for using scale, land cover, and other watershed properties to predict rainfall thresholds in ungauged streams.

References


