



A top-down model to generate ensembles of runoff from a large number of hillslopes

P. R. Furey¹, V. K. Gupta², and B. M. Troutman³

¹NorthWest Research Associates, Boulder, CO, USA

²Dept. of Civil, Environmental and Architectural Engineering, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

³Lakewood, CO, USA

Correspondence to: P. R. Furey (prfurey@gmail.com)

Received: 31 October 2012 – Revised: 13 June 2013 – Accepted: 30 July 2013 – Published: 25 September 2013

Abstract. We hypothesize that total hillslope water loss for a rainfall–runoff event is inversely related to a function of a lognormal random variable, based on basin- and point-scale observations taken from the 21 km² Goodwin Creek Experimental Watershed (GCEW) in Mississippi, USA. A top-down approach is used to develop a new runoff generation model both to test our physical-statistical hypothesis and to provide a method of generating ensembles of runoff from a large number of hillslopes in a basin. The model is based on the assumption that the probability distributions of a runoff/loss ratio have a space–time rescaling property. We test this assumption using streamflow and rainfall data from GCEW. For over 100 rainfall–runoff events, we find that the spatial probability distributions of a runoff/loss ratio can be rescaled to a new distribution that is common to all events. We interpret random within-event differences in runoff/loss ratios in the model to arise from soil moisture spatial variability. Observations of water loss during events in GCEW support this interpretation. Our model preserves water balance in a mean statistical sense and supports our hypothesis. As an example, we use the model to generate ensembles of runoff at a large number of hillslopes for a rainfall–runoff event in GCEW.

at the hillslope scale of resolution, is necessary to simulate streamflows for the purpose of understanding streamflow formation in a diverse range of applied and theoretical research contexts. A few important examples are understanding the physical basis of observed scaling in floods (Gupta et al., 2010; Sharma et al., 2012), understanding the bio-physical basis of scaling in riparian vegetation (Dunn et al., 2011), and predicting soil erosion and water quality from hillslopes during floods under conventional and organic agricultural practices (Lowery et al., 2009). Direct measurements of processes that produce runoff generation are generally unavailable or spatially limited within the drainage area of a study basin. Consequently, runoff generation must be represented using an analytical or a numerical model, and estimates of model variables and parameters must be made using available data. The problem requires a holistic and highly interdisciplinary approach (Wilby, 1997). Two challenges stand out in developing a runoff generation model: (1) representing physical processes that produce runoff from rainfall and (2) representing their variability in space and time throughout a basin. The following overarching question captures these issues and serves as the focus of our paper: how can space–time variable runoff generation in a river basin be modeled at a large number of hillslopes when the finest scale of observed runoff is substantially larger, and the scale of existing infiltration equations and related measurements is much smaller?

We introduce a physical-statistical hypothesis, developed from basin-scale and point-scale observations, that total hillslope water loss for a rainfall-runoff event is inversely related to a function of a lognormal random variable. A top-down statistical model is developed to test our hypothesis and to

1 Introduction

Runoff generation is the net result of separating rainfall into a surface runoff component and a “loss” component that includes infiltration, interception, and evapotranspiration. A spatial representation of runoff generation in a river basin,

provide a theoretical framework for distributing total volume of runoff in space among a large number of hillslopes in a river basin for a rainfall–runoff event. This approach does not require any calibration of model parameters. A key assumption of the model is that the “runoff/loss ratio”, a dimensionless metric that describes the relationship between total event runoff and water loss, has a space–time rescaling property. Because this metric depends solely on water balance, the assumption imposes a constraint on the possible nature and magnitude of processes that govern water loss and runoff generation over time during an event. Yet, the model does not specify the processes themselves. We test the rescaling assumption against observations from the 21 km² Goodwin Creek Experimental Watershed (GCEW) in Mississippi, USA, and find that observations support it. We also show that, for a given rainfall event, the magnitudes of total basin runoff from the model are, on the average, equal to those obtained from observations. Our model can be used in both applied and theoretical research contexts mentioned above.

The rest of the paper is organized as follows. In Sect. 2, we provide some background on the characteristics of and relationships between bottom-up and top-down models. In Sect. 3, we introduce the key variable in this study, the runoff/loss ratio ψ , and explain the data and data-processing steps required to obtain estimates of it for GCEW and for the unnested sub-basins within it. We also introduce our physical-statistical hypothesis in this section. In Sect. 4, we present a pattern in data among unnested basins indicating that rescaled probability distributions of ψ are the same among events. We introduce and explain the model using this pattern as motivation. Finally, we test the model against data in Sect. 5, discuss test results and show an application of our model for a selected rainfall–runoff event. We show that the model supports our hypothesis in Sect. 6, and summarize results in Sect. 7.

2 Background

Spatial scales in hydrology extend upward from a point, to a plot, to a hillslope, to a sub-basin and beyond. Conceptually, the point scale is about 0.01 m² (100 cm²) and the other scales increase in succession roughly by a factor of 10³. We define the third scale, hillslope, as the land surface area that drains into a single channel link of a river network (Shreve, 1967); it can also be called a hillslope-link scale (Mantilla and Gupta, 2005). The hillslope scale represents an important transition in land surface form and process. At this scale, surface runoff from a hillslope enters a channel link in a river network. Above this scale, runoff occurs in multiple links draining a sub-basin. The connectivity of links, as a channel network, aggregates runoff and affects its space–time structure. Scale problems have been recognized in hydrology literature for quite some time (Amerman and McGuinness, 1967; Pilgrim, 1983).

All rainfall–runoff models of engineering hydrology confront the challenge of representing runoff generation accurately at multiple spatial scales, from a point to a plot to a hillslope to a basin. The approaches taken to develop such models have led to two types of models: bottom-up (upward) and top-down (downward). Bottom-up models originate from observations at a point or plot scale (Kavvas et al., 2004; Govindaraju et al., 2006), while top-down models originate from observations at the basin scale (Klemes, 1983; Sivapalan et al., 2003). Most approaches taken to model runoff generation as a space–time variable phenomenon within a basin have been bottom-up. Kirkby (1988) has given an insightful review of the early literature devoted to bottom-up hillslope runoff processes and models.

2.1 Two limitations to bottom-up modeling of runoff generation

Bottom-up models of runoff generation in real river basins are developed from observations having a spatial resolution that is much finer than that of the model. Results in Gutmann and Small (2007); Gutmann and Small (2010) indicate that the uncertainty in representing runoff generation in deterministic land surface models can be substantial partly because there is a great disparity, at least 8 orders of magnitude, between the spatial scale of estimated soil hydraulic properties that are derived from soil texture class data (100 cm²) and the spatial scale (resolution) of most land surface models (≥ 1 km²). A similar scale disparity is found in applications of many rainfall–runoff models (Vieux, 2004). Also, observations show that soil moisture and infiltration tend to differ between points as well (e.g. Bell et al., 1980; Achouri and Gifford, 1984). For an area where the mean saturated hydraulic conductivity (K_s) is fixed, numerical studies indicate that the time series of mean infiltration changes as the spatial variability of K_s within the area increases (e.g. Smith and Hebbert, 1979). Arguably, calibration occurs with both bottom-up and top-down models because of a disparity between observational and model scales.

Bottom-up models of runoff generation also tend to be founded on theoretical formulations of processes at the point scale, while they are run at scales that are much larger. Thus, the spatial resolution of the governing equations is much finer than that of the model. For example, point-scale equations for infiltration are derived from a combination of physical principles under certain idealized conditions and laboratory experimental results, e.g. Philip’s infiltration equation (Philip, 1969). Yet, most bottom-up models assume that these equations extend to the plot scale or larger (e.g. Peck et al., 1977; Merz and Bardossy, 1998), and thus point-scale parameters are replaced by “effective parameters”. The existing literature suggests that such an assumption is not well founded for infiltration under natural conditions. For example, an “effective” hydraulic conductivity for the plot scale, or larger, is not easy to determine because hydraulic conductivity can change

between points within a plot by many orders of magnitude (e.g. Nielsen et al., 1973; Carvallo et al., 1976). Similarly, analytical studies have shown that the functional form of point- and hillslope-scale infiltration equations are different (Maller and Sharma, 1981; Chen et al., 1994; Govindaraju et al., 2006). Kavvas et al. (2004) consider these issues in formulating the WHEY model.

2.2 Top-down modeling of runoff generation

Top-down models have limitations that are similar but opposite of those in bottom-up models, yet they also have advantages that can serve to complement bottom-up modeling results. In particular, the spatial resolution of observations and governing equations on which a top-down model is based is much coarser than the spatial resolution of the model itself. A great advantage of a top-down model is that the model equations represent the collective effect of processes that produce runoff generation. Thus, a top-down model groups together the influences of vegetation, soil type, surface topography, etc., on runoff, while a bottom-up model must treat these influences separately.

There are a relatively few studies on runoff generation that may be considered top-down. Clark and Hebbert (1971) used the phi index model to illustrate that spatially variable infiltration within a basin impacts basin-mean infiltration and its relationship to basin-mean rainfall intensity. Gargouri-Ellouze and Bargaoui (2009) used the phi index to identify the primary physical factors that influence runoff generation among 22 basins; two important factors were found to be maximum rainfall intensity and percent of forest cover. Lan-Anh and Willems (2011) developed a top-down rainfall-runoff model that requires calibration and treats runoff generation as a spatial mean process. A top-down perspective has been taken in efforts to interpolate runoff in space, though not in tandem with the components that lead to its generation (i.e. rainfall and water loss). For example, Sauquet et al. (2000) and Gottschalk et al. (2006) developed a stochastic interpolation method to distribute runoff in space, among pixels in a basin, using observed runoff in nested sub-basins.

3 Estimating runoff/loss ratios for unnested basins in GCEW

Our idea is to define a basin-wide event-based runoff generation metric for GCEW that can be estimated from observations and then to distribute random values of the metric down to the scale of hillslopes that cover GCEW. Our strategy for distributing the metric down to smaller scales is to honor water balance such that, on average, total event runoff for the basin equals the summation of total event runoff at the smaller scales. Down-scaling models of precipitation capture this idea (e.g. Over and Gupta, 1996; Perica and Fofoula-Georgiou, 1996) as explained later in Sect. 4.2.

The basin-wide metric in this paper, ψ , depends on the depth of total event runoff, $\tilde{q} = r - l$, where r is total event rainfall and l is total event water loss. For a given rainfall event over a basin of area a ,

$$\psi = \frac{\tilde{q}}{l} = \frac{r}{l} - 1 \quad ; \quad r = \int_{s_1}^{s_2} r(t) dt,$$

$$l = 1/a \int_{s_1}^{s_2} (r(t)a - \tilde{q}(t)) dt, \tag{1}$$

where t is time, (s_1, s_2) is a time period of duration $s_2 - s_1$ over which rainfall-runoff occurs such that post-event streamflow conditions return to pre-event conditions, $r(t)$ is average rainfall rate over the basin, and $\tilde{q}(t)$ is stream discharge at the outlet of the basin derived solely from event rainfall. Water balance is the basis for l , and thus it does not require an assumption about the processes that affect water loss during a rainfall event. Rather, l represents the collective effects that soil and vegetation (type and spatial distribution) and rainfall (rate, duration, and spatial distribution) have on water that contributes to runoff and water that does not. The first equality in Eq. (1) obeys mass conservation, and shows that $\psi \geq 0$ given that total water loss l cannot exceed total rainfall r . We refer to ψ as the runoff/loss ratio because it equals the ratio of total runoff, $\tilde{q} = r - l$, to total loss, l .

GCEW has a drainage area of 21.19 km², and the rainfall-runoff data for the basin come from 13 stream gauges and 31 rain gauges (Blackmarr and the Channel and Watershed Processes Research Unit, 1995). For this study, we excluded another stream gauge in the basin (gauge 10) because of problems with its data. Figure 1 shows three maps of GCEW that represent basin, unnested basin, and hillslope scales. Map A shows only the drainage area of GCEW, defined here as the upstream area of stream gauge 1. Map B shows the drainage areas of the 13 stream gauges, which can overlap (e.g. gauges 4 and 7). Map C shows the drainage areas of 544 hillslopes that compose GCEW. Rainfall and streamflow data were used to evaluate ψ for the entire GCEW area (see Map A) and for each unnested sub-basin of GCEW (see Map B). Values of ψ were determined for numerous events, and the methods for selecting events, estimating total event rainfall and streamflow, and determining ψ are given below.

3.1 Selecting rainfall-runoff events from streamflow data

We selected rainfall-runoff events having moderately simple rainfall and streamflow conditions. Consecutive 2-day periods from 1981 to 1995 were identified where records indicate that (1) rainfall occurred on the first day of the 2-day period at all rainfall stations, (2) streamflow was present in all gauged channels on both days, and (3) streamflow at gauge 1, the outlet of GCEW, had a single distinct peak. These 2-day periods capture either an entire rainfall event or a subset of one

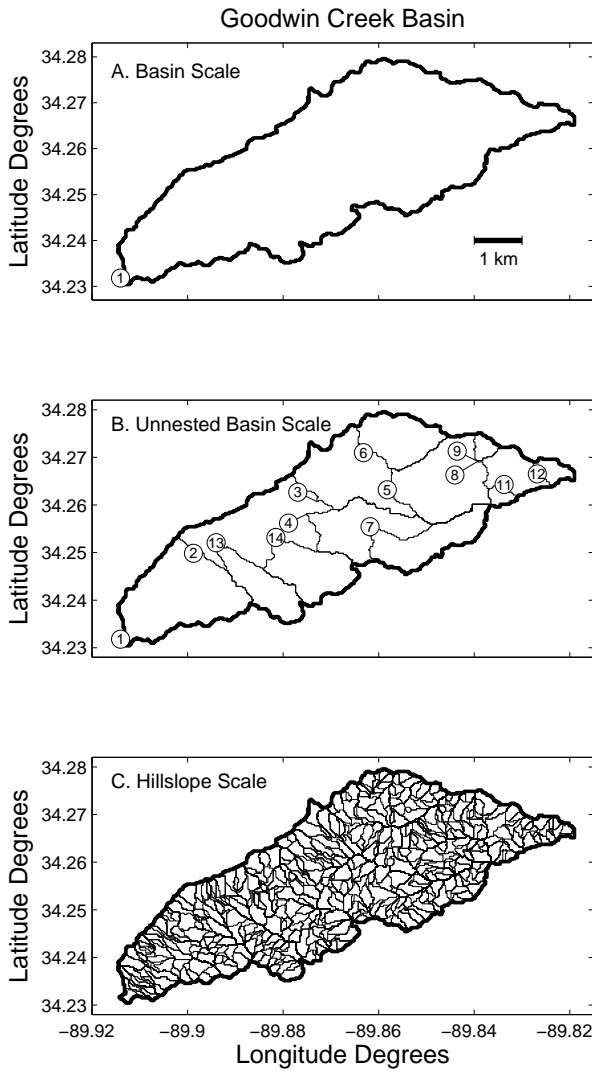


Fig. 1. Decomposition of GCEW into three spatial scales. Plot A shows the basin scale ($a = 21.19 \text{ km}^2$) along with the location of stream gauge 1, the outlet of GCEW. Plot B shows the unnested basin scale ($a/n = 1.63 \text{ km}^2$) and the locations of 13 stream gauges. Plot C shows the hillslope scale ($a/m = 0.04 \text{ km}^2$).

that lasts more than two days. In the latter case, it is possible that rainfall contributing to the outlet peak occurs prior to the 2-day period. Therefore, we used the time of peak streamflow at gauge 1, t_i , as a reference time for an event i . We selected only those 2-day periods where (4) rainfall did not occur at any of the stations from 48 to 24 h before t_i , and (5) rainfall did occur at all rainfall stations during the 24 h period leading up to t_i . A computer algorithm found 148 events that meet criteria (1) to (5), about 10 events per year. For each event, step (4) requires that antecedent conditions include a 24 h period of zero rainfall. This step provides little constraint on antecedent soil moisture, which can be near zero in late summer and near 1 in winter (Furey and Gupta, 2005).

3.2 Estimating rainfall and runoff for events

We generated rainfall fields using data from the 31 rain stations in GCEW. Rain station data for the basin record accumulated rainfall depth per event, and the time interval between data points is not a constant. Therefore, we first produced a new 5-minute-interval time series of rainfall rates for each station by expressing accumulated rainfall values in terms of rain rate and then linearly interpolating between data points. We set interpolated rain rates that were negative to zero. Then, for each event, we generated a time series of rainfall fields beginning at $t_i - d$ and ending at $t_i + d$, where d is a 24 h time period. At each 5-minute time step, we produced a rainfall field by triangulating rain rates at the stations and then linearly interpolating between values; other more-involved approaches could be used that characterize rainfall fields better when rainfall spatial variability is high. For each event in this study, there are 576 five-minute fields ($2 \text{ days} \times 24 \text{ h day}^{-1} \times 12 \text{ five-minute fields h}^{-1}$) that describe the space-time structure of rainfall, including 5-minute periods where there is no rain. Each field covers GCEW and can be partitioned into areas corresponding to the sub-basins (unnested or otherwise) and hillslopes in the basin.

Using stream gauge data, we estimated the runoff generated during an event i from each basin j by subtracting baseflow as

$$\tilde{q}_{i,j}(t) = q_{i,j}(t) - \underline{q}_{i,j}; s_1 \leq t \leq s_2, j = 1, 2, \dots, 13 \quad (2)$$

where

$$\begin{aligned} \underline{q}_{i,j} &= \min_t(q_{i,j}(t)); (t_i - d) \leq t \leq t_i \\ s_1 &= \max(t) \text{ where } q_{i,j}(t) = \underline{q}_{i,j}; (t_i - d) \leq t \leq t_i \\ s_2 &= \min(t) \text{ where } q_{i,j}(t) = \underline{q}_{i,j}; t_i < t. \end{aligned} \quad (3)$$

Here, $q_{i,j}(t)$ represents observed streamflow and $\underline{q}_{i,j}$ represents antecedent streamflow. We interpret $\underline{q}_{i,j}$ as baseflow that is constant during an event and subtract it from $q_{i,j}(t)$ to obtain event runoff, $\tilde{q}_{i,j}(t)$.

3.3 Determining ψ

We estimated runoff/loss ratios at the basin scale, ψ_i , and unnested basin scale, $\psi_{i,j}$ where $j = 1, 2, \dots, 13$ denotes an unnested basin. For the basin scale (Fig. 1, Map A), we used Eq. (1) where the variables for runoff/loss ratio, total rainfall, and streamflow are each given a subscript i . We averaged rainfall over the drainage area of stream gauge 1 at each time step to obtain rainfall rate, denoted $r_i(t)$, and used Eq. (2) to estimate streamflow, denoted $\tilde{q}_i(t)$. For the unnested basin scale (Fig. 1, Map B), we used a general form of Eq. (1) given as

$$\psi_{i,j} = \frac{r_{i,j}}{l_{i,j}} - 1 \quad ; \quad r_{i,j} = \int_{s_1}^{s_2} r_{i,j}(t) dt,$$

$$l_{i,j} = 1/a_j \int_{s_1}^{s_2} \left(r_{i,j}(t)a_j + \sum_{k \in B_j} \tilde{q}_{i,k}(t) - \tilde{q}_{i,j}(t) \right) dt, \quad (4)$$

where B_j is a set of gauged sub-basins of basin j . If B_j is an empty set, then $\sum_{k \in B_j} \tilde{q}_{i,k}(t) = 0$.

For the unnested basin scale, we unnested data from a gauged basin j using data from a set B_j of gauged sub-basins. Table 1 lists the 13 stream-gauged basins in GCEW ($j = 1, 2, \dots, 13$) and shows sub-basin gauges in set B_j , if any, and drainage areas before and after unnesting. We obtained the unnested drainage area for basin j by subtracting the areas of the sub-basin gauges in B_j from the area of basin j ; e.g. for basin $j = 2$, the unnested drainage area is $a_2 = 17.92 - (8.78 + 3.57 + 1.24 + 1.63) = 2.7 \text{ km}^2$. We used CUENCAS (Mantilla and Gupta, 2005) to determine an unnested rainfall time series for event i and basin j , denoted $r_{i,j}(t)$, by averaging rainfall over the unnested area of j at each time step. Finally, we evaluated the unnested total runoff for event i and basin j as the difference between total runoff exiting and total runoff entering the unnested area.

3.4 Selecting a final collection of events

For many of the 148 events, there are unnested basins for which water balance is unresolved. This feature arises in two ways. For certain events, there is at least one unnested basin for which the runoff exiting its area is less than the runoff entering its area at upstream gauges. In Eq. (4), this situation means $\tilde{q}_{i,j}(t) < \sum_{k \in B_j} \tilde{q}_{i,k}(t)$, which suggests that water loss occurs within the river channels of unnested basin j . It is most often found with basin $j = 2$, which coincidentally has the largest set B_j of sub-basin gauges. Also, for some events, there is an unnested basin for which the integrand in Eq. (4) is negative, meaning $\tilde{q}_{i,j}(t) > r_{i,j}(t)a_j + \sum_{k \in B_j} \tilde{q}_{i,k}(t)$. This situation suggests that there is an unknown source of water that contributes to runoff exiting unnested basin j . These two features could have natural or human-induced physical origins, but they also could originate from inaccuracies in streamflow measurements, rainfall estimates, and baseflow $q_{i,j}$ estimates.

We selected events for which water balance is resolved in a minimum of 11 unnested basins. This brings the number of events to 112, still a large number by which to develop and test a top-down model. These events were put into three groups: 112 events with 11 unnested basins (Case I), 75 events with 12 unnested basins (Case II), and 20 events with 13 unnested basins (Case III). Case I events include Case II events which include Case III events. We grouped the data to test our model. As explained in Sect. 4, our model predicts

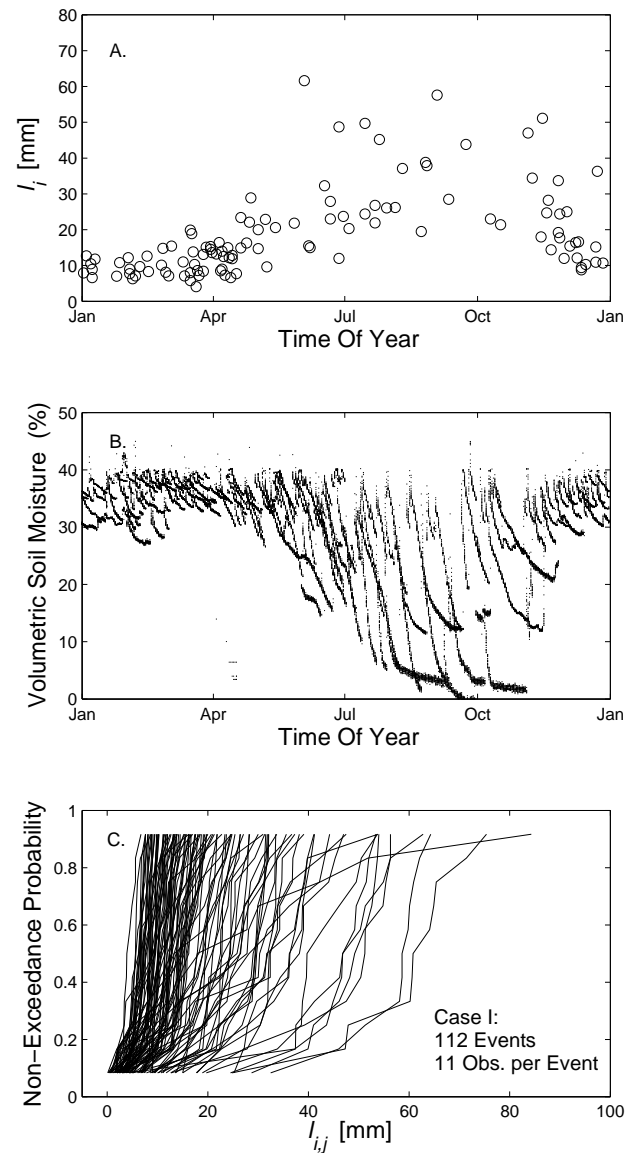


Fig. 2. Plot A shows water-loss depths at the basin scale, l_i , for 112 events versus time of year. Plot B shows volumetric soil moisture content versus time of year from Furey and Gupta (2005). Plot C shows empirical CDFs of water-loss depths at the unnested basin scale, $l_{i,j}$, for 112 events. Each distribution represents a rainfall–runoff event.

similarity between rescaled cumulative distribution functions (CDFs). Statistical tests of this feature can be sensitive to the number of samples that comprise empirical CDFs. Thus, while Case I events include those for Case III, it is possible that differences between the cases in the number of samples per empirical CDF lead to contradictory test results that invalidate the model.

The process we used to select events was designed to provide a large number of events but, as expected, it eliminated many events from analysis. Nonetheless, events for each case

Table 1. Properties of gauges used in this study.

ARS gauge ID	Basin ID j	Drainage area [km ²]	Sub-basin gauges Set B_j	Unnested drainage area a_j [km ²]
1	1	21.39 (<i>a</i>)	2	3.47
2	2	17.92	3, 4, 13, 14	2.70
3	3	8.78	5, 6	3.29
4	4	3.57	7	1.97
5	5	4.30	8, 9	2.57
6	6	1.19	–	1.19
7	7	1.60	–	1.60
8	8	1.55	11, 12	0.97
9	9	0.18	–	0.18
11	10	0.28	–	0.28
12	11	0.30	–	0.30
13	12	1.24	–	1.24
14	13	1.63	–	1.63

span a broad range of streamflow conditions. Among events, the smallest streamflow peaks at the outlet of GCEW are 0.21, 0.25, and 0.40 m³ s⁻¹ for Cases I to III, respectively. The largest streamflow peaks at the outlet of GCEW are 102, 47.8, and 43.7 m³ s⁻¹ for Cases I to III, respectively. The small events have return periods of less than 1 yr, while the large events have return periods of approximately 2 yr for Cases II and III and 5 yr for Case I. Both the event selection process and the period of time we investigated, 15 yr, are responsible for the lack of events having return periods that exceed 5 yr.

3.5 Soil moisture in GCEW and a physical-statistical hypothesis to test against the model

Before developing our model, we compared observations of water-loss depths at the basin scale, l_i , to historical observations of soil moisture in GCEW at the point scale. Figure 2a presents values of l_i versus time of year for the Case I events. Values are low from December to March and high from July to October. The peak water-loss depth for GCEW occurs around late August. Figure 2b comes from Furey and Gupta (2005) and shows observed values of volumetric soil moisture content taken 5 cm below soil surface at two point locations in and near GCEW from January 1999 to June 2004. The general relationship between l_i and time is a mirror image of the general relationship between soil moisture and time.

Spatial distributions of water-loss depth in GCEW also appear to depend on soil moisture. Figure 2c shows the empirical CDFs of water-loss depth at the unnested basin scale, $l_{i,j}$, for the 112 events. Each CDF consists of 11 values of water-loss depth corresponding to 11 unnested basins. The figure shows that distributions that have a large mean tend to have a large variance and vice versa. Because a large $l_{i,j}$ implies that soil moisture is low (dry), it suggests that the

spatial variability of soil moisture in GCEW is greatest when the spatial-mean soil moisture is low. This feature is consistent with soil moisture observations in other humid climate basins (Brocca et al., 2007). The patterns in Fig. 2 show that soil moisture plays an important role in runoff and loss in GCEW and should be a component of our model.

The observations presented in Fig. 2 also lead to a hypothesis, as follows. Plots A and B in Fig. 2 suggest that l_i is roughly proportional to some function of $1/\theta$, where θ denotes point-scale soil moisture. It also suggests that l_i characterizes surface infiltration amount as a first-order effect in GCEW. Based on the observations, we infer that $l_{i,j,k}$ for any hillslope within an unnested sub-basin is also proportional to some function of $1/\theta$; there are no data sets available to test this idea directly. Hydraulic conductivity $K(\theta)$ is observed to be proportional to a function of θ because, as soil moisture decreases, conductivity decreases (Brutsaert, 2005, p. 274–275, Figs. 8.23 and 8.25). A widely used parametric relationship for conductivity is $K(\theta) = K_s f(\theta)$, where K_s is hydraulic conductivity at saturation, and f denotes a function of θ (Brutsaert, 2005, p. 279–280). Spatially variable K_s is commonly modeled as a lognormal random variable (e.g. Govindaraju et al., 2006; Meng et al., 2006), which implies that $K(\theta)$ is lognormal within a hillslope. These relationships serve as the basis for a physical-statistical hypothesis: hillslope water loss is inversely related to a function of a lognormal random variable. We show that our model supports this hypothesis at the end of Sect. 6.

4 Runoff–loss model

Our goal is to develop a runoff–loss model that describes the variability of ψ temporally, from event to event, and spatially, both within a given scale, e.g. among unnested basins, and between scales, e.g. from basin to unnested basin to hillslope.

Such a model provides a method of using observations of ψ at the basin and unnested basin scales to estimate runoff generation and losses at the hillslope scale, where practically no observations are available. Our first step in developing the model was to explore the temporal and spatial variability of runoff/loss ratios at the unnested basin scale, $\psi_{i,j}$, for Case I events. We wanted to determine if a simple rescaling of $\psi_{i,j}$ with respect to a measure of an overall event magnitude could account for event-to-event changes.

4.1 Establishing model form – temporal structure

Figure 3a shows the empirical CDFs of $\psi_{i,j}$ for Case I events. There are 112 CDFs that correspond to 112 events, and each CDF consists of 11 unnested basins. For each event i , we evaluated the geometric mean of observed runoff/loss ratios, g_i , and then rescaled the observed ratios as $\psi_{i,j}/g_i$. Figure 3b shows the CDFs of this rescaled runoff/loss ratio. The curves in Fig. 3b collapse nicely, indicating that temporal variability can, to a first order, be accounted for by the use of the geometric mean as a simple scale parameter. An alternative approach is to rescale $\psi_{i,j}$ by the arithmetic mean, but test results (not shown here) indicate that a weaker collapse occurs. Rescaling $\psi_{i,j}$ by its geometric mean is equivalent to shifting $\log(\psi_{i,j})$ by its arithmetic mean, i.e. computing $\log(\psi_{i,j}) - \log(g_i)$. Thus, the results in Fig. 3 indicate that an event-by-event first-moment shift yields similar distributions.

We took a step further to see if there is improvement when incorporating second moments (variances). If the event-by-event variances are all the same, then the first-moment shift will suffice. Otherwise, we must take variances into consideration. We incorporated a second-moment correction by computing $(\log(\psi_{i,j}) - \log(g_i))/\hat{\sigma}_i$, where $\hat{\sigma}_i$ is the standard deviation of $\log(\psi_{i,j})$. Figure 3c shows the CDFs of the exponential of this quantity, which is $(\psi_{i,j}/g_i)^{1/\hat{\sigma}_i}$. The collapse improves compared to that in Fig. 3b, indicating that rescaling should account for temporal differences in variances. The implication of the results presented in Fig. 3 is that event-to-event temporal variability in the spatial distribution of $\psi_{i,j}$ can be characterized quite well by the use of two parameters, the mean and variance of $\log(\psi_{i,j})$. The results in Fig. 3 serve as the basis for developing our model.

4.2 Establishing model form – spatial structure

The model we develop needs to capture the features presented in Fig. 3 but also connect runoff/loss ratios across scales in a way that respects their nested spatial correlation structure. At the unnested basin scale, we expect a high degree of spatial correlation among values of $\psi_{i,j}$ for event i because they will be influenced by overall event magnitude. If we compare the spatial distribution of $\psi_{i,j}$ for a large event to one for a small event, e.g. large vs. small rainfall amounts, then we expect that the two distributions will be distinct.

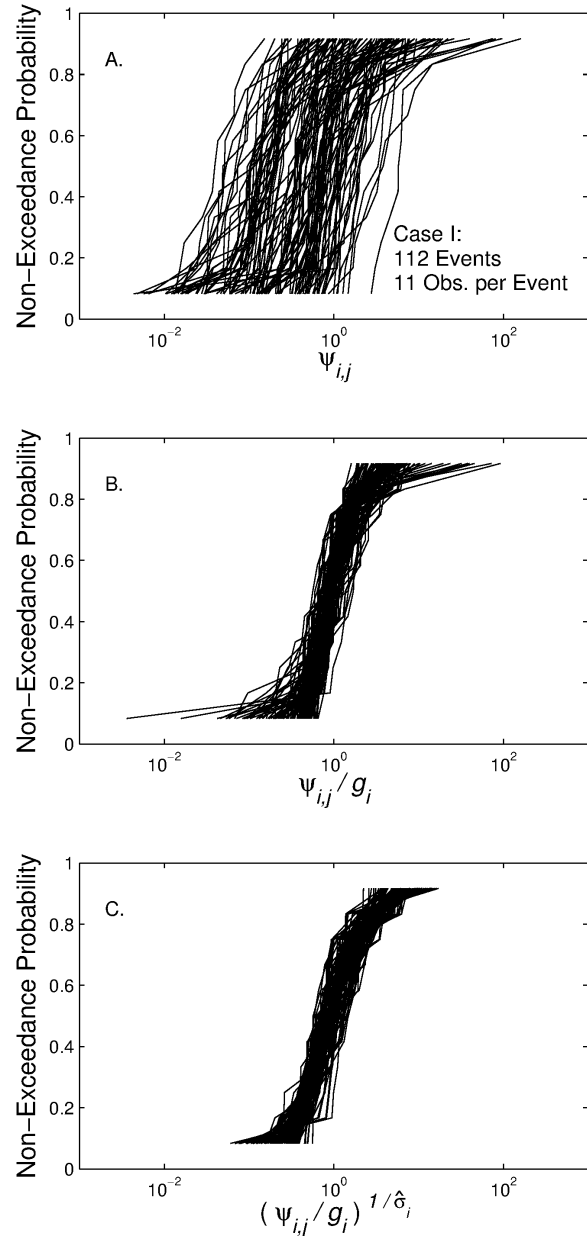


Fig. 3. Plot A shows a semi-log plot of empirical CDFs of runoff/loss ratios at the unnested basin scale, $\psi_{i,j}$, for 112 events. Each distribution represents a rainfall–runoff event. Plot B shows corresponding CDFs of $\psi_{i,j}/g_i$, and Plot C shows corresponding CDFs of $(\psi_{i,j}/g_i)^{1/\hat{\sigma}_i}$.

This effect can be modeled by taking $\psi_{i,j}$ to be a function of ψ_i where the value of ψ_i reflects the overall event effect. Likewise, for unnested basin j , we can model hillslope runoff/loss ratios to be a function of $\psi_{i,j}$ where $\psi_{i,j}$ reflects the overall event effect.

The nested structure of runoff/loss ratios must also be weighted by the drainage areas with which they are associated. Suppose a basin is partitioned into five un-nested basins, $j = 1, 2, \dots, 5$, where $\psi_{i,1}$ occupies 80 % of the total basin area while the remaining $\psi_{i,j}$ occupy 20 %. The dependency of $\psi_{i,1}$ on the overall event effect, represented by ψ_i , will be stronger than that of the remaining $\psi_{i,j}$ simply because $\psi_{i,1}$ represents a larger fraction of the basin.

Discrete random cascade models (Gupta and Waymire, 1993) have a nested structure with spatial characteristics like those described above. Therefore, we can use the equations that define a cascade model to help establish the equations that define our model. A brief self-contained explanation of random cascades is given below that motivates the functional form of our model.

Consider a discrete random cascade spatial model with a branching number b ; i.e. a region of area $a_0 = 1$ is divided into b subregions of equal size (called level-1 subregions), each of these subregions is subdivided into b subregions of equal size (called level-2 subregions), and so on. Denote total mass in the whole region by M_0 , and denote mass in the j th level-1 subregion by $M_{1,j}$, $j = 1, \dots, b$. Then, under the random cascade model, $M_{1,j}$ and M_0 are related by

$$M_{1,j} \stackrel{d}{=} (1/b)W_{1,j}M_0, \quad (5)$$

where $\stackrel{d}{=}$ means “equality in distribution”, and $W_{1,j}$ are independent and identically distributed (iid) level-1 cascade generators with expectation 1. To understand the implications of this equation in terms of mean mass conservation, we can take the expectation of Eq. (5) to obtain

$$E[M_{1,j}] = E[M_0]/b. \quad (6)$$

Therefore, $E[M_0] = bE[M_{1,j}]$, showing that mass conservation holds in a mean sense. The area of a level-1 subregion is $a_{1,j} = 1/b$, and we can take the logarithm of equation (5) to obtain

$$\ln(M_{1,j}) \stackrel{d}{=} \ln(M_0) + \ln(a_{1,j}) + \ln(W_{1,j}). \quad (7)$$

Here, total mass $M_{1,j}$ depends on M_0 , where M_0 reflects the overall mass, and $M_{1,j}$ depends on the fraction of the overall area. These relationships parallel those we want for a model of ψ . An expression like Eq. (7) can also be written for each level- n subregion. For example, it can be modified to a relation between total mass for a level-2 subregion, with area $1/b^2$, and total mass $M_{1,j}$. Finally, notice that Eq. (7) involves a Markov-like property. In a Markov process, a variable at the present time that is conditioned on a set of variables representing past times has the same distribution as a variable that is conditioned on only the most recent past time (Feller, 1968). Random cascades have a Markov-like property because a given spatial scale influences only the next finer spatial scale (Over, 1995).

We can model ψ like a random cascade, though not exactly as one. A random cascade model leads to a limit mass distribution after a large number of subdivisions of space. It describes a process, e.g. rainfall, at a large number of spatial scales. By contrast, the smallest spatial scale for a model of ψ is the hillslope; physical reasons are given in Gupta et al. (2010). Also, cascades are random spatial measures that obey mass conservation in a mean sense, as shown by Eq. (6), but ψ does not obey this condition. Rather, it is the components of ψ (i.e. total rainfall and total loss) that obey mass conservation, as shown in Sect. 4.7.

In the sections that follow, we develop a runoff–loss model based on two equations that are each similar in form to Eq. (7), but are more general. Each equation characterizes how ψ at one spatial scale depends on both ψ at the next coarser spatial scale and drainage area fraction. We introduce the model in terms of random and then fixed un-nested basins and hillslopes. Before testing the model, we also discuss how it connects to water balance.

4.3 Definition of model variables

We develop a model under the postulate that certain properties of ψ observed at the basin scale are preserved down to the hillslope scale. Consider a basin of area a and divide it into n un-nested basins of similar size and m hillslopes of similar size where $m > n$. We let m_j denote the number of hillslopes in an un-nested basin j such that $m = \sum_{j=1}^n m_j$. We let a_j denote the area of un-nested basin j where $j = 1, 2, \dots, n$ and $a_{j,k}$ denote the area of hillslope k in un-nested basin j where $k = 1, 2, \dots, m_j$. It follows that $a = \sum_{j=1}^n a_j = \sum_{j=1}^n \sum_{k=1}^{m_j} a_{j,k}$. Finally, we let a/n define the scale associated with the un-nested basins and a/m define the scale associated with the hillslopes. Thus, we consider three different spatial scales: basin scale a , un-nested basin scale a/n where $n > 1$, and hillslope scale a/m where $m > n$. Figure 1 shows the decomposition of GCEW into these scales.

Random variables R_i and L_i express basin scale conditions in the model; see Table 2. They represent total rainfall and total water loss for event i in a basin, respectively. Realizations are denoted by r_i and l_i . A given event is not associated with a particular rainfall amount unless r_i is specified. Rather, a given event simply represents a distinct period of time where rainfall from a single storm produces runoff from a basin and streamflow conditions at the end of the time period are equal to those at the beginning.

Random variables J , a_J , $R_{i,J}$, and $L_{i,J}$ express spatial variability at the un-nested basin scale; see Table 2. Random variable J represents the stream-gauged un-nested basins in GCEW so that a realization, j , represents a specific un-nested basin. Random variable a_J represents the drainage areas of un-nested basins so that a realization a_j represents the area of a specific un-nested basin. Random variables $R_{i,J}$ and $L_{i,J}$ represent total event rainfall and total water loss for event i

Table 2. Variables that represent random and fixed locations (in space) for area a , total rainfall R , total loss L , and runoff/loss ratio Y . Realizations of R , L , and Y (not shown) are obtained using a lowercase letter; e.g. a realization of $R_{i,j}$ is given by $r_{i,j}$.

Scale	Location	Area	Total rainfall	Total loss	Runoff/loss ratio
Basin	Fixed	a	R_i	L_i	Y_i
Unnested basin	Random	a_J	$R_{i,J}$	$L_{i,J}$	$Y_{i,J}$
	Fixed	a_j	$R_{i,j}$	$L_{i,j}$	$Y_{i,j}$
Hillslope	Random	$a_{j,K}$	$R_{i,j,K}$	$L_{i,j,K}$	$Y_{i,j,K}$
	Fixed	$a_{j,k}$	$R_{i,j,k}$	$L_{i,j,k}$	$Y_{i,j,k}$

at the unnested basin scale. Realizations $r_{i,J}$ and $l_{i,J}$ refer to an unnested basin with an outlet at a random location, while realizations $r_{i,j}$ and $l_{i,j}$ pertain to a specific unnested basin, j .

Random variables K , $a_{j,K}$, $R_{i,j,K}$, and $L_{i,j,K}$ represent spatial variability at the hillslope scale, and their meanings parallel those for the unnested basin scale; see Table 2. Thus, K represents the hillslopes in an unnested basin so that a realization, k , represents a specific hillslope. Random variable $a_{j,K}$ represents the drainage areas of hillslopes in basin j , and a realization $a_{j,k}$ represents the drainage area of a specific hillslope in unnested basin j . Lastly, random variables $R_{i,j,K}$ and $L_{i,j,K}$ represent total event rainfall and total water loss for event i in unnested basin j at the hillslope scale. Realizations $r_{i,j,K}$ and $l_{i,j,K}$ refer to a hillslope at a random location in unnested basin j , and realizations $r_{i,j,k}$ and $l_{i,j,k}$ refer to a hillslope at a given location in unnested basin j .

Table 2 provides a comparison between the notations used to represent the basin, unnested basin, and hillslope scales. The connection between variables across scales has been made systematic. The first index corresponds to an event (time), the second index corresponds to the unnested basin scale (space), and the third index corresponds to the hillslope scale (space). Thus, the number of indices attached to a random variable defines its spatial scale. One index refers to an event i at the basin scale, e.g. L_i ; two indices refers to an event i at the unnested basin scale J , e.g. $L_{i,J}$; and three indices refers to an event i for a fixed unnested basin j at the hillslope scale K , e.g. $L_{i,j,K}$.

4.4 Model equations for random unnested basins and hillslopes

The model equations introduced below represent a generalization of the random cascade model given in Eq. (7). The nature of the generalization is explained after the model is introduced. Before introducing the equations, we make some assumptions about unnested basin and hillslope distributions. Assume that J has a uniform distribution on the integers $1, \dots, n$, such that all unnested basins have the same probability of being selected at random. This definition is consistent with our analysis of empirical CDFs in Sect. 4.1. Likewise, assume that $K|J = j$ has a uniform distribution on the integers $1, \dots, m_j$, such that all hillslopes in unnested basin j

have the same probability of being selected at random. Also, assume that J and K are independent of all random variables R and L .

Let $Y_i = R_i/L_i - 1$ be a random variable for the runoff/loss ratio at the basin scale. Similarly, let $Y_{i,J} = R_{i,J}/L_{i,J} - 1$ and $Y_{i,j,K} = R_{i,j,K}/L_{i,j,K} - 1$ be random variables at the unnested basin and hillslope scales. For a random variable X , let $G[X]$ denote the geometric mean of X . Then, assume that for a given event i

$$\ln(Y_{i,J}) = \alpha_i + \beta_i \ln(Y_i) + \gamma_i \ln(a_J/G[a_J]) + \sigma_i \ln(U_{i,J}) \tag{8a}$$

$$\ln(Y_{i,j,K}) = \alpha_{i,j} + \beta_{i,j} \ln(Y_{i,j}) + \gamma_{i,j} \ln(a_{j,K}/G[a_{j,K}]) + \sigma_{i,j} \ln(H_{i,j,k}). \tag{8b}$$

Equation (8a) describes the variability in runoff/loss ratios between unnested basins for event i , while Eq. (8b) describes the variability in runoff/loss ratios between hillslopes in unnested basin j for event i . Here, in Eq. (8a), $G[a_J]$ is a geometric mean of unnested basin areas, $\ln(U_{i,J})$ is a random variable that has a mean of zero and is iid among events, Y_i is independent of $\ln(U_{i,J})$, and parameters α_i , β_i , γ_i , and σ_i can change from event to event because they depend on i . The iid condition on $\ln(U_{i,J})$ means that a common probability distribution connects events. The subscript i in $\ln(U_{i,J})$ is needed because realizations of $\ln(U_{i,J}|J = j)$ for one event are not necessarily the same as those for another. Similarly, in Eq. (8b), $G[a_{j,K}]$ is a geometric mean of hillslope areas in unnested basin j , $\ln(H_{i,j,K})$ is a random variable that has a mean of zero and is iid among events, $Y_{i,j}$ is independent of $\ln(H_{i,j,K})$, and parameters $\alpha_{i,j}$, $\beta_{i,j}$, $\gamma_{i,j}$, and $\sigma_{i,j}$ can change from event to event because they depend on i .

Equations (8a) and (8b) are general expressions of Eq. (7). Equation (8a) includes a random term $U_{i,j}$ that corresponds to the cascade generator $W_{1,j}$. It also includes additional parameters, α_i , β_i , γ_i , and σ_i , and an additional subscript, i , which accounts for changes from event to event. If values of α_i , β_i , γ_i and σ_i are 0, 1, 1, and 1, then the resulting equation looks like Eq. (7). However, values of these parameters are not specified a priori in the equation. Equation (8b) is also a general expression of Eq. (7) but only when it is modified to a relation between a level-2 subregion and a level-1 subregion. Later, in Sect. 4.6, we make an assumption that the distributions of error terms for Eqs. (8a) and (8b) are the same, which

is essentially a scale invariance assumption. This assumption is found in random cascades where $W_{n,j}$ is the same distribution for all n . Thus, taken together, Eqs. (8a) and (8b) are analogous to a “two-level” discrete cascade formulation where intermediate scales and scales finer than the hillslope scale are not explored. By including the event subscript i , Eqs. (8a) and (8b) comprise a space–time model rather than simply a spatial model. Over and Gupta (1996) considered a similar space–time extension of the random cascade model for space–time rainfall.

We next examine the physical implications of Eq. (8a) and then show how it explains the observations presented in Sect. 4.1. Equation (8a) characterizes runoff/loss ratios at the unnested basin scale. It captures two features that should be upheld, as described in Sect. 4.2. Namely, the equation shows that $\ln(Y_{i,J})$ depends on the runoff/loss ratio at the next larger spatial scale, Y_i , and only at this scale, a Markov-like property. It also depends on unnested basin drainage area, $a_J/G[a_J]$. In the equation, the distributions of a_J and $U_{i,J}$ do not change between events, and $U_{i,J}$ accounts for both spatial variability and “fixed event” temporal variability. The equation represents spatial variability through differences in drainage area, via a_J , and through differences in physical conditions that affect water loss and runoff, via $U_{i,J}$. For this paper, we interpret $U_{i,J}$ to represent the influence of antecedent soil moisture conditions for event i . Temporal variability, as expressed by $U_{i,J}$, is represented as an ensemble where realizations of $U_{i,J}$ change under repeated occurrences of the same event.

Equation (8a) helps to explain the observations presented in Fig. 3 and discussed in Sect. 4.1. Taking the expectation of Eq. (8a), conditioned on $Y_i = y_i$, gives

$$E[\ln(Y_{i,J})|Y_i = y_i] = \alpha_i + \beta_i \ln(y_i) + \gamma_i (E[\ln(a_J)] - \ln(G[a_J])).$$

Because $\ln(G[X]) = E[\ln(X)]$ for any random variable X , the expression above can be rewritten as

$$\ln(G_i(y_i)) = \alpha_i + \beta_i \ln(y_i) \quad ; \quad G_i(y_i) = G[Y_{i,J}|Y_i = y_i].$$

Here, $E[\ln(a_J)] - \ln(G[a_J]) = 0$ because $\ln(G[a_J]) = E[\ln(a_J)]$, and $G_i(y_i)$ corresponds to the sample estimates of the geometric mean used to produce the CDFs in Fig. 3. Taking the exponential of both $(\ln(Y_{i,J})|Y_i = y_i)$, from Eq. (8a), and $\ln(G_i(y_i))$, from the expression above, leads to a rescaled runoff/loss ratio given as

$$\frac{Y_{i,J}|Y_i = y_i}{G_i(y_i)} = \left(\frac{a_J}{G[a_J]}\right)^{\gamma_i} U_{i,J}^{\sigma_i},$$

or

$$\left(\frac{Y_{i,J}|Y_i = y_i}{G_i(y_i)}\right)^{1/\sigma_i} = \left(\frac{a_J}{G[a_J]}\right)^{\gamma_i/\sigma_i} U_{i,J}. \tag{9}$$

Equation (9), in conjunction with our discussion in Sect. 4.1, indicates that distributions of $\ln(\psi_{i,j}/g_i)$ can be

normal for each event but dissimilar between events because of event-to-event changes in γ_i and/or σ_i ; $U_{i,j}$ cannot be a source of dissimilarity because it is iid among events. In Sect. 4.1, we showed that CDFs of $\psi_{i,j}$ collapse to a common distribution when rescaling accounts for event-to-event changes in both $G_i(y_i)$ and σ_i (Fig. 3c). Based on Eq. (9), this result occurs when $\left(\frac{a_J}{G[a_J]}\right)^{\gamma_i/\sigma_i}$ is constant among events because $\frac{a_J}{G[a_J]} = 1$, $\gamma_i = 0$, or $\gamma_i/\sigma_i > 0$ is constant among events. In any case, only changes in $G_i(y_i)$ and σ_i are important for relating CDFs of $\psi_{i,j}$ to a common distribution.

4.5 Model equations for fixed unnested basins and hillslopes

Our objective is to develop an expression for the runoff/loss ratio at the hillslope scale that provides a method of simulating event-based total runoff volume for each hillslope draining an unnested basin. To meet this objective, we first need to describe $\ln(Y_{i,J})$ for a given unnested basin j and $\ln(Y_{i,j,K})$ for a given hillslope k within unnested basin j . It is being assumed that J and K are independent of all random variables Y . Also, subscript j for Y means that we are conditioning on unnested basin j , and a subscript k for Y means that we are conditioning on hillslope k . Consequently, $(Y_{i,j}|Y_i = y_i)$ is equivalent to $(Y_{i,J}|Y_i = y_i, J = j)$, $(Y_{i,j,k}|Y_{i,j} = y_{i,j})$ is equivalent to $(Y_{i,j,K}|Y_{i,j} = y_{i,j}, K = k)$, and Eqs. (8a) and (8b) provide the description we need when expressed conditionally as $(\ln(Y_{i,j})|Y_i = y_i)$ and $(\ln(Y_{i,j,k})|Y_{i,j} = y_{i,j})$. In these equations, unnested basin and hillslope drainage areas, respectively, are fixed. Because of this feature, we denote

$$\begin{aligned} a'_j &= a_j/G[a_J] \\ a'_{j,k} &= a_{j,k}/G[a_{j,k}] \end{aligned} \tag{10}$$

as the area terms in the equations.

The expression obtained from Eq. (8a) for $(\ln(Y_{i,j})|Y_i = y_i)$ describes the variability in runoff/loss ratios for a given unnested basin j and event i . The area term, a'_j , expresses the relative contribution that the runoff/loss ratio for the unnested basin makes to the basin-wide value $\ln(y_i)$. If $a'_j = 1$, then $(\ln(Y_{i,j})|Y_i = y_i)$ is independent of area. This situation can occur if all unnested basins have the same area, regardless of size. However, gauged unnested basins rarely, if ever, have the same unnested areas, including GCEW. When the unnested basin is given as j , $U_{i,J}$ becomes $U_{i,j}$. The distribution of $U_{i,j}$ does not change between events, and, for a given event i , $U_{i,j}$ accounts for both spatial variability and fixed event temporal variability. The spatial variability described by $U_{i,j}$ represents an ensemble where realizations of $U_{i,j}$ change under repeated copies of unnested basin j for the same event. Differences in antecedent soil moisture between copies are postulated to represent this change.

Similarly, the expression obtained from Eq. (8b) for $(\ln(Y_{i,j,k})|Y_{i,j} = y_{i,j})$ describes variability for a given

hillslope k in unnested basin j and for event i . Its area term, $a'_{j,k}$, represents a relative contribution to $\ln(y_{i,j})$. Given a hillslope, $H_{i,j,K}$ becomes $H_{i,j,k}$. The distribution of $H_{i,j,k}$ does not change between events, but $H_{i,j,k}$ accounts for both spatial variability and fixed event temporal variability.

The rescaled runoff/loss ratio for a fixed basin is comparable to but significantly different from Eq. (9). From Eq. (8a), the expected runoff/loss ratio for a fixed basin is

$$E[\ln(Y_{i,j})|Y_i = y_i] = \alpha_i + \beta_i \ln(y_i) + \gamma_i \ln(a'_j),$$

which can be written as

$$\ln(\tilde{G}_i(y_i, j)) = \alpha_i + \beta_i \ln(y_i) + \gamma_i \ln(a'_j);$$

$$\tilde{G}_i(y_i, j) = G[Y_{i,j}|Y_i = y_i],$$

because $\ln(G[X]) = E[\ln(X)]$ for any random variable X , as mentioned previously. Since $\tilde{G}_i(y_i, j)$ is given for a basin j , it does not correspond to the sample estimates of the geometric mean used to produce the CDFs in Fig. 3. Taking the exponential of $(\ln(Y_{i,j})|Y_i = y_i)$, from Eq. (8a), and $\ln(\tilde{G}_i(y_i, j))$ leads to a rescaled runoff/loss ratio for a fixed basin given as

$$\frac{Y_{i,j}|Y_i = y_i}{\tilde{G}_i(y_i, j)} = U_{i,j}^{\sigma_i}$$

or

$$\left(\frac{Y_{i,j}|Y_i = y_i}{\tilde{G}_i(y_i, j)} \right)^{1/\sigma_i} = U_{i,j}. \tag{11}$$

Unlike Eq. (9), Eq. (11) shows that the distribution of $(Y_{i,j}|Y_i = y_i)$, after a rescaling that includes σ_i , equals $U_{i,j}$ and thus is iid among events.

4.6 Assumptions for assigning water loss to hillslopes

To reach our final objective, we need to relate $\alpha_{i,j}$, $\beta_{i,j}$, and $\gamma_{i,j}$ in Eq. (8b) to α_i , β_i , and γ_i in Eq. (8a) because hillslope observations needed to estimate $\alpha_{i,j}$, $\beta_{i,j}$, and $\gamma_{i,j}$ are unavailable at present and are not expected to be available in the foreseeable future. Therefore, we let

$$\bar{\alpha}_i = 1/n \sum_{j=1}^n \alpha_{i,j} \quad v_{\alpha,i} = 1/n \sum_{j=1}^n (\alpha_{i,j} - \bar{\alpha}_i)^2, \tag{12}$$

$$\bar{\beta}_i = 1/n \sum_{j=1}^n \beta_{i,j} \quad v_{\beta,i} = 1/n \sum_{j=1}^n (\beta_{i,j} - \bar{\beta}_i)^2,$$

$$\bar{\gamma}_i = 1/n \sum_{j=1}^n \gamma_{i,j} \quad v_{\gamma,i} = 1/n \sum_{j=1}^n (\gamma_{i,j} - \bar{\gamma}_i)^2,$$

where $\bar{\alpha}_i$ is the average of $\alpha_{i,j}$ over j for event i , and $v_{\alpha,i}$ is the variance of $\alpha_{i,j}$ over j for event i . Likewise, $\bar{\beta}_i$ and $\bar{\gamma}_i$ are averages and $v_{\beta,i}$ and $v_{\gamma,i}$ are variances. We assume that, to a first order, $v_{\alpha,i} = v_{\beta,i} = v_{\gamma,i} = 0$, which implies

that $\bar{\alpha}_i = \alpha_{i,j}$, $\bar{\beta}_i = \beta_{i,j}$, and $\bar{\gamma}_i = \gamma_{i,j}$. We also assume that $\bar{\alpha}_i = \alpha_i$, $\bar{\beta}_i = \beta_i$, and $\bar{\gamma}_i = \gamma_i$, which means that $\alpha_{i,j} = \alpha_i$, $\beta_{i,j} = \beta_i$, and $\gamma_{i,j} = \gamma_i$. We later assess the impact of these assumptions by examining model results in the context of water balance. Substituting these last three relationships into Eq. (8b), conditioning on $Y_{i,j} = y_{i,j}$ and $K = k$, and exponentiating the result gives

$$(Y_{i,j,k}|Y_{i,j} = y_{i,j}) = \exp(\alpha_i) y_{i,j}^{\beta_i} a'_{j,k}{}^{\gamma_i} H_{i,j,k}^{\sigma_{i,j}}.$$

In this expression, the right-hand side still depends on the hillslope scale, through $H_{i,j,k}^{\sigma_{i,j}}$. To remove this dependence, we assume that the random variables $H_{i,j,k}^{\sigma_{i,j}}$, $k = 1, 2, \dots, m_j$ are iid and

$$U_{i,j}^{\sigma_i} \stackrel{d}{=} H_{i,j,k}^{\sigma_{i,j}}. \tag{13}$$

This equation holds for each k and gives

$$(Y_{i,j,k}|Y_{i,j} = y_{i,j}) \stackrel{d}{=} \exp(\alpha_i) y_{i,j}^{\beta_i} a'_{j,k}{}^{\gamma_i} U_{i,j}^{\sigma_i};$$

$$k = 1, 2, \dots, m_j.$$

Now, $(Y_{i,j,k}|Y_{i,j} = y_{i,j})$ is a function of variables at the unnested basin and basin scales where observations exist.

The distributions of $(Y_{i,j}|Y_i = y_i)$ and $(Y_{i,j,k}|Y_{i,j} = y_{i,j})$ must be specified to assign water loss to hillslopes. We assume that

$$(Y_{i,j}|Y_i = y_i) \stackrel{d}{=} \text{LN}(\mu_{i,j}, \sigma_i^2) \tag{14}$$

$$(Y_{i,j,k}|Y_{i,j} = y_{i,j}) \stackrel{d}{=} \text{LN}(\mu_{i,j,k}, \sigma_i^2);$$

$$\mu_{i,j,k} = \ln(\exp(\alpha_i) y_{i,j}^{\beta_i} a'_{j,k}{}^{\gamma_i}), \tag{15}$$

where LN refers to a lognormal random variable and the distribution parameters in parentheses denote the mean and variance of the associated normal distribution. Here, normal distribution variances are constant in space (among hillslopes and unnested basins). The assumption that distributions are lognormal requires that total rainfall exceeds total loss for a rain event; i.e. $(Y_{i,j}|Y_i = y_i) > 0$, $(Y_{i,j,k}|Y_{i,j} = y_{i,j}) > 0$, and runoff occurs. It also requires that total rainfall is positive. Under this assumption, we have

$$U_{i,j} \stackrel{d}{=} H_{i,j,k} \stackrel{d}{=} \text{LN}(0, 1) \tag{16}$$

$$(Y_{i,j,k}|Y_{i,j} = y_{i,j}) \stackrel{d}{=} \exp(\alpha_i) y_{i,j}^{\beta_i} a'_{j,k}{}^{\gamma_i} U_{i,j}^{\sigma_i},$$

indicating that $(Y_{i,j,k}|Y_{i,j} = y_{i,j})$ is lognormally distributed for a given event and unnested basin. The assumption also means that $\sigma_{i,j} = \sigma_i$, where $\sigma_i = \sqrt{\text{Var}(\ln(Y_{i,j})|Y_i = y_i)}$ and $\sigma_{i,j} = \sqrt{\text{Var}(\ln(Y_{i,j,k})|Y_{i,j} = y_{i,j})}$. Finally, the assumption means that Eq. (8a), given $J = j$, and Eq. (8b), given $K = k$, take the form of a linear regression model. Notice that the equality in σ 's follows from Eq. (13). It is a first-order approximation and represents the simplest case of the expression, $\sigma_{i,j} = c\sigma_i$. Given the difference in scale between σ_i and $\sigma_{i,j}$, it is possible that $c > 1$.

Table 3. Results of the k -sample Kolmogorov–Smirnov (KS), k -sample Z_c , and Lilliefors tests for $(\psi_{i,j}/g_i)^{1/\hat{\sigma}_i}$.

Test	H_0 – null hypothesis	Do not reject H_0 if ...	Case : result
k -sample KS	Distributions are similar	$c < 0.462$	I: $c = 0.091$ II: $c = 0.083$ III: $c = 0.154$
k -sample Z_C	Distributions are similar	$p > 0.05$	I: $p = 1.0$ II: $p = 1.0$ III: $p = 1.0$
Lilliefors	Distributions of the logarithm are normal	$l = 0$	I: $l = 0$ for 90 events of 112 II: $l = 0$ for 60 events of 75 III: $l = 0$ for 17 events of 20

The observations given in Fig. 3b, which show that rescaling $\psi_{i,j}$ as $(\psi_{i,j}/g_i)^{1/\hat{\sigma}_i}$ leads to a collapse in distributions, supports the assumption that $\psi_{i,j}$ is lognormally distributed. The physical basis of this feature is linked to the hypothesis explained in Sect. 3.5, where it is stated that the spatial distribution of saturated hydraulic conductivity, which influences soil moisture and thus water loss, is commonly found to be lognormal. As mentioned in Sect. 4.4, we interpret $U_{i,J}$ to represent antecedent soil moisture conditions for event i . A similar interpretation is made for $U_{i,j}$ and $H_{i,j,k}$ where fixed event spatial variability arises through spatial changes in soil moisture.

If there is a need to generate total runoff at the hillslope scale for an observed rainfall event, then we need to make an assumption about the statistical connection between the runoff/loss ratio and total rainfall. We shall assume that the runoff/loss ratio and rainfall are statistically independent at both the unnested basin and hillslope scales; that is, the random variables $Y_{i,j}$ and $Y_{i,j,k}$ are all independent of the random variables $R_{i,j}$ and $R_{i,j,k}$. Thus, the model in Eq. (16) for the runoff/loss ratio will remain the same if total rainfall at both scales is known and fixed.

Equation (16), together with the independence assumption in the previous paragraph, is the model that we set out to develop as it provides a way of assigning total water loss to hillslopes in three steps. First, we obtain values for $y_{i,j}$ from observations of $\psi_{i,j}$ at the unnested basin scale. Second, based on Eq. (8a) when conditioned on $J = j$, we obtain estimates of α_i , β_i , γ_i , and σ_i via linear regression using observations of ψ_i and $\psi_{i,j}$ and drainage areas of unnested basins. Results from these two steps allow us to use Eq. (16) to generate realizations $y_{i,j,k}$ for each hillslope k . Third, we determine values of total loss $l_{i,j,k} = r_{i,j,k}/(1 + y_{i,j,k})$ given that we have a realization $y_{i,j,k}$ and an observation-based estimate of $r_{i,j,k}$. An example is given below in Sect. 5.3.

4.7 Connection to water balance

We examined how the runoff–loss model is connected to water balance under the following simplified conditions. For event i and basin j , let hillslope area $a_{j,k}$ be the same for

all k , which means that $a'_{j,k} = 1$ in Eq. (16). Water balance for unequal hillslope areas needs a more complicated formulation and is beyond the scope of this paper. Let the total rainfall for event i be known at the unnested basin and hillslope scales; that is, $r_{i,j}$ and $r_{i,j,k}$ are fixed and nonrandom. Also, let event i produce runoff, meaning that total rainfall over basin j exceeds total water loss or $r_{i,j} > l_{i,j}$ and $y_{i,j} > 0$.

Under the conditions above, water balance for an event i requires that (1) observed and modeled total runoff from basin j are equal and that (2) modeled totals of rainfall, runoff, and water loss in basin j equal the summation of corresponding hillslope totals, over all k hillslopes in basin j . Requirement 1 must be satisfied on average because water loss is treated probabilistically. This situation is similar to mean mass conservation, which is used to construct canonical ensembles in statistical mechanics and is central to the theory of random cascades, as illustrated in Sect. 4.2.

Requirement 2 is satisfied if

$$r_{i,j} = 1/m_j \sum_{k=1}^{m_j} r_{i,j,k}, \tag{17a}$$

$$l_{i,j} = 1/m_j \sum_{k=1}^{m_j} E[L_{i,j,k} | L_{i,j} = l_{i,j}], \tag{17b}$$

where $(L_{i,j,k} | L_{i,j} = l_{i,j}) = (L_{i,j,K} | L_{i,j} = l_{i,j}, K = k)$. These equations represent conservation of mass with respect to total rainfall depth and total depth of water loss. The condition under which Eq. (17b) holds is seen by noting that the definition of $Y_{i,j,K}$ gives $L_{i,j,K} = R_{i,j,K}/(Y_{i,j,K} + 1)$. When rainfall $r_{i,j,k}$ is fixed for all k , as shown in Eq. (17a), we have $L_{i,j,k} = r_{i,j,k}/(Y_{i,j,k} + 1)$ in Eq. (17b). Also, the definition of $Y_{i,j}$ gives $L_{i,j} = R_{i,j}/(Y_{i,j} + 1)$. When rainfall $r_{i,j}$ is fixed for all j , conditioning on $L_{i,j} = l_{i,j}$, as shown in Eq. (17b), is equivalent to conditioning on $Y_{i,j} = y_{i,j}$; this feature follows because $L_{i,j} = R_{i,j}/(Y_{i,j} + 1)$ becomes $l_{i,j} = r_{i,j}/(y_{i,j} + 1)$. Thus, with rainfall known at the unnested basin and hillslope scales, Eq. (17b) yields

$$\begin{aligned}
 l_{i,j} &= 1/m_j \sum_{k=1}^{m_j} E \left[\frac{r_{i,j,k}}{Y_{i,j,k} + 1} \middle| Y_{i,j} = y_{i,j} \right], \\
 &= 1/m_j \sum_{k=1}^{m_j} E \left[\frac{r_{i,j,k}}{\exp(\alpha_i) y_{i,j}^{\beta_i} \text{LN}(0, \sigma_i^2) + 1} \right], \\
 &= r_{i,j} E \left[\frac{1}{\exp(\alpha_i) y_{i,j}^{\beta_i} \text{LN}(0, \sigma_i^2) + 1} \right].
 \end{aligned}$$

Conservation of mass with respect to water loss holds when the expectation in the last equality equals $1/(y_{i,j} + 1)$. This expectation takes the form of $E[1/(X + 1)]$ with X log-normal, for which there is no analytical expression. In the limit as $\sigma_i^2 \rightarrow 0$ the expectation is given by $1/(\exp(\alpha_i) y_{i,j}^{\beta_i} + 1)$. Further, for small σ_i , an asymptotic approximation involving powers of σ_i can be made using standard expansion techniques for integrals, but these results are beyond the scope of this paper. To honor conservation of mass in the small variance case, it is necessary for $\alpha_i \rightarrow 0$ and $\beta_i \rightarrow 1$ as $\sigma_i^2 \rightarrow 0$ so that, in the limit, the expectation equals $1/(y_{i,j} + 1)$. Observations presented in Sect. 5.1 show that α_i and β_i are near these limits.

Conservation of mass with respect to runoff, a component of Requirement 2, follows from Eqs. (17a) and (17b). To see the logic of this connection, consider that randomness in hillslope water-loss depth must lead to randomness in hillslope and basin runoff. Let $Q_{i,j} = (Q_{i,J} | J = j)$ be a random variable for the total volume of runoff for event i and basin j . An observed value of runoff, a realization of $Q_{i,j}$, is given as

$$q_{i,j} = (r_{i,j} - l_{i,j}) a_j \quad ; \quad r_{i,j} > l_{i,j}. \tag{18}$$

Substituting Eqs. (17a) and (17b) into Eq. (18) gives

$$\begin{aligned}
 q_{i,j} &= \left(1/m_j \sum_{k=1}^{m_j} r_{i,j,k} - 1/m_j \sum_{k=1}^{m_j} E[L_{i,j,k} | L_{i,j} = l_{i,j}] \right) a_j \tag{19} \\
 &= \sum_{k=1}^{m_j} (r_{i,j,k} - E[L_{i,j,k} | L_{i,j} = l_{i,j}]) a_{j,k}, \\
 &= \sum_{k=1}^{m_j} E[Q_{i,j,k} | L_{i,j} = l_{i,j}].
 \end{aligned}$$

This equation shows that the observed total basin runoff equals the summation of expected total runoff from hillslopes, which supports Requirement 2, i.e. mean mass conservation in terms of runoff. If Eqs. (17a) and (17b) hold true, then so does Eq. (19) and the runoff–loss model preserves water balance.

5 Analysis and results

5.1 Testing model relationships and assumptions

5.1.1 Equation (9): distributions of rescaled runoff/loss ratios at the unnested basin scale for random unnested basins

Figure 4 shows the empirical CDFs of runoff/loss ratios at the unnested basin scale, $\psi_{i,j}$, and the CDFs of rescaled ratios given as $(\psi_{i,j}/g_i)^{1/\hat{\sigma}_i}$, where g_i and $\hat{\sigma}_i$ denote observation-based estimates of $G_i(y_i)$ and σ_i in Eq. (9). The figure presents results for Cases I, II, and III. As explained in Sect. 3.4, the number of events and unnested basins for these cases are, respectively, (112,11), (75,12), and (20,13). The first case consists of the same events examined in Sect. 4.1. Results in the figure suggest, qualitatively, that rescaled ratios come from the same probability distribution.

Table 3 shows the results of comparing CDFs of rescaled ratios using the k -sample Kolmogorov–Smirnov test (Conover, 1999) and the k -sample Z_C test (Zhang and Wu, 2007), which is shown to be more powerful. The first test requires that the number of points in each CDF is the same in the group of CDFs to be compared. Plots in Fig. 4 meet this requirement. Results indicate that rescaled distributions are statistically identical among all events.

Table 3 also shows the results of applying a Lilliefors test to the distribution of $\ln((\psi_{i,j}/g_i)^{1/\hat{\sigma}_i})$, the logarithm of rescaled runoff/loss ratios, for each event. The Lilliefors test allows for both the mean and variance of $\ln(\psi_{i,j})$ to change from event to event, and thus it gives identical results when applied to $(\psi_{i,j}/g_i)^{1/\hat{\sigma}_i}$, which is both mean- and variance-corrected. Results in the table indicate whether a correction by the mean and variance yields identical distributions. For Case I events, they show that the hypothesis of normality is rejected at the 5 % level for 22 of the 112 CDFs. Some or all of the rejections could be erroneous, a type I error. Applying the same tests using a Bonferroni correction to account for the large number of individual tests (Hsu, 1996) indicates that all CDFs are normal except one. Similar results are found for Cases II and III.

Figure 5 shows quantile–quantile (QQ) plots (Wilk and Gnanadesikan, 1968) for $\ln((\psi_{i,j}/g_i)^{1/\hat{\sigma}_i})$, representing the logarithm of the rescaled ratios in Fig. 4. For each case, we grouped the ratios from all events to make a plot. The lines in the QQ plots for Cases I and II are relatively straight, suggesting normality or approximate normality among rescaled ratios; this linearity disappears if σ_i is not used to rescale observations. Results for Case III deviate from normality. Based on Eq. (9) in Sect. 4.4, the deviations could mean that $(\gamma_i/\sigma_i) \ln\left(\frac{a_j}{G[a_j]}\right)$ varies among some or all events. Alternatively, $(\gamma_i/\sigma_i) \ln\left(\frac{a_j}{G[a_j]}\right)$ is constant among all events, but $\ln(U_{i,j})$ is not iid among all events, as assumed. The important issue of lognormality is analyzed later in Sect. 5.1.3.

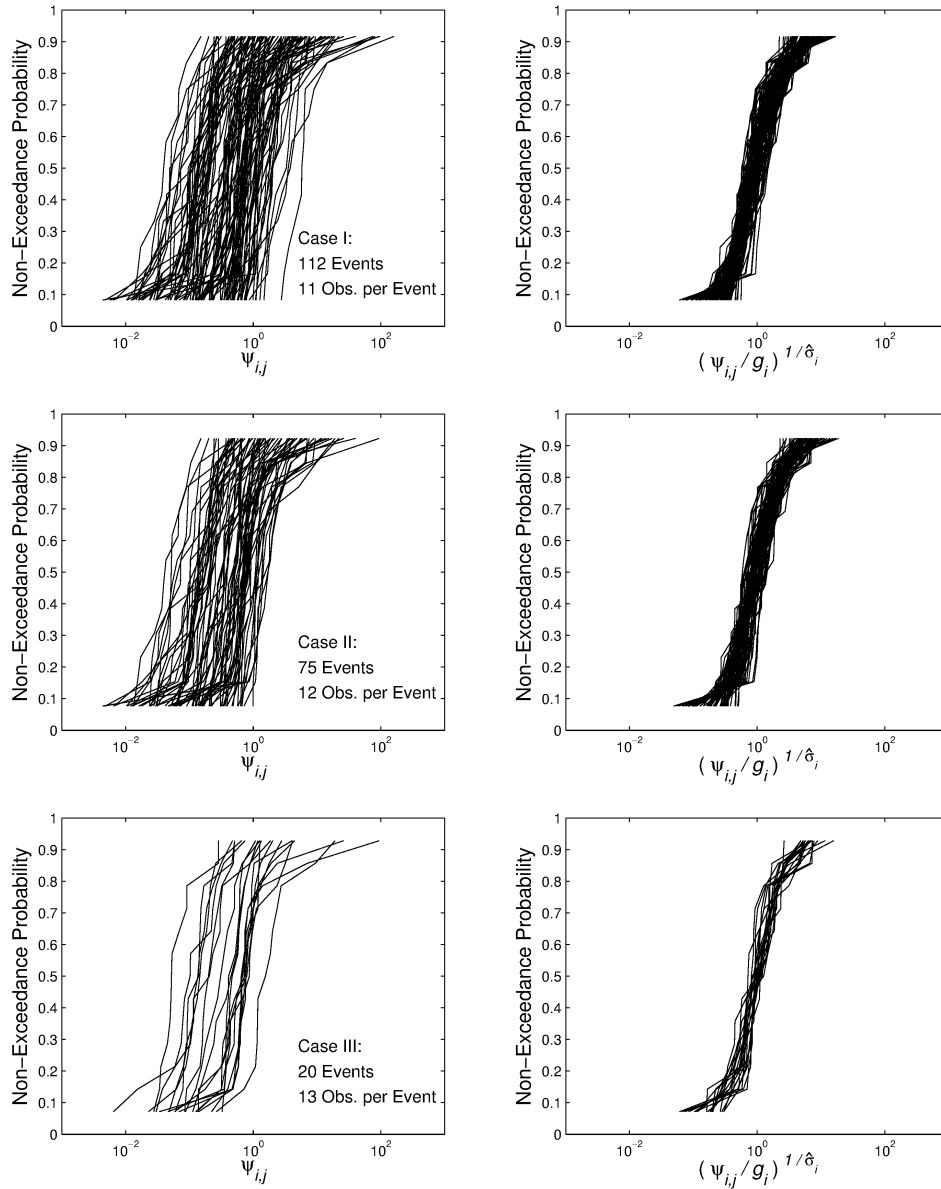


Fig. 4. Left: Semi-log plot of empirical CDFs of runoff/loss ratios at the unnested basin scale, $\psi_{i,j}$, for Cases I to III. Each distribution represents a rainfall–runoff event. Right: Semi-log plot of corresponding CDFs of rescaled runoff/loss ratios, $(\psi_{i,j}/g_i)^{1/\hat{\delta}_i}$.

5.1.2 Equation (8): relationship between runoff/loss ratios at the unnested basin and basin scales for fixed unnested basins

Equation (8) provides an expression for $\ln(Y_{i,j})$ for event i and unnested basin j , which takes the form of a multiple weighted linear regression model where there are two explanatory variables, $\ln(Y_i)$ and $\ln(a'_j)$, and a weight given by $1/\sigma_i^2$. This model assumes that, on average, $\ln(Y_{i,j})$ depends linearly on $\ln(Y_i)$ and $\ln(a'_j)$ for an event, but also allows for the possibility that values of the coefficients α_i , β_i ,

and γ_i change among events. The model is not identifiable if all three coefficients α_i , β_i , and γ_i are allowed to change with event; the problem that arises is similar to trying to fit a straight line to data where there is only one value for the explanatory variable. Thus, two of the three parameters must be fixed (no variation from event to event). In our analysis, we fixed the value of $\beta_i = \beta$ and $\gamma_i = \gamma$ and treated the intercept, α_i , as a random variable (random from event to event). The analysis also considered the influence that changes in σ_i between events can have on results. We used the nlme package in R to obtain results (Gentleman et al., 2013).

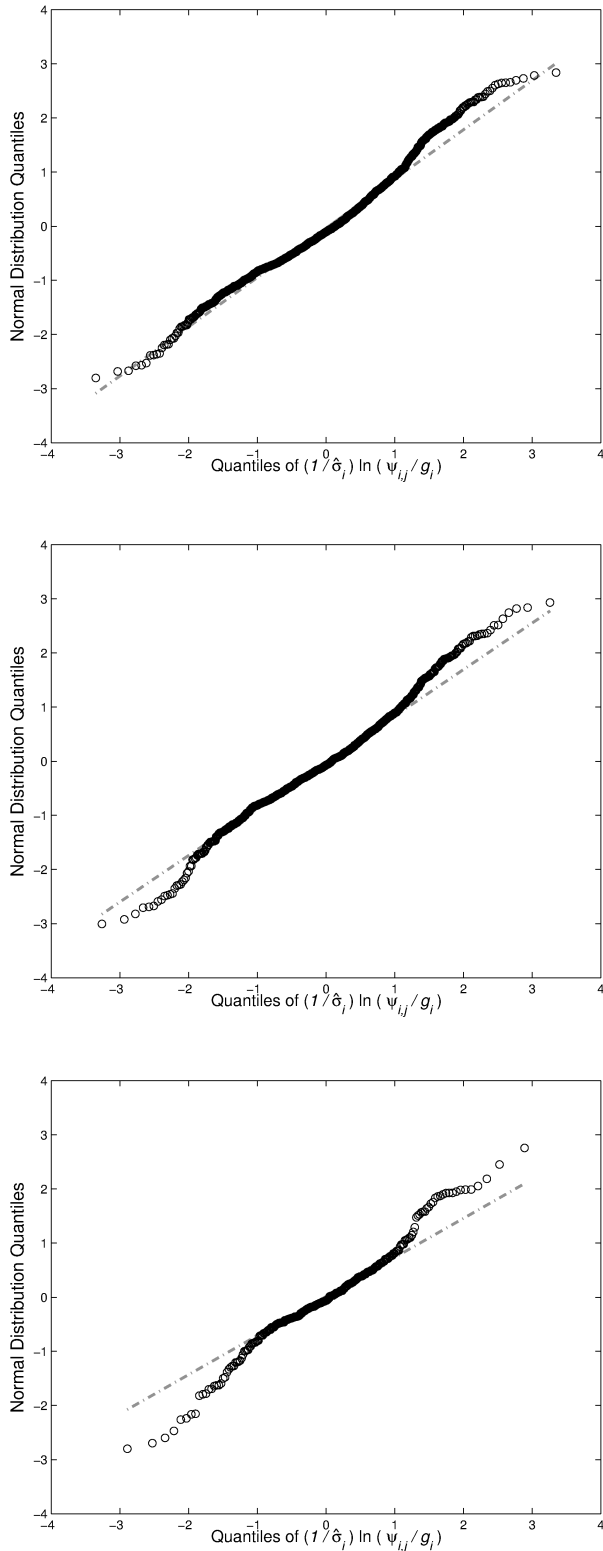


Fig. 5. QQ plot of $\ln((\psi_{i,j}/g_i)^{1/\hat{\sigma}_i})$, the rescaled runoff/loss ratios in Fig. 4 after a log-transformation.

Results from the statistical analysis indicate that there is significant variability in the value of α_i between events for Cases I and II but not Case III. For Case I, $\hat{\beta} = 1.027$, $\hat{\gamma} = -0.364$, and the mean of α_i is estimated to be 0.316. The variance of α_i among events is estimated to be 0.009 and, because this value exceeds zero, α_i changes between events as assumed in our model. For Case II, $\hat{\beta} = 1.104$, $\hat{\gamma} = -0.342$, and the mean of α_i is estimated to be 0.449. The variance of α_i among events is estimated to be 0.002, and thus α_i changes between events. By contrast, results for the events in Case III indicate that the variance of α_i is not significantly greater than zero, so that all the α_i are equal to a single value α . For Case III, $\hat{\beta} = 1.034$, $\hat{\gamma} = -0.415$, and $\hat{\alpha} = 0.185$.

The left-hand column of Fig. 6 shows that observations of $\ln(\psi_{i,j})$ and $\ln(\psi_i)$ for Cases I to III are, on average, linearly related across events. The lines presented in the plots illustrate the influence of α_i on this relationship. To plot the lines, we used values of α_i and $\beta_i = \beta$ obtained from our statistical analysis but assumed $\gamma_i \ln(a'_i) = 0$ so that, based on Eq. (8), $\ln(\psi_{i,j})$ depends only on $\ln(\psi_i)$ and a line can be plotted. The bottom line in a plot represents the minimum value of α_i among events, while the top line represents the maximum value of α_i .

The left-hand column of Fig. 7 presents the same results shown in Fig. 6 but using only events that pass the Lilliefors test. The reduced number of events can be found in Table 3. Comparing the left-hand columns of Figs. 6 and 7 reveals that many of the events removed include “extreme” values of $\psi_{i,j}$. This result suggests that events where the distribution of $\ln(U_{i,j})$ is not Normal, according to the Lilliefors test, tend to have extreme values. Non-normality could arise if there are correlations among observations during these events that are unaccounted for in the model, if $\ln(U_{i,j})$ is not Normal due to some unique physical conditions in the rainfall–runoff process, or if the model does not well represent extreme events. Alternatively, the extreme values may simply represent measurement errors.

We have applied Eq. (8) as a linear mixed-effects statistical model where randomness is treated separately for each event. A measure of goodness of fit, like R^2 , can be difficult to interpret for such a model (Nakagawa and Schielzeth, 2013) and is not provided in Figs. 6 and 7. However, if a simple linear model is applied across all events for each of the cases in the two figures, then R^2 values are around 0.6. A comparable and possibly better goodness of fit can be expected for a linear mixed-effects model. One indication that this situation holds is that the Akaike information criterion (Akaike, 1974) is slightly better when treating events individually (linear mixed effects) instead of collectively.

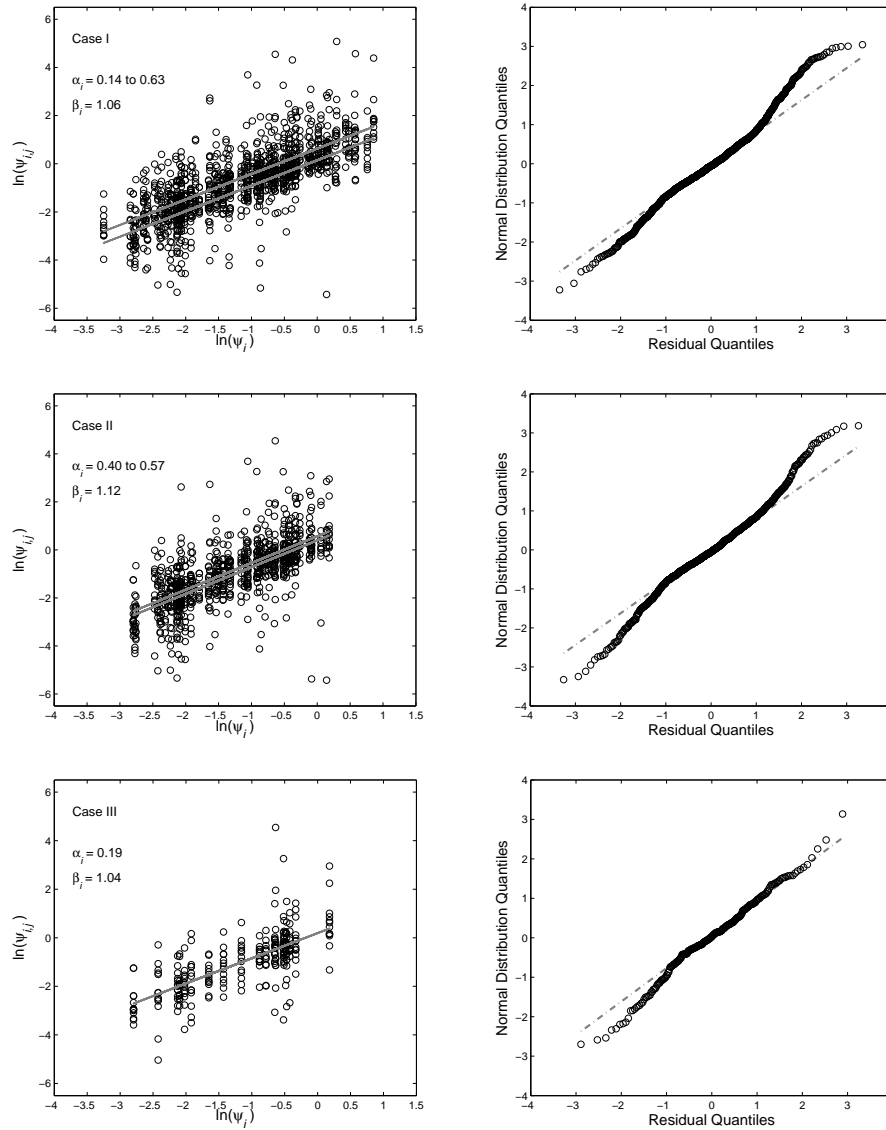


Fig. 6. Left: Plot of $\ln(\psi_{i,j})$ versus $\ln(\psi_i)$ for Cases I to III. To plot the lines, we used values of α_i (intercept) and $\beta_i = \beta$ (slope) obtained from our statistical analysis, but assumed $\gamma_i \ln(a'_j) = 0$; see Eq. (8). The bottom line in a plot represents the minimum value of α_i among events, while the top line represents the maximum value of α_i . Right: QQ plot of residuals obtained from statistical analysis results.

5.1.3 Equation (14): lognormality of runoff/loss ratios at the unnested basin scale for fixed unnested basins

We tested the assumption given by Eq. (14) that $(Y_{i,j}|Y_i = y_i)$ is equal in distribution to a lognormal. For this test, we compared the relationship between quantiles of a normal distribution and those of residuals from the statistical analysis results described in Sect. 5.1.2. If Eq. (14) is correct, then the residuals are normally distributed as $\ln(U) - E[\ln(U)]$. The right-hand column of Fig. 7 presents this comparison in a QQ plot for Cases I to III and indicates that the distribution of residuals is close to normality. This result supports

the lognormal assumption given by Eq. (14). By comparison, the QQ plots in Figs. 5 and 6 are, overall, further from normality. Differences in QQ plots between the three figures underscore the importance of both the area term in the model and Lilliefors test results.

A QQ plot provides a qualitative indication of normality, but the correlation coefficient of such a plot could be used to quantify the degree to which normality is achieved. Accordingly, we could compare the correlation coefficients of the QQ plots in Fig. 6 against those in Fig. 7 to quantify changes between the figures. This kind of analysis, however, is beyond the scope of our paper.

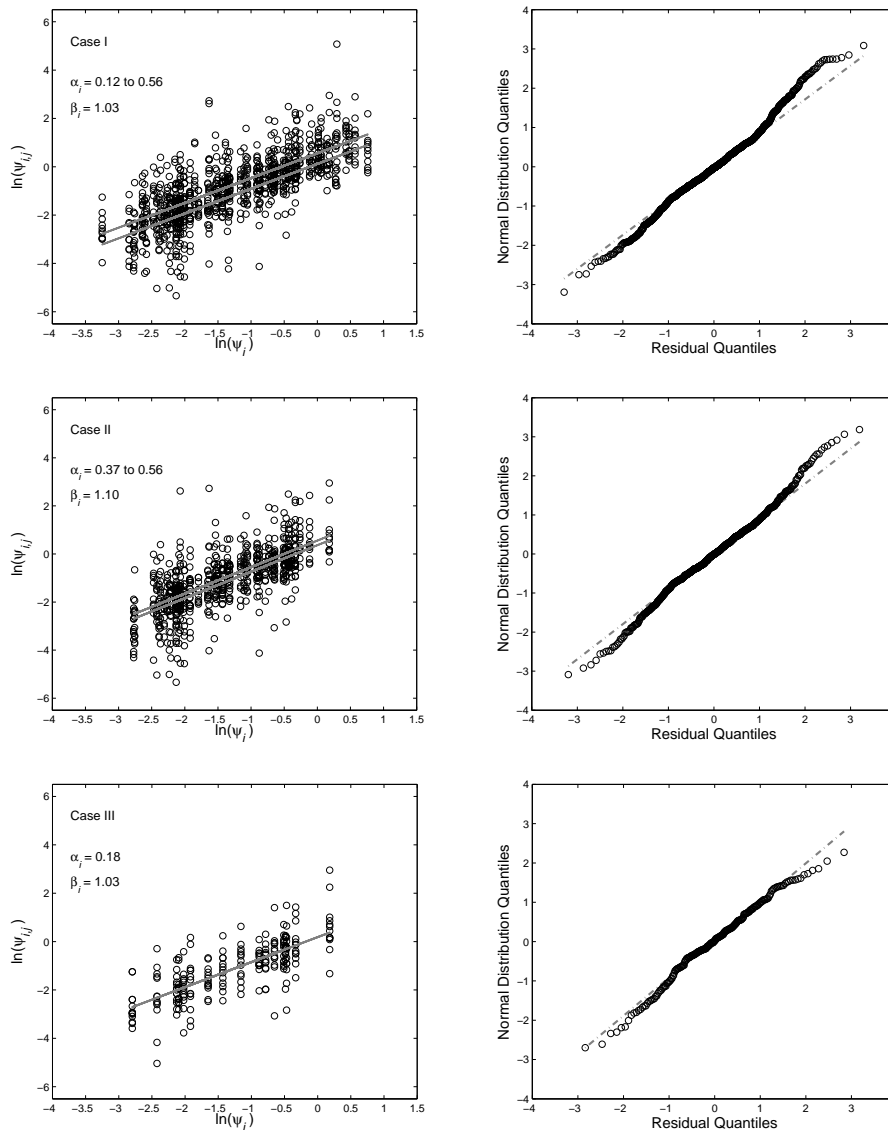


Fig. 7. Left: Plot of $\ln(\psi_{i,j})$ versus $\ln(\psi_i)$ for Cases I to III using only those events that passed the Lilliefors test. To plot the lines, we used values of α_i (intercept) and $\beta_i = \beta$ (slope) obtained from our statistical analysis, but assumed $\gamma_i \ln(a'_j) = 0$. The bottom line in a plot represents the minimum value of α_i among events, while the top line represents the maximum value of α_i . Right: QQ plot of residuals obtained from statistical analysis results. See Fig. 6 for comparison.

5.1.4 Statistical independence between runoff/loss ratios and rainfall depths at the unnested basin scale

We tested the assumption given after Eq. (16) that the runoff loss ratio is statistically independent of total rainfall. Figure 8 shows test results for Case I events that passed the Lilliefors test. The first plot in the figure indicates that observed rainfall at the unnested basin scale, $r_{i,j}$, is largely independent of the runoff/loss ratio at the basin scale, ψ_i . The slight positive trend in the plot is not observed for Case II and III events. The second plot compares $r_{i,j}$ to residuals from the fitted model

described in Sect. 5.1.2 and derived from the data presented in Fig. 7. The residuals represent the unexplained variability in values of $y_{i,j}$ after accounting for the influence of y_i and a'_j . Residuals appear to be independent of $r_{i,j}$, and this same result is found for the Case II and III events. Thus, observations support our assumption.

5.2 Testing water balance conditions of data

The water balance formulation in Sect. 4.7 depends on two mass conservation conditions, Eqs. (17a) and (17b), that relate the unnested basin and hillslope scales. For simplicity,

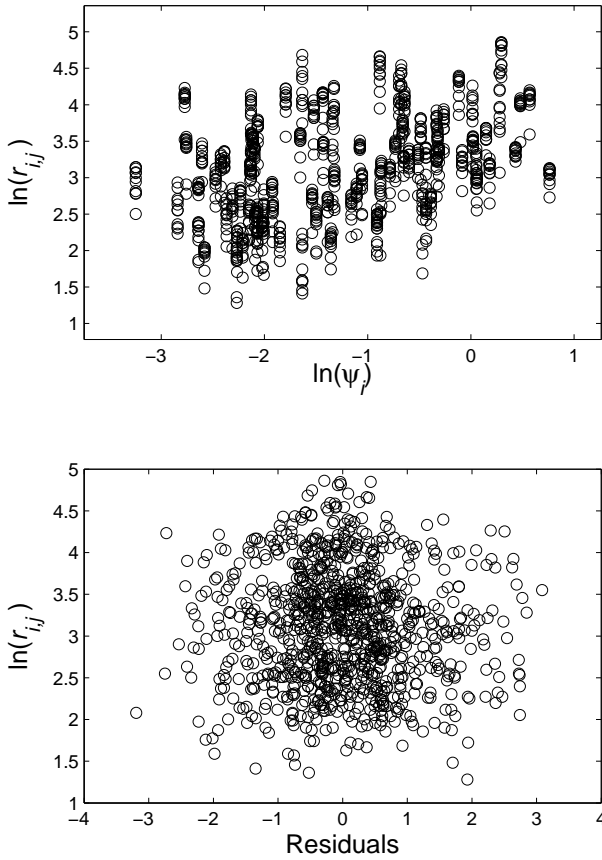


Fig. 8. Plot A shows the relationship between the logarithm of rainfall depth, $\ln(r_{i,j})$, and $\ln(\psi_i)$ for Case I events. Plot B shows the relationship between $\ln(r_{i,j})$ and regression residuals from the first plot (Case I) of Fig. 6.

we formulated both equations under the condition that hillslopes have the same area, but they hold without this constraint when rainfall and water loss are weighted by area. Neither equation can be tested directly because rainfall observations and water-loss observations are unavailable at the hillslope scale. However, similar conservation conditions must hold that relate the basin and unnested basin scales.

We examined whether data at the basin and unnested basins scales honor mass conservation. Plot A in Fig. 9 shows that total rainfall at the basin scale, r_i , equals total area-weighted rainfall at the unnested basin scale, $\bar{r}_{i,j} = \sum_{j=1}^n r_{i,j} a_j / a$. Similarly, Plot B in the figure shows that total water loss at the basin scale, l_i , equals total area-weighted loss at the unnested basin scale, $\bar{l}_{i,j} = \sum_{j=1}^n l_{i,j} a_j / a$. Thus, our data for total event rainfall and water loss preserve mass conservation for Case III events. Cases I and II involve 11 and 12 unnested basins, respectively, and their areas do not cover GCEW. In these cases, total water loss at the basin scale should exceed or equal the total area-weighted loss at the unnested-basin scale. Observations indicate that this condition holds.

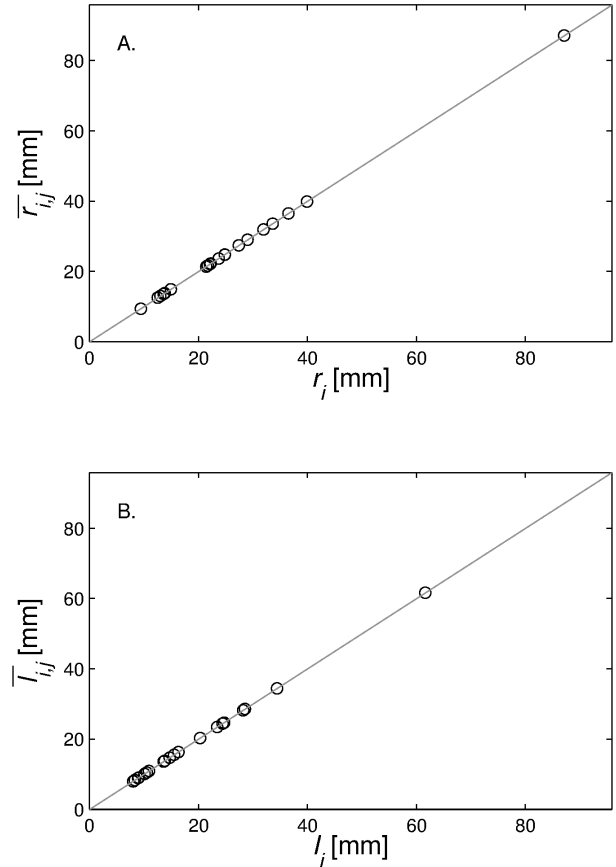


Fig. 9. Relationship between $\bar{r}_{i,j}$ and r_i (Plot A) and $\bar{l}_{i,j}$ and l_i (Plot B) for Case III events. The 1 : 1 relationship in both plots means that mass is conserved.

5.3 Model application

We applied the runoff–loss model given by Eq. (16) to event $i = 2$, a Case III rainfall–runoff event that occurred on 17 March 1982. We chose this event because the magnitude of its runoff/loss ratios are close to the median of the range of those found for all Case I to III events. For this event, we let $\alpha_2 = 0.185$, $\beta_2 = 1.034$, $\gamma_2 = -0.415$, and $\sigma_2^2 = 0.768$. We obtained these values from the statistical analysis results above. They are not event-specific values insofar that results for Case III show that parameter values do not differ significantly between events.

We first used Eq. (16) to produce a map of expected hillslope loss for event $i = 2$; see Fig. 10. Three steps were taken to make this map. We calculated the expected runoff/loss ratio for each of the 544 hillslopes in GCEW. In the model, this quantity is $E[Y_{2,j,k} | Y_{2,j} = y_{2,j}]$, where $k = 1, 2, \dots, m_j$ and $\sum_{j=1}^{13} m_j = 544$. We then determined corresponding values of hillslope loss,

$$l_{2,j,k} = r_{2,j,k} / (1 + E[Y_{2,j,k} | Y_{2,j} = y_{2,j}]).$$

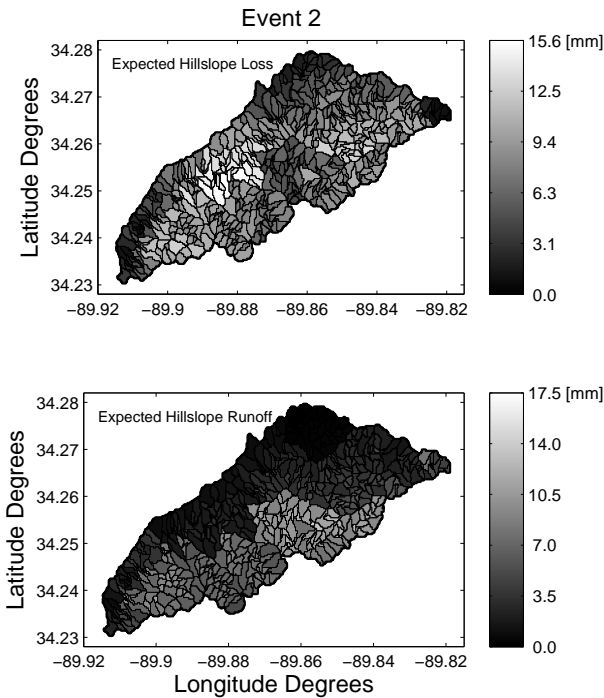


Fig. 10. The expected water loss and runoff of each hillslope for event $i = 2$; i.e. $E[L_{2,j,k}]$ and $r_{2,j,k} - E[L_{2,j,k}]$.

Finally, we added a bias correction factor to each value of $l_{2,j,k}$ so that the new value approximates $E[L_{2,j,k}]$. The difference between $l_{2,j,k}$ and $E[L_{2,j,k}]$ turns out to be rather small, numerically. This last step is taken for mathematical consistency and does not involve a model fitting procedure, i.e. calibration.

We also produced model realizations of water loss and runoff at both the hillslope and unnested basin scales. First, we used Eq. (16) to produce 50 hillslope-loss sets, each representing a possible spatial configuration of hillslope water loss for event $i = 2$. Then, for each hillslope-loss set, we determined realizations of both water loss and runoff at the unnested basin scale. For water loss, we added the hillslope-loss values within each unnested basin j . For runoff, we calculated the difference between total hillslope rainfall and hillslope loss and then added the differences to obtain total runoff for each unnested basin.

Figure 11 compares water-loss and runoff model realizations to observations. Plot A in the figure shows that the volume of rainfall exceeds the volume of model loss at the unnested basin scale, for all realizations; i.e. $r_{2,j} > l_{2,j}$ for each basin j and each of the 50 hillslope-loss sets. Likewise, hillslope rainfall always exceeds or equals hillslope loss. Plot B shows that total runoff values from the model at the unnested basin scale are, on average, equal to those obtained from observations. For a given unnested basin, the variability in total runoff between realizations originates

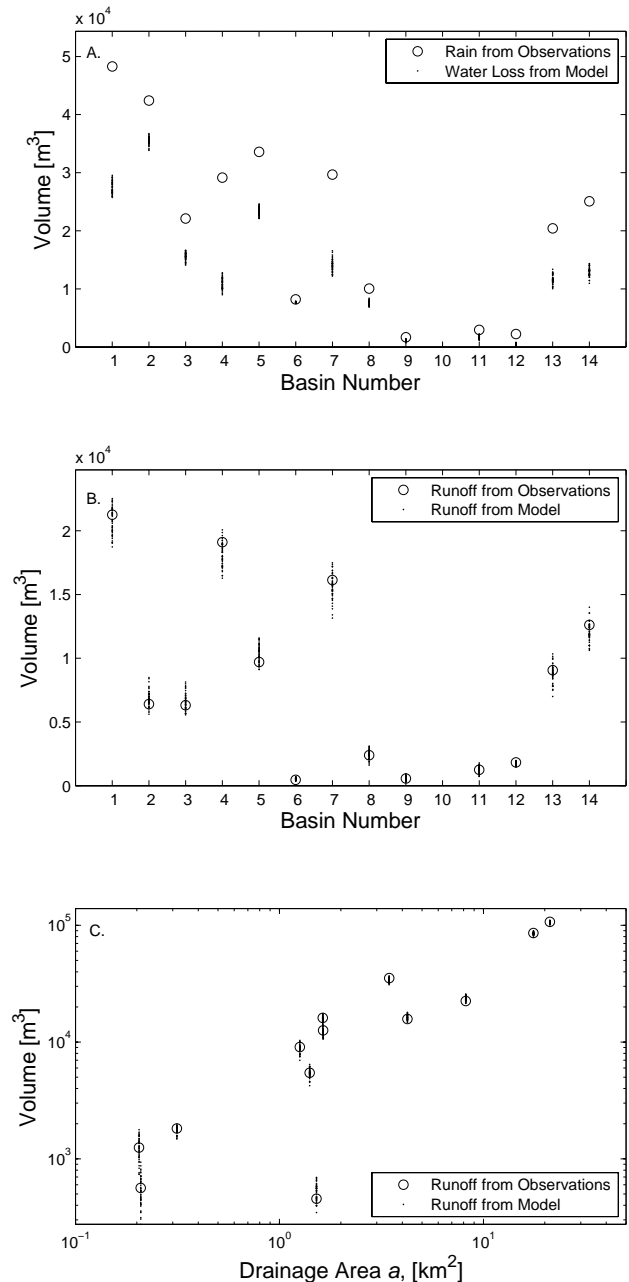


Fig. 11. Plot A shows the volume of observed rainfall and modeled water loss for each unnested basin. Plot B shows the volume of observed and modeled runoff for each unnested basin. On average, model realizations of runoff equal observed runoff. Plot C shows the log–log relationship between nested-basin runoff and upstream drainage area for both observations and model realizations. Note that runoff values do not include a baseflow component and thus are not equivalent to streamflow.

from differences in the spatial configuration and magnitudes of hillslope losses between realizations. Physically, such differences in hillslope loss could represent differences in antecedent (pre-rainfall) soil moisture. Plot C indicates that

different realizations of total runoff for event $i = 2$ produce slightly different scaling relationships. In other words, hillslope loss conditions impact scaling.

6 Discussion

We have used space–time patterns in $\psi_{i,j} = \tilde{q}_{i,j}/l_{i,j} = r_{i,j}/l_{i,j} - 1$ to develop an approach for mapping water loss and runoff at the hillslope scale. This metric is positive and has no upper bound given that $\tilde{q}_{i,j} > 0$ and $l_{i,j} > 0$. Spatial variability in $\tilde{q}_{i,j}/l_{i,j}$ for an event can be modeled by a probability distribution with support on the set of all positive numbers. Another metric could be used such as the ratio of total event runoff to rainfall, $\tilde{q}_{i,j}/r_{i,j}$. However, this metric has an upper bound of 1 given that total runoff cannot exceed total rainfall or $\tilde{q}_{i,j} \leq r_{i,j}$. Thus, by comparison, this metric would need to be modeled by a probability distribution with support restricted to values less than one. A transformation of a commonly used distribution would be needed in this case.

Model results given in Sect. 5.3 are encouraging given the different possible sources of error. Sources include model assumptions, parameter estimation, and our representation of space–time rainfall. In particular, in the model application we have used linear regression to obtain α_i for event $i = 2$ and, by model construction, have assumed that this value is the same for each unnested basin j . Using this value, model results do conserve mass in a mean sense. It can be shown that model results are sensitive to the value of α_i ; for instance, if α_2 is arbitrarily set to 0.4 then model runoff noticeably exceeds observed runoff. We also estimated rainfall fields using a simple spatial-interpolation method applied to rainfall data from 31 rain gauges. Other methods that better estimate the high spatial variability of rainfall, when it exists, could change model results. Nonetheless, there were no obvious problems in applying our model due to rainfall estimates, possibly because rainfall tends to cover GCEW completely during rainfall–runoff events.

We have developed the runoff–loss model in the context of a lognormal distribution. The events we examined support such a distribution, but it should also be possible to use an empirical distribution, one which is not necessarily lognormal. Empirical CDFs of rescaled runoff/loss ratios could be used instead of assuming lognormality. An algorithm could then be used to sample rescaled runoff/loss ratios from the empirical distributions.

Observations presented in Sect. 3.5 were used to hypothesize that hillslope water loss should be inversely related to a function of a lognormal distribution. Indeed, our model indicates that the distribution of hillslope loss for an event i in an unnested basin j is given by

$$(L_{i,j,K} | Y_{i,j} = y_{i,j}) \stackrel{d}{=} r_{i,j,K} / (\exp(\alpha_i) y_{i,j}^{\beta_i} a'_{j,K} \gamma_i \text{LN}(0, \sigma_i^2) + 1), \quad (20)$$

which shows this relationship. This result, obtained from a top-down approach, supports our physical-statistical

hypothesis. We note that the events used to formulate our hypothesis are the same as those used to develop our model, yet the hypothesis is general and not specific to any events. A more rigorous test of the hypothesis could be made using other events and perhaps different types of data.

7 Conclusions

Understanding the physical basis of observed scaling in peak flows in rainfall–runoff events (Ogden and Dawdy, 2003; Gupta et al., 2010) requires that runoff generation be modeled at a large number of hillslopes in a river basin. This need is challenging to meet because direct measurements of processes that produce runoff generation at hillslopes are generally unavailable or spatially limited within a basin. In addition, most rainfall–runoff models use point-scale equations and observations to represent runoff generation at larger scales, e.g. the hillslope scale, yet the validity of this bottom-up representation is unclear.

Based on basin- and point-scale observations in GCEW, we have hypothesized that hillslope water loss is inversely related to a function of a lognormal random variable. We have then developed a top-down model both to test this physical-statistical hypothesis and to provide a method of generating ensembles of runoff from a large number of hillslopes in a basin. The basis of the model is an observed, previously unreported, rescaling property of runoff/loss ratios. The model obeys water balance in a mean statistical sense, a mode of mass conservation found both in the theory of random cascades for modeling rainfall and turbulence and in statistical physics. The runoff–loss model supports our physical-statistical hypothesis, as shown by Eq. (20).

We have given an example application of the model using a rainfall–runoff event in GCEW, where we have observations of streamflow at multiple locations and reasonable estimates of spatial-mean rainfall depth at the basin, unnested basin, and hillslope scales. Model-generated values of water loss and runoff for hillslopes in GCEW were used to obtain corresponding model values for unnested basins. Comparison of model-generated and observed values for unnested basins support the model. Our results also reveal that hillslope water losses for an event can impact the spatial scaling of event runoff. Future tests of the model should use other rainfall–runoff events that collectively span a broad range of rainfall conditions. Unresolved issues have been identified in testing the model against data, and these also require future research.

The runoff–loss model can be used to constrain simulations of time-varying runoff at the hillslope scale where data are unavailable. For example, suppose there is a need to simulate water loss in time at the hillslope scale during a known rainfall–runoff event in GCEW. Such a simulation, involving infiltration, ET, interception, and subsequent runoff, will provide values of total loss and total runoff for each hillslope in the basin. These values will depend on the parameterization

and description of processes in the simulation. The simulation, its parameterization and formulation, could be constrained by matching simulated values of total loss and total runoff with those determined from the runoff–loss model. Our approach has potential applications in a diverse range of applied and theoretical research contexts where simulation of runoff generation from a large number of hillslopes is required.

Acknowledgements. The National Science Foundation supported this research through grant EAR-1005311 to the University of Colorado and EAR-10007324 to NorthWest Research Associates. We thank the three anonymous referees for their insightful input that greatly helped to improve our paper.

Edited by: S. Lovejoy

Reviewed by: three anonymous referees

References

- Achouri, M. and Gifford, G. F.: Spatial and seasonal variability of field measured infiltration rates on a rangeland site in Utah, *J. Rangeland Manage.*, 37, 451–455, 1984.
- Akaike, H.: A new look at the statistical model identification, *IEEE Trans. Automatic Control*, 19, 716–723, 1974.
- Amerman, C. R. and McGuinness, J. L.: Plot and small watershed runoff: Its relation to larger areas, *Trans. ASABE*, 10, 464–466, 1967.
- Bell, K. R., Blanchard, B. J., Schmutge, T. J., and Witzczak, M. W.: Analysis of surface moisture variations within large-field sites, *Water Resour. Res.*, 16, 796–810, 1980.
- Blackmarr, W. M. and the Channel and Watershed Processes Research Unit: Documentation of hydrologic, geomorphic, and sediment transport measurements on the Goodwin Creek Experimental Watershed, Northern Mississippi, for the period 1982–1993 – preliminary release, Tech. Rep. 3, Natl. Sediment Lab., Agric. Res. Serv., US Dept. of Agric., Oxford, MS., 1995.
- Brocca, L., Morbidelli, R., Melone, F., and Moramarco, T.: Soil moisture spatial variability in experimental areas of central Italy, *J. Hydrol.*, 333, 356–373, 2007.
- Brutsaert, W.: *Hydrology: An Introduction*, Cambridge University Press, 2005.
- Carvalho, H. O., Cassel, D. K., Hammond, J., and Bauer, A.: Spatial variability of in-situ unsaturated hydraulic conductivity of Mad-dock sandy loam, *Soil Science*, 121, 1–8, 1976.
- Chen, Z., Govindaraju, R. S., and Kavvas, M. L.: Spatial averaging of unsaturated flow equations under infiltration conditions over areally heterogeneous fields, 1. Development of models, *Water Resour. Res.*, 30, 523–533, 1994.
- Clark, R. D. S. and Hebbert, R. H. B.: An areal variability model for predicting storm runoff, Tech. rep., Inst. of Eng., Australia, 1971.
- Conover, W. J.: *Practical Nonparametric Statistics*, Wiley, 1999.
- Dunn, W. C., Milne, B. T., Mantilla, R., and Gupta, V. K.: Scaling relations between riparian vegetation and stream order in the Whitewater River network, Kansas, USA, *Landscape Ecology*, 26, 983–997, 2011.
- Feller, W.: *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd Edn., Wiley, 1968.
- Furey, P. R. and Gupta, V. K.: Effects of excess rainfall on the temporal variability of observed peak-discharge power laws, *Adv. Water Resour.*, 28, 1240–1253, 2005.
- Gargouri-Ellouze, E. and Bargaoui, Z.: Investigation with Kendall plots of infiltration index – maximum rainfall intensity relationship for regionalization, *Phys. Chem. Earth, Parts A/B/C*, 34, 642–653, 2009.
- Gentleman, R., Ihaka, R., and R Project Contributors: The R Foundation for Statistical Computing, available at: <http://www.r-project.org>, last access: August 2013.
- Gottschalk, L., Krasovskaia, I., Leblois, E., and Sauquet, E.: Mapping mean and variance of runoff in a river basin, *Hydrol. Earth Syst. Sci.*, 10, 469–484, doi:10.5194/hess-10-469-2006, 2006.
- Govindaraju, R. S., Corradini, C., and Morbidelli, R.: A semi-analytical model of expected areal-average infiltration under spatial heterogeneity of rainfall and soil saturated hydraulic conductivity, *J. Hydrol.*, 316, 184–194, 2006.
- Gupta, V. K. and Waymire, E. C.: A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteor.*, 32, 251–267, 1993.
- Gupta, V. K., Mantilla, R., Troutman, B. M., Dawdy, D., and Krajewski, W. F.: Generalizing a nonlinear geophysical flood theory to medium-sized river networks, *Geophys. Res. Lett.*, 37, L11402, doi:10.1029/2009GL041540, 2010.
- Gutmann, E. D. and Small, E. E.: A comparison of land surface model soil hydraulic properties estimated by inverse modeling and pedotransfer functions, *Water Res. Resour.*, 43, W05418, doi:10.1029/2006WR005135, 2007.
- Gutmann, E. D. and Small, E. E.: A method for the determination of the hydraulic properties of soil from MODIS surface temperature for use in land-surface models, *Water Res. Resour.*, 46, W06520, doi:10.1029/2009WR008203, 2010.
- Hsu, J. C.: *Multiple Comparisons: Theory and Methods*, Chapman and Hall – CRC, 1996.
- Kavvas, M. L., Chen, Z. Q., Dogrul, C., Yoon, J. Y., Ohara, N., Liang, L., Aksoy, H., Anderson, M. L., Yoshitani, J., Fukami, K., and Matsuura, T.: Watershed Environmental Hydrology (WHEY) model based on upscaled conservation equations: hydrologic module, *J. Hydrol. Eng.*, 9, 450–464, 2004.
- Kirkby, M.: Hillslope runoff processes and models, *J. Hydrol.*, 100, 315–339, 1988.
- Klemes, V.: Conceptualization and scale in hydrology, *J. Hydrol.*, 65, 1–23, 1983.
- Lan-Anh, N. T. and Willems, P.: Adopting the downward approach in hydrological model development: the Bradford catchment case study, *Hydrol. Process.*, 25, 1681–1693, 2011.
- Lowery, B., Cox, C., Lemke, D., Noowak, P., Olson, K., and Strock, J.: The 2008 Midwest flooding impact on soil erosion and water quality: Implications for soil erosion control practices, *J. Soil Water Conservation*, 64, 166A, doi:10.2489/jswc.64.6.166A, 2009.
- Maller, R. A. and Sharma, M. L.: An analysis of areal infiltration considering spatial variability, *J. Hydrol.*, 52, 25–37, 1981.
- Mantilla, R. and Gupta, V. K.: A GIS numerical framework to study the process basis of scaling statistics in river networks, *Geosci. Remote Sens. Lett.*, IEEE, 2, 404–408, 2005.

- Meng, H., Salas, J. D., Green, T. R., and Ahuja, L. R.: Scaling analysis of space-time infiltration based on the universal multifractal model, *J. Hydrol.*, 322, 220–235, 2006.
- Merz, B. and Bardossy, A.: Effects of spatial variability on the rainfall runoff process in a small loess catchment, *J. Hydrol.*, 212–213, 304–317, 1998.
- Nakagawa, S. and Schielzeth, H.: A general and simple method for obtaining R^2 from generalized linear mixed-effects models, *Methods Ecol. Evolution*, 4, 133–142, 2013.
- Nielsen, D. R., Biggar, J. W., and Erh, K. T.: Spatial variability of field-measured soil-water profiles, *Hilgardia*, 42, 215–260, 1973.
- Ogden, F. L. and Dawdy, D. R.: Peak discharge scaling in small Hortonian watershed, *J. Hydrologic Eng.*, 8, 64–73, 2003.
- Over, T. M.: Modeling space-time rainfall at the mesoscale using random cascades, Ph.D. thesis, U. of Colorado, Boulder, 1995.
- Over, T. M. and Gupta, V. K.: A space-time theory of mesoscale rainfall using random cascades, *J. Geophys. Res.*, 101, 26319–26331, 1996.
- Peck, A. J., Luxmoore, R. J., and Stolzy, J. L.: Effects of spatial variability of soil hydraulic properties in water budget modeling, *Wat. Resour. Res.*, 13, 348–354, 1977.
- Perica, S. and Fofoula-Georgiou, E.: Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions, *J. Geophys. Res.*, 101, 26347–26361, 1996.
- Philip, J. R.: Theory of infiltration, *Advan. Hydrosoci.*, 5, 215–96, 1969.
- Pilgrim, D. H.: Some problems in transferring hydrological relationships between small and large drainage basins and between regions, in: *Scale Problems in Hydrology*, edited by: Rodriguez-Iturbe, I. and Gupta, V. K., *J. Hydrol.*, 65, 49–72, 1983.
- Sauquet, E., Gottschalk, L., and Leblois, E.: Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme, *Hydrol. Sci. J.*, 45, 799–815, 2000.
- Sharma, A. S., Baker, D. N., Bhattacharyya, A., Bunde, A., Dimri, V. P., Gupta, H. K., Gupta, V. K., Lovejoy, S., Main, I. G., Schertzer, D., von Storch, H., and Watkins, N. W.: Complexity and Extreme Events in Geosciences: An Overview, in: *Extreme Events and Natural Hazards: The Complexity Perspective*, edited by: Sharma, A. S., A. Bunde, V. P. D., and Baker, D. N., 196, 371, American Geophysical Union, Geophysical Monograph Series, 2012.
- Shreve, R. L.: Infinite topologically random channel networks, *J. Geol.*, 75, 178–186, 1967.
- Sivapalan, M., Günter Blöschl, Zhang, L., and Vertessy, R.: Downward approach to hydrological prediction, *Hydrol. Process.*, 17, 2101–2111, 2003.
- Smith, R. E. and Hebbert, R. H. B.: A Monte Carlo analysis of the hydrologic effects of spatial variability of infiltration, *Water Resour. Res.*, 15, 419–429, 1979.
- Vieux, B. E.: *Distributed Hydrologic Modeling Using GIS*, Springer, 2004.
- Wilby, R.: *Contemporary Hydrology: Towards Holistic Environmental Science*, John Wiley, 1997.
- Wilk, M. B. and Gnanadesikan, R.: Probability plotting methods for the analysis of data, *Biometrika*, 55, 1–17, 1968.
- Zhang, J. and Wu, Y.: k-Sample tests based on the likelihood ratio, *Comput. Stat. Data Analysis*, 51, 4682–4691, 2007.