

EGUSPHERE-2025-5110

Title: Influence of rainfall event characteristics and antecedent conditions on subsurface stormflow response of two forested hillslopes

Authors: Emanuel Thoenes, Theresa Blume, Markus Weiler, Bernhard Kohl, Luisa Hopp, Stefan Achleitner

Point-by-point reply

We thank the reviewer for the careful reading of our manuscript and for the useful comments.

Reviewer 1 (Karl Auerswald)

General Comments

... In particular, I found the definition of rain events (break 6 h) critical, given that the duration of runoff events was much longer, with the falling limb of the hydrograph lasting 240 h. These short rain events were visible on the hydrograph, but they were not independent of the previous rain that caused the falling limb. The authors use the antecedent soil moisture conditions (ASI) to correct for this discrepancy between definitions of rain events and runoff events, but I wonder whether this attempt to correct one deficit by introducing another yields helpful and applicable findings. This notion can be supported by several arguments:

i) Whether a peak in rain intensity becomes visible in the hydrograph or not depends on the catchment size. Event definition and ASI cannot replace catchment size.

ii) I would expect the most prominent SSL event to occur when a large rain falls on very dry soil because then the largest contrast in hydraulic conductivity, differing by several orders of magnitude, will result above and below the wetting front. Hence, a low ASI will promote SSL, which is opposite to what happens when one rain event is split into two. Splitting a rain in two will thus cause an erroneous relation to ASI.

iii) For surface runoff events on very small plots with the falling limb lasting only minutes, it was found that even a break of 23 h was not sufficient to classify the second part as a separate rain because the hydrograph continued as if no break had occurred (for details see Auerswald and Gu, 2021). A modelling as a separate event at a higher ASI would have produced a different shape of the hydrograph that did not agree with the measured hydrograph.

Author's reply:

We thank you for your valuable comments. The rainfall events were delineated using the minimum inter-event time (MIT) method, with an MIT ("break") of 6 h, which is a widely adopted value (Dunkerley, 2008)(see also reply to L 218 ff). Due to our rainfall event and SSF event definition, multiple rainfall events can occur during a single SSF event. However, all rainfall events (occurring during an SSF event) are considered when analyzing the total volume of an SSF event (i.e. the cumulative total depth of all rainfall events occurring during an SSF event is used in the comparison with the generated SSF event volume). Furthermore, the single rainfall events are used when analyzing the single peaks (events/sub-events) visible on the hydrograph (see also reply to comment L 218 ff). We agree that if such a peak interrupts the falling limb of a previous SSF event, it will be influenced by the previous event as well; the conditions preceding the peak

(ASI and initial discharge) are therefore also considered in the analysis. Our method allowed us to also analyze the smaller/single peaks of larger (multi-peak) SSF events.

Detailed Comments

L 18-19: *No; rainfall and moisture were the only variables analysed. Hence, it is not known whether they primarily control SSF.*

Author's reply:

Besides total rainfall and antecedent wetness, we also analyzed the rainfall intensity as variable, which was secondary concerning the SSF volume. However, because before L 18-19 we do not list the analyzed rainfall characteristics, the reader would have not known this at this point in the text. To overcome this, we will add the analyzed rainfall characteristics before L 18-19.

L 22: *This number is misleading because it depends on definitions and assumptions that are not explicitly stated in the abstract.*

Author's reply:

We agree that the number could be misleading because the underlying definitions and assumptions were not defined in the abstract. To avoid confusion, we will remove the sentence from the abstract.

L 35: *Doesn't the condition "overlain by a more permeable layer" always exist due to the wetting front? Isn't the more general condition leading to SSF that conductivity plus sorptivity (sensu Philip) is smaller than rain intensity?*

Author's reply:

In the manuscript, we meant a structural (persistent) contrast in hydraulic permeability (i.e. saturated hydraulic conductivity) rather than a transient contrast associated with the passage of a wetting front. We agree that in a homogeneous medium the unsaturated hydraulic conductivity above the wetting front will be greater than below it due to the higher water content.

In sloping settings, a deflection from vertical to lateral flow generally requires a decrease in permeability with depth that limits vertical drainage. Given the presence of such an impeding horizon/interface, when the downward flux reaching this interface exceeds the percolation capacity across it, pressure increases and water will perch above it. The sloping setting will then result in a downslope-oriented hydraulic gradient and thus lateral flow.

SSF can occur for rainfall intensities below surface infiltration capacity, because the limiting process is the vertical drainage at depth rather than infiltration at the surface. In materials with low hydraulic conductivity (e.g. silt- and clay-rich sediments), perching and lateral flow can therefore occur even under low rainfall intensities when the downward flux exceeds the small percolation capacity of the impeding zone. Moreover, lateral subsurface flow can also occur when hydraulic conductivity is anisotropic, with horizontal conductivity exceeding vertical conductivity, because

vertical drainage is then impeded relative to lateral drainage. We will revise the text highlighting the *persistent* contrast in hydraulic conductivity and the percolation-excess mechanism.

L 56: *The event is not yet defined. Furthermore, this sentence suggests that the events of precipitation and SSF agree, whereas later, SSF lasts much longer than precipitation.*

Author's reply:

Here we refer to studies found in the literature that analyze the relationship between total precipitation amount of rainfall events and volumes of SSF events. Each study uses its own definition of an "event". Besides the different definition of "event", we also study the total depth of rainfall events; hence, the term "event" is used in a more general sense in the introduction. To avoid any confusion that rainfall events and SSF events necessarily coincide ("agree"), we will revise the text specifying which type of event we are referring to.

L 72: *ASI is not defined yet (only in the abstract).*

Author's reply:

The antecedent soil moisture index (ASI) was defined in L 64.

L 86: *I would argue that the complex relationships are a result of a complex definition of rain events being part of an SSL event.*

Author's reply:

In the introduction we are presenting the results of many authors, each of them having a different definition of "event". Despite the different approaches used to define the events, the relationships between the SSF response, rainfall characteristics and antecedent conditions appear complex.

L 121: Does 'concave' apply along or across the slope? To me, Fig. 1 suggests a convex shape across the slope (diverging water), not a concave slope. I may misinterpret Fig. 1.

Author's reply:

"Concave" referees to the curvature across the slope (perpendicular to the hillslope's main dip direction). In Fig. 1, the concave curvature promotes converging water flow. We will revise the text specifying that the concavity is across the slope.

L 125: *The unit should be mm/yr.*

Author's reply:

In the sentence, we stated that the mean annual precipitation is 836 mm. For more clarity, we will change the unit to mm/yr.

L 130 ff: *The soil material is either described as periglacial deposits or as colluvium. Both terms appear to be used interchangeably. In my understanding, both are different (Pleistocene vs. Holocene; existing along the entire slope vs. existing in toe-slope positions; coarse stones are possible vs. coarse stones are usually depleted compared to upslope soils).*

Author's reply:

We agree that the terminology can be ambiguous. In the manuscript, we used colluvium interchangeably with periglacial slope deposits, applying colluvium in the broad geomorphological sense (*sensu lato*) as a general term for unconsolidated slope deposits (Selby, 1995); although other sources specify that these deposits are primarily transported by gravity and/or sheetwash (American Geological Institute, 1976; Britannica Editors, 2009; Miller and Juilleret, 2020). We point out that the broad English definition does not align with the German term "Kolluvium", which has a more strict definition (Kleber, 2006). To avoid confusion, we will remove the term colluvium from the manuscript when referring to the slope deposits.

L 160: *I wonder why Robinia is defined to the species level while all other trees are only defined to the genus level.*

Author's reply:

The species-level identification for *Robinia* had no particular reason in this context. To ensure consistency, we will remove the species-level and will report *Robinia* at the genus level.

Table 1: *Depth is not clear. Is this the upper or the lower boundary or the midpoint? It may be worth noting whether the size classes follow the 2/50, the 2/63 or any other system.*

Author's reply:

The reported depth denotes the midpoint of the sampling interval in the vertical profile. Each sample represents material collected within ± 2.5 cm around the stated depth (e.g. 0.1 m BGL corresponds to 0.075–0.125 m BGL). This interval reflects the height of the cylindrical core sampler (5 cm).

Particle-size classes follow a 2 mm gravel–sand boundary and a 63 μm sand–silt boundary (i.e. 2/63 system):

gravel > 2 mm, sand 2–0.063 mm, silt 0.063–0.0059 mm, clay < 0.0059 mm.

The silt-clay cutoff of 5.9 μm was applied instead of the 2 μm cutoff to better align laser diffraction results with those obtained from traditional sedimentation analyses (Faé et al., 2019).

Laser diffraction particle size analysis, also referred to as laser granulometric measurement, denotes the determination of the size distribution of solid or liquid particles suspended in a liquid or gaseous medium by means of the deflection (diffraction) of laser light waves. The fundamentally different measurement principle compared to conventional sedimentation analyses results in consistently lower proportions in the clay fraction (≤ 2 μm) when using laser diffraction. The underlying causes are likely related to the predominantly non-spherical nature of these particle size spectra and their suspension and refraction behavior. This phenomenon is well documented in the scientific literature (e.g., Fae et al., 2019). To compensate for this methodological discrepancy or to align laser diffraction results with those obtained from traditional sedimentation analyses, a modification of the silt–clay boundary is often proposed. Fae et al. (2019) suggests a threshold of 5.9 μm , assigning all fractions below this value to the clay fraction.

We will revise the table to clarify that the reported depths are midpoints of the 5 cm sampling interval; we will also report the grain-size boundaries used for gravel, sand, silt, and clay.

L 218 ff: *This is a strange definition of an event that must yield a complex behaviour, as complained about on L. 86. Please note that for erosion events, also a minimum dry period of 6 h is until today a generally accepted requirement going back to Wischmeier (1959), who developed this from only 20 m long plots. These two identical definitions of a dry period separating runoff events are in conflict because the travel time of surface runoff is several orders of magnitude shorter than that of SSF.*

Author's reply:

We agree that defining a rainfall event using a minimum inter-event time (MIT) of 6 h can result in rainfall events that are shorter than the resulting SSF event. Our definition also implies that multiple rainfall events can be associated to a multi-peak SSF event. In any case, when the total volume, mean Q and MCA of the SSF event is analyzed versus rainfall depth, the cumulative depth of *all* rainfall events occurring during the SSF event is considered (e.g. if two rainfall events occur during the span of an SSF event the depths of the two rainfall events are added together, even if the two events are separated by a dry period much greater than 6 h). This ensures that all rainfall entering the system and leading to the generation of an SSF event is considered. Figure 6 panel b shows this cumulative rainfall.

When the volume, mean Q and MCA of an SSF event is analyzed versus rainfall intensity, the average of the rainfall intensity of all triggering rainfall events (i.e. events that produce a noticeable increase in SSF) associated with the respective SSF event is considered. For example, if two triggering rainfall events are associated with an SSF event, the I_5 used to study the relationship with the SSF volume is the average I_5 of the two rainfall events. With this approach the intensities of different rainfall pulses (rainfall events) occurring during the SSF event are considered. Figure 5 panel b shows the averaged intensities.

The other SSF metrics $Q_{\Delta\max}$, time to peak, rising rate, response time and lag time were used to characterize simple and sub- SSF events. These metrics were analyzed against rainfall characteristics of the triggering rainfall event associated with the simple/sub- SSF event. For example, the $Q_{\Delta\max}$ of a simple event (or subevent) is analyzed against the characteristics of the triggering rainfall event which induced the peak in the hydrograph.

Details on how rainfall and SSF events characteristics are defined and linked are provided in Section 2.3.6. We will further clarify how rainfall characteristics were derived and explicitly state when rainfall characteristics from a single rainfall event are used, and when characteristics derived from multiple rainfall events (occurring during the same SSF event) are considered.

L 264: *'Non-triggering' is also a strange category. It assumes that rain does not contribute to SSF when it does not trigger a noticeable peak in SSF. From water balance considerations, the non-triggering events during the falling limb of the hydrograph should also contribute to SSF. Or where else does the water go? Ignoring these events must increase the unexplained fraction of SSF.*

Author's reply:

We defined "non-triggering" events as events that occur *outside* of delineated SSF events (i.e. no SSF is detected during the occurrence of such a non-triggering event). These non-triggering events may contribute to sustaining the baseflow but not to SSF (as defined in our study). Rainfall events that occur during an SSF event, but that do not trigger a noticeable peak in SSF, are

defined as “ambiguous” (since in other circumstances they could have potentially produced a peak in SSF). These ambiguous events, which can for example occur during the falling limb of the hydrograph, likely contribute to SSF and are therefore included in the water balance. When the total SSF volume is analyzed, the cumulative rainfall depth of all rainfall events (triggering + ambiguous) occurring during the SSF event is considered; see also the reply to L 218 ff.

L 277: *Shouldn't it read "trenchflow rate"? The omission of the term 'rate' is particularly confusing, as rates are often abbreviated with lowercase letters, while sums are abbreviated with uppercase letters.*

Author's reply:

We agree that it should read “trenchflow rate”, since we refer to the discharge. We will revise the manuscript accordingly and consistently use “trenchflow rate” when referring to the discharge.

L 281: *This assumption is in conflict with the general understanding of hydrology, which suggests that soil moisture changes due to rain. Why not use $V_{tot}/(P_{tot}-VWC \text{ change})$? This would lead to a more realistic estimate of the contributing area.*

Author's reply:

Equation 1 ($MCA = V_{tot} / P_{tot}$) is not an assumption but follows the definition of minimum contributing area (stormflow ratio) after Dickinson and Whiteley (1970) (e.g. Fujieda et al., 1997; Zwartendijk et al., 2023). In our study, MCA is not intended to represent the actual contributing area; rather, it represents an absolute lower bound on the contributing area for a storm event, corresponding to the limiting case were 100% of rainfall on the contributing area becomes SSF. The true contributing area is likely larger than the MCA because we do not expect that all rainfall falling on the MCA is transformed into SSF.

Using $V_{tot} / (P_{tot} - \Delta VWC)$ would define a different, storage-corrected metric and would require additional assumptions about representativeness and conversion of VWC change to areal storage change.

We will revise the text clarifying the MCA definition in Section 2.3.6.

L 301: *Why should the average VWC be important and not the total? In your assessment, H_{tot} is constant, and hence the difference between average and total does not appear. However, when transferring your results to different catchments, the difference becomes important. I would even expect that a contrast leading to contrasting conductivities is important. A contrast would neither be captured by the average nor the total but by SD or similar metrics.*

Author's reply:

We acknowledge that the way the weighted average VWC was calculated may require some additional explanation. First, directly summing the VWC values measured by each soil moisture sensor (“total VWC”) would lead to unrepresentative values, as this would be equivalent to sum percentages or ratios (volume of water/total volume).

To obtain a meaningful “total water content”, the VWC values of each sensor must first be converted to an equivalent water depth, considering the layer thickness (H) that each sensor is assumed to represent/characterize. The sum of the resulting equivalent water depths can then be

calculated. This is essentially what is done in the numerator of Eq. 2. We multiplied the VWC measured by each sensor by the depth (H) that the sensor was assumed to represent to obtain an equivalent water depth, and then summed these water depths. Finally, to derive the depth-averaged VWC we divided the equivalent total water depth by the total depth (H_{tot}) represented by the sensors. Hence, the total equivalent water depth (water storage level) can be obtained by multiplying the depth-averaged VWC by H_{tot} .

As an example, if sensors are installed at depths of 10, 30 and 60 cm BGL the first sensor installed at 10 cm BGL is assumed to represent the conditions between 0–20 cm BGL ($H_1 = 20$), the second sensor 20–45 cm BGL ($H_2 = 25$), and the third sensor 45–75 cm BGL ($H_3 = 30$); hence, in this case, the total depth ($H_{tot} = H_1 + H_2 + H_3$) that our sensors characterize would be 75 cm. The total depth (H_{tot}) is constant because the installation depths of the sensor, and the depths (H) that these sensors are assumed to represent, do not change.

The weighted average VWC (Eq. 2) was chosen instead of the total equivalent water depth to make our result somewhat more comparable to other studies, as the total equivalent water depth depends directly on the total depth represented by the sensors. For example, given the same VWC at each sensor, a soil moisture profile characterizing the soil moisture up to 1 m depth will yield a smaller total equivalent water depth than a profile characterizing soil moisture up to 10 m depth. By dividing the total equivalent water depth by the total depth (H_{tot}), we account for the considered H_{tot} and make our results more comparable with studies that use different soil moisture profile setups.

The vertical contrast in VWC could be quantified using additional metrics. For the sake of simplicity, and because our focus is on overall antecedent wetness, we consider only the depth-weighted mean VWC.

We will revise the text specifying how to obtain the total equivalent water depth (water storage level).

L 313: *Why is D used here and not H_{tot} ? Is there a difference between the two? Explain.*

Author's reply:

We agree that the way ASI (Eq. 3) was linked to Eq. 2 was confusing. Therefore, we will reformulate Eq. 3 so that it is independent of Eq. 2; the new formulation will yield the same ASI values as the previous one.

The ASI represents the antecedent equivalent water depth (water storage level).

In the previous formulation, a shallow average VWC ($\bar{\theta}_s$) was computed using Eq. 2 but considering only the three upper soil moisture sensors (10, 30, 60 cm BGL); hence, the corresponding total represented depth (H_{tot}) was 75 cm (see example in the reply to L 301). To convert $\bar{\theta}_s$ into an equivalent water depth (water storage level), the $\bar{\theta}_s$ was multiplied by the total represented depth (D) of 750 mm. In the case where only the three upper sensors are considered, H_{tot} equals D (75 cm = 750 mm).

In the new formulation, ASI will be computed in a more direct way. The water content measured by the sensors at depths of 10, 30 and 60 cm BGL will be directly multiplied by the layer thickness (in mm) that the sensors are assumed to represent (i.e. 200 mm, 250 mm and 300 mm, respectively), yielding equivalent water depths. Summing these equivalent water depths will result in a total water depth (water storage level) representative of the shallow soil profile (0–75 cm BGL).

We acknowledge that ASI could have also been calculated using all sensors. The decision to only use the upper three sensors is arbitrary, and gives more weight to the shallow (0–75 cm BGL) soil moisture conditions.

L 318 + 328: *Shouldn't ASI minus PWP (permanent wilting point) be used instead of arbitrarily adjusting ASI? From a theoretical perspective, ASI should not be the correct parameter. Furthermore, I do not see a justification for a fractional adjustment factor to correct this.*

Author's reply:

We use ASI as an indicator of antecedent wetness conditions of the soil. Subtracting the PWP from ASI would reduce the absolute values; however, the wet–dry dynamics would remain the same.

In order to study the combined influence of total rainfall (P_{tot}) and antecedent wetness conditions, we transformed the antecedent volumetric water content into an equivalent water depth (storage level) (ASI), which as a depth, can be added to the total rainfall depth (i.e. $P_{tot} + ASI$).

Since ASI is strongly dependent on the thickness (depth) of the considered soil profile (750 mm in our case), ASI can become large relative to P_{tot} when deep profiles are considered, so antecedent conditions may receive disproportionate weight in the composite variable $P_{tot} + ASI$. When $P_{tot} + ASI$ is analyzed against an SSF metric (X), a disproportionately large (dominant) ASI implies that variability in the composite variable is driven primarily by ASI rather than P_{tot} . In situations where ASI is extremely disproportionate, the relationship (and correlation) between X and $P_{tot} + ASI$ will tend to resemble that between X and ASI alone; accordingly, as the thickness of the profile used to compute ASI increases, the correlation of X with $P_{tot} + ASI$ approaches the correlation of X with ASI. By scaling ASI in the composite variable $P_{tot} + ASI$, the possible disproportionate weight of ASI can be accounted for.

Furthermore, the weighting factor indicates how important ASI is within the composite variable with respect to the analyzed metric. For example, if the weighting factor is very small, it indicates that the correlation between the SSF metric and P_{tot} is improved by considering only a marginal contribution of the antecedent wetness conditions. Conversely, a large weighting factor indicates that the antecedent wetness conditions play a more important role in improving the correlation.

In conclusion, in the composite variable $P_{tot} + ASI_{adj}$, we scaled ASI to quantify its contribution when analyzing relationships between the composite variable and SSF metrics.

We will revise Section 2.3.7 and 2.3.8 to clarify why ASI was adjusted and how this was used in the analysis. We will also update Figure 8, as we realized that the heatmap contained some wrong (older) correlation values, that were obtained using another method. The updated values do not affect the conclusions and discussion points, which remain the same.

L 334: *This is clearly a spurious correlation that cannot be interpreted because an identical variable is used on both sides of the equation (see Pearson, 1897, or specifically in hydrological research: Kenney, 1982). For some specific cases, a correction exists (Kenney, 1982; Kanaroglou, 1996).*

Author's reply:

We agree that correlating V/P to P , as done in Eq. 4 ($MCA=V/P = a \cdot P + b$) would lead to a spurious correlation of the special case where a ratio ($MCA=V/P$) is being correlated with its own denominator (P). Therefore, we will fit Eq. 5 directly.

Eq. 5 is used to test whether the observed $V-P$ relationship is consistent with an approximately linear increase of MCA with P , while avoiding direct regression of the ratio MCA on its denominator. Since MCA is (indirectly) linked to the area that contributes to SSF generation, the good fit of Eq. 5 will support the interpretation that the contributing area likely increases with P .

We will fit Eq. 5 using the same P range for both trenches. Table 3 will be updated to report the revised coefficients a and b (with confidence intervals) and r^2 of the fitted Eq. 5. Figure 14 will be updated to show the fit of Eq. 5 and the linear fit, both including confidence intervals.

Fig. 8: Many of these correlations are at least partly spurious because both sides of the equation contain common elements. E.g., the calculation of P_{tot} depends on the intensities. The significance, as denoted, is then meaningless. This makes interpretation of the figure impossible because the degree to which common numbers were used on both sides differs (I_5 contributes considerably less to P_{tot} than I_{60} ; hence, the difference between correlation coefficients may only be due to a different fraction of the spurious component). Furthermore, the table is clearly multiple testing, which would require an adjustment to maintain the alpha level. Less adjustment would be necessary if the authors did not weaken the statistical power of their analysis by testing parameters like I_5 or I_{10} , for which I can see no hydrological argument as to why these parameters should generally drive SSF . Also, no confidence intervals are given. Hence, it cannot be decided whether one correlation coefficient is larger than the other. The text wrongly assumes that the larger value is always better.

Author's reply:

We agree that the correlations between $MCA-V$, $MCA-P+ASI$, and $MCA-P+ASl_{adj}$ are susceptible to spurious self-correlation (see also reply to comment L334). The same applies to the correlations between $Q\Delta_{max}-Q_i$, rising rate- Q_i , and rising rate- $Q\Delta_{max}$. Therefore, we will remove the correlation coefficients for these correlations from Fig. 8 and delete the text dealing with those correlation coefficients.

The remaining correlations are not susceptible to spurious self-correlation (*sensu stricto*; Kenney, 1982). We agree, however, that some strong correlations are expected. For example, SSF events with larger volumes are more likely to be characterized by higher Q_{max} , so the high correlation between these two metrics is not surprising and was not given much weight in the study. Similarly, rainfall amount and intensity are partly related with each other; however large rainfall depths can result from long, low-intensity events (stratiform precipitation) as well as short, high-intensity events (convective thunderstorms). Parameters like I_5 or I_{30} were analyzed to explore if and how they influence SSF . Flow through well-connected macropore networks (pipeflow) can be an important contributor to SSF and was shown to be influenced by the maximum rainfall intensity (Uchida et al., 2005).

We agree that given the large number of tested correlations the α -level needs to be corrected. To keep the overall α -level at 0.05, we will follow the strict Bonferroni's correction (Curtin and Schulz, 1998). The per-test significance level will be obtained by dividing 0.05 by the total number of tested correlations (97 for each trench). Correlations will be considered statistically significant if the p-value is smaller than 0.0005.

The missing confidence intervals of the Spearman rank correlations will be quantified using a bootstrap approach (Haukoos and Lewis, 2005): for each pair of metrics, the dataset is resampled with replacement (i.e. data points were randomly selected from the original dataset and could be selected multiple times) to generate 1000 bootstrap datasets (recommended ≥ 250 datasets; Efron and Tibshirani, 1986). The Spearman's rank correlation coefficient is then computed for each resample, and a 95% confidence interval is obtained using the percentile method as the 2.5 and 97.5 percentiles of the bootstrap distribution of the rank correlation coefficient. The resulting confidence intervals will be presented in a Table in a (new) appendix C.

Moreover, to evaluate whether the correlation between Y and two alternative metrics (X1 and X2) improved, we will compute the difference between the corresponding Spearman rank correlations, $\Delta r_s = r_{s2} - r_{s1}$, where $r_{s1} = r_s(Y, X1)$ and $r_{s2} = r_s(Y, X2)$. Uncertainty in Δr_s will be quantified using a bootstrap approach (Wilcox, 2016): the dataset (Y, X1, X2) is resampled with replacement to generate 1000 bootstrap datasets; for each resample, Δr_s is computed and stored. A 95% confidence interval for Δr_s is then obtained using the percentile method. An improvement is considered supported when the 95% confidence interval for Δr_s (CI Δr_s) is greater than zero. This analysis will be performed only for selected metric pairs (highlighted in the results).

The corresponding statements in the Results/Discussion will be modified to reflect the results of these additional analyses.

L 440: *Does the area really change, or does this only reflect the shortcomings of Eqn 1?*

Author's reply:

The MCA does change. Eq. 1 does not yield the actual (active) contributing area, but the minimum contributing area (MCA), i.e., a lower-bound estimate. To avoid confusion, we will revise the passage to clearly distinguish MCA from the actual contributing are, and avoid the use of the ambiguous term "area" for the MCA.

Table 3: *r^2 and the parameters of Eqn 4 are spurious. How can the r^2 of Eqn 5 be calculated when the text says that Eqn 5 was obtained by rearranging Eqn 4? Is r^2 then the Nash-Sutcliffe efficiency? Calculation Eqn 5 directly would be advantageous because it avoids the spurious step.*

Author's reply:

We agree that Eq. 4 is affected by spurious self-correlation. We therefore revisited the analysis fitting directly Eq. 5, avoiding the spurious step (see also the reply to L 334). The reported r^2 for Eq. 5 was calculated as the squared Pearson correlation between V-observed and V-simulated (obtained from Eq. 5). In the future we plan to directly fit Eq. 5.

L 466: *No confidence intervals are given to support this statement.*

Author's reply:

We agree that the confidence intervals were missing. We will report the computed 95% confidence intervals for the fitted parameters in Table 3. In addition, we will update Figure 14 to include the corresponding confidence bands.

L 478: *same remark as above*

Author's reply:

In order to support stated improvement in correlation we will calculate the confidence intervals (as described in the reply to Fig. 8; specifically CIΔrs) and will report them in the text.

L 508-509: *Where is this shown? A test like the Hotelling test must be applied to support such a statement. I am unsure whether the Hotelling test can be applied to Spearman's correlation or if an analogous test exists. If not, Spearman appears unsuitable for the intended task.*

Author's reply:

The statement "I60 appeared to be most suited to explain the differences in rising rates among events with similar Ptot ..." was not supported by a statistical test. To avoid misunderstandings, we will remove this statement from the manuscript.

Fig. 19: *Again, confidence intervals would be required to support the interpretation made in the text.*

Author's reply:

We agree. We will add uncertainty information (in a new table in the appendix) reporting bootstrap confidence intervals for all seasonal trimmed means (percentile bootstrap, 1000 resamples). We will also revise the corresponding text to explicitly frame the seasonal results as descriptive patterns and avoid overstating differences where confidence intervals overlap. In particular, we will replace causal formulations (e.g., "explained by") with more cautious, process-consistent phrasing ("consistent with") and will add statements highlighting when uncertainty is large.

L 738: *This statement is only true for SSF volume as defined in this study. Whether this is generally true remains unknown.*

Author's reply:

We agree. We will revise the sentence to clarify that this inference is specific to SSF volume as defined in our study.

L 742: *The contributing area was unfortunately not quantified. Only a minimum area was quantified. By omitting the adjective, a wrong impression is given (not to mention that the adjective is wrong as well, because the equation makes an unrealistic assumption).*

Author's reply:

We agree. The contributing area was not quantified; only the minimum contributing area (MCA) was quantified. The text will be revised accordingly to avoid implying otherwise.

MCA is a lower bound derived from the observed SSF event volume and precipitation under the limiting assumption that 100% of rainfall over that area is converted to SSF. This assumption is not realistic because interception, soil-moisture storage, and vertical leakage reduce the fraction of rainfall that becomes SSF; therefore, the true contributing area is expected to be larger than MCA. We will refer to MCA (rather than contributing area) when reporting results, and will interpret increases in MCA only as indicative of a likely expansion of the contributing area, without claiming that the contributing area itself was quantified.

L 764: *R. C.?*

Author's reply:

Our citation entry was wrong. "R.C." stands for the first name of the author. We will revise the citation in the manuscript. Furthermore, we will correct a typo in equation A2: the exponent of the last term is 3 (not 2). Our results are not affected, as the calculations were performed using the correct formula.

L 767: *and beta.*

Author's reply:

We agree. The text will be revised to include beta.

Table A1: *The assumed porosities for the two lowest depths of T3-B3 are very unlikely. For the measured bulk density, this would only be possible if the density of the solids were 3.2 kg/L, which is considerably above that of the main minerals in soils. Something else must be wrong in the calculations if they require unreasonable porosity to yield reasonable results. In contrast, the measured porosities are in line with the measured BDs.*

Author's reply:

We thank the reviewer for the in-depth analysis of our data. We agree that, given the measured bulk density (1.9 g/cm³) and density of the solids (2.7 g/cm³), the assumed porosities are not possible for the two deepest T3-B3 samples if they are interpreted as representative of the sampling location.

However, the soil moisture sensors were not installed at exactly the same location where the soil samples were collected. In addition, the high fraction of large grains (77% gravel) made the collection of "undisturbed" soil cores (later used for the bulk density and porosity analysis) with a 250 cm³ cylindrical sampler difficult. To minimize soil disturbance areas with visibly "lower" gravel content, and thus probably lower porosity, were selected to carry out the "undisturbed" soil sampling. Moreover, the high gravel content made the installation of the soil moisture sensor

technically challenging, such that an installation-related disturbance, leading to a small increase in porosity of the volume surrounding the sensor, cannot be excluded. Combined with the small-scale heterogeneity that can characterize unconsolidated sediments, these factors provide a plausible explanation for the higher (required) porosity values.

The used porosity values are consistent with the maximum volumetric water content (VWC) calculated using the widely applied Topp equation (Topp et al., 1980), under the assumption that during saturated or near-saturated conditions the volumetric water content approximates porosity. The Topp-based VWC values were obtained directly from the SMT100 (Truebner) sensor output (internal calculation). Without the porosity correction the VWC obtained with the Topp and CRIM equation would have yield max VWC values greater than the lab-measured porosity. To overcome this, the porosity was adjusted using the max. VWC observed VWC (Topp) as a guideline and also ensuring that the max. CRIM-calculated VWC approximated the adjusted porosity.

As acknowledged in the manuscript, absolute VWC estimates derived from the CRIM approach can deviate from the “true” VWC due to uncertainties in the permittivity of the solids (K_s) and porosity. Nevertheless, we used the CRIM approach primarily to reduce temperature sensitivity in the permittivity–VWC conversion. In this study, VWC was solely used as an indicator for wet/dry conditions (relative temporal changes), such that VWC dynamics are more relevant than the absolute VWC values.

We will revise the manuscript to clarify why the porosity correction was necessary and what criteria were used to carry it out. Moreover, we will explain why porosities derived from laboratory samples may differ from the porosities used for the CRIM-based calculations (spatial offset between sampling and sensor locations, small-scale heterogeneity, and potential installation-related disturbance in gravel-rich material).

References

American Geological Institute: Dictionary of geological terms, Rev. ed., Anchor Books, 472 pp., 1976.

Britannica Editors: Encyclopedia Britannica, 2009.

Curtin, F. and Schulz, P.: Multiple correlations and bonferroni’s correction, *Biol. Psychiatry*, 44, 775–777, [https://doi.org/10.1016/S0006-3223\(98\)00043-2](https://doi.org/10.1016/S0006-3223(98)00043-2), 1998.

Dickinson, W. T. and Whiteley, H.: Watershed areas contributing to runoff, *IAHS Publ*, 96, 12–26, 1970.

Dunkerley, D.: Identifying individual rain events from pluviograph records: a review with analysis of data from an Australian dryland site, *Hydrol. Process.*, 22, 5024–5036, <https://doi.org/10.1002/hyp.7122>, 2008.

Faé, G. S., Montes, F., Bazilevskaya, E., Añó, R. M., and Kemanian, A. R.: Making Soil Particle Size Analysis by Laser Diffraction Compatible with Standard Soil Texture Determination Methods, *Soil Sci. Soc. Am. J.*, 83, 1244–1252, <https://doi.org/10.2136/sssaj2018.10.0385>, 2019.

Fujieda, M., Kudoh, T., De Cicco, V., and De Calvarcho, J. L.: Hydrological processes at two subtropical forest catchments: the Serra do Mar, São Paulo, Brazil, *J. Hydrol.*, 196, 26–46, [https://doi.org/10.1016/S0022-1694\(97\)00015-2](https://doi.org/10.1016/S0022-1694(97)00015-2), 1997.

Kenney, B. C.: Beware of spurious self-correlations!, *Water Resour. Res.*, 18, 1041–1048, <https://doi.org/10.1029/WR018i004p01041>, 1982.

Kleber, A.: “Kolluvium” does not equal “colluvium,” *Z. Für Geomorphol.*, 50, 541–542, <https://doi.org/10.1127/zfg/50/2006/541>, 2006.

Miller, B. A. and Juilleret, J.: The colluvium and alluvium problem: Historical review and current state of definitions, *Earth-Sci. Rev.*, 209, 103316, <https://doi.org/10.1016/j.earscirev.2020.103316>, 2020.

Selby, M. J.: *Earths’ Changing Surface*, Oxford University Press, 607 pp., 1995.

Topp, G. C., Davis, J. L., and Annan, A. P.: Electromagnetic determination of soil water content: Measurements in coaxial transmission lines, *Water Resour. Res.*, 16, 574–582, <https://doi.org/10.1029/WR016i003p00574>, 1980.

Uchida, T., Tromp-van Meerveld, I., and McDonnell, J. J.: The role of lateral pipe flow in hillslope runoff response: an intercomparison of non-linear hillslope response, *J. Hydrol.*, 311, 117–133, <https://doi.org/10.1016/j.jhydrol.2005.01.012>, 2005.

Wilcox, R. R.: Comparing dependent robust correlations, *Br. J. Math. Stat. Psychol.*, 69, 215–224, <https://doi.org/10.1111/bmsp.12069>, 2016.

Zwartendijk, B. W., Van Meerveld, H. J., Teuling, A. J., Ghimire, C. P., and Bruijnzeel, L. A.: Rainfall-runoff responses and hillslope moisture thresholds for an upland tropical catchment in Eastern Madagascar subject to long-term slash-and-burn practices, *Hydrol. Process.*, 37, e14937, <https://doi.org/10.1002/hyp.14937>, 2023.