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 two reviewers for the careful reading of our manuscript and for the useful comments.

Reviewer 1

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. 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 that point in the text. To address this, we added the analyzed rainfall characteristics before L 18-19.

Changes in manuscript:

We added the analysed 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 removed the sentence from the abstract.

Changes in manuscript:

The sentence containing the number in question was removed 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.

Changes in manuscript:

We have revised 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". Hence, the term "event" is used in a more general sense in the introduction. To avoid any confusion, we revised the text and removed the acronyms Ptot and Vtot from the introduction.

Changes in manuscript:

The text was revised and the acronyms Ptot and Vtot removed from the introduction.

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.

Changes in manuscript:

The text was not revised.

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.

Changes in manuscript:

The text was not revised.

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). The concave curvature shown in Fig. 1 for trench 3 promotes converging water flow.

Changes in manuscript:

The text was revised, 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 changed the unit to mm/yr.

Changes in manuscript:

The text was revised changing 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 have removed the term colluvium from the manuscript.

Changes in manuscript:

We revised the text and removed the term colluvium.

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 removed the species-level and now report *Robinia* at the genus level.

Changes in manuscript:

We revised the text by removing "*pseudoacacia*" and now 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.

Changes in manuscript:

We revised the table to clarify that the reported depths are midpoints of the 5 cm sampling interval; we 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 are 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 further clarified how rainfall characteristics were derived and explicitly stated 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.

Changes in manuscript:

We revised Section 2.3.6 to clearly distinguish when characteristics derived from a single rainfall event, and when characteristics derived from multiple aggregated events (i.e. rainfall events occurring during an 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 ra4infall events (triggering + ambiguous) occurring during the SSF event is considered; see also the reply to L 218 ff.

Changes in manuscript:

The text was not revised.

***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.

Changes in manuscript:

We revised the manuscript accordingly and now 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 Vtot/(Ptot-VWC 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 a definition based on the minimum contributing area (stormflow ratio) concept 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 of a 100% runoff coefficient (i.e. all rainfall on the contributing area becomes stormflow). The true contributing area is therefore expected to be larger, or at least as large as the MCA.

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

Changes in manuscript:

We revised 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, Htot 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 leasing 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.

Changes in manuscript:

The text was revised specifying how to obtain the total equivalent water depth (water storage level).

L 313*: Why is D used here and not Htot? 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, Eq. 3 was reformulated; the new formulation yields 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 is computed in a more direct way. The water content measured by the sensors at depths of 10, 30 and 60 cm BGL are 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 results 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.

Changes in manuscript:

We revised the manuscript reformulating Eq. 3 and clarifying how ASI is calculated.

***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.

Changes in manuscript:

We revised Section 2.3.7 and 2.3.8 to clarify why ASI was adjusted and how this was used in the analysis.

We also updated 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*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 now 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 supports the interpretation that the contributing area likely increases with P.

Changes in manuscript:

We revised the manuscript accordingly: Eq. 5 was fitted instead of Eq. 4 (Eq. 4 was removed) using the same P range for both trenches. Table 3 was updated to report the revised coefficients *a* and b (with confidence intervals) and r^2 of the fitted Eq. 5. Figure 14 was updated and now shows the fit of Eq. 5 and the linear fit, both including confidence intervals. The text describing the relationship between Eq. 4 and Eq. 5 and the interpretation in terms of contributing area was revised.

***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 Ptot 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 (I5 contributes considerably less to Ptot than I60; 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 I5 or I10, 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–P, MCA–V, MCA–P+ASI, and MCA–P+ASIadj are susceptible to spurious self-correlation (see also reply to comment L334). The same applies to the correlations between QΔmax–Qi, rising rate–Qi, and rising rate–QΔmax. Therefore, we removed the correlation coefficients for these correlations from Fig. 8 and deleted 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 Qmax, 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 I5 or I30 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 followed the strict Bonferroni's correction (Curtin and Schulz, 1998). The per-test significance level was obtained by dividing 0.05 by the total number of tested correlations (97 for each trench). Correlations were considered statistically significant when the p-value was smaller than 0.0005.

The missing confidence intervals of the Spearman rank correlations were quantified using a bootstrap approach (Haukoos and Lewis, 2005). For each pair of metrics, the dataset was 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). Spearman's rank correlation coefficient was then computed for each resample, and a 95% confidence interval was 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 are presented in a Table in the (new) appendix D.

Moreover, to evaluate whether the correlation between Y and two alternative metrics (X1 and X2) improved, we computed the difference between the corresponding Spearman rank correlations, $\Delta rs = rs2 - rs1$, where $rs1 = rs(Y, X1)$ and $rs2 = rs(Y, X2)$. Uncertainty in Δrs was quantified using a bootstrap approach (Wilcox, 2016). The dataset (Y, X1, X2) was resampled with replacement to generate 1000 bootstrap datasets; for each resample, Δrs was computed and stored. A 95% confidence interval for Δrs was then obtained using the percentile method. An improvement was considered supported when the 95% confidence interval for Δrs ($CI\Delta rs$) was greater than zero. This analysis was performed only for selected metric pairs (highlighted in the results).

Changes in manuscript:

We revised the correlation analysis to address potential spurious self-correlation and multiple-testing issues. Specifically, we removed correlation coefficients affected by spurious self-correlation (defined in the 1st paragraph above) from Fig. 8 and deleted the text dealing with those correlation coefficients. We corrected the α-level applying Bonferroni's correction and updated the significance classification of the resulting coefficients in Fig. 8. We also calculated the confidence intervals of the Spearman rank correlations using a bootstrap approach, and report them in a Table in the (new) appendix D. Finally, for selected metric triplets sharing a common variable, we assessed whether correlations differed between alternative predictors by bootstrapping the difference in Spearman correlations (Δrs) and computing 95% confidence intervals. The corresponding statements in the Results/Discussion were updated to reflect these revised 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 revised the passage to clearly distinguish MCA from the actual contributing area.

Changes in manuscript:

We revised the passage avoiding 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*). Previously, the reported r^2 for Eq. 5 was calculated as the squared Pearson correlation between V-observed and V-simulated (obtained from Eq. 5).

Changes in manuscript:

The Table 3 was revised by removing Eq. 4 and reporting the fitted coefficients (with confidence intervals) and r^2 for 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 now report the computed 95% confidence intervals for the fitted parameters in Table 3. In addition, Figure 14 has been updated (now showing the fit of Eq. 5) and includes the corresponding confidence bands.

Changes in manuscript:

Table 3 and Figure 14 were revised to include 95% confidence intervals to support the statement.

L 478: *same remark as above*

Author's reply:

In order to support stated improvement in correlation we calculated the confidence intervals (as described in the reply to Fig. 8); specifically $CI_{\Delta rs}$ and reported them in the text.

Changes in manuscript:

We added the confidence intervals to support the statement on the improved correlation.

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 removed this statement from the manuscript.

Changes in manuscript:

We removed the statement from the text.

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

Author's reply:

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

Changes in manuscript:

We computed and reported 95% confidence intervals for all seasonal trimmed means using a percentile bootstrap (1000 resamples) (Appendix E) and updated the corresponding text to explicitly acknowledge uncertainty and overlapping intervals.

***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.

Changes in manuscript:

We revised 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 was 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 now refer to MCA (rather than contributing area) when reporting results, and 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 revised the citation in the manuscript. Furthermore, we corrected 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.

Changes in manuscript:

We revised the citation and Eq. A2 in the manuscript.

***L 767**: and beta.*

Author's reply:

We agree.

Changes in manuscript:

The text was 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 soilds (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.

Changes in manuscript:

We revised the manuscript to clarify why the porosity correction was necessary and what criteria were used to carry it out. Moreover, we 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).

Reviewer 2

General Comments

From a general perspective, the study would benefit from a clear perceptual model of hillslope hydrology underlying the analysis. The correlation analysis presented here probably does not aim at investigating the general causality of rainfall and runoff, which is not in question, but quantifying their possible relationships under different conditions. An exploratory data analysis alone, however, is limited in its explanatory power, if not combined with a perceptual model that enables testable hypotheses. The manuscript does not clearly describe on which perceptual model the analysis is based. For example, the authors derive a quadratic increase of runoff with rainfall, which is hardly physically interpretable over the entire range of possible flow conditions. A large part of the analysis uses the concept of variable contributing area/MCA, but also the concept of varying subsurface runoff coefficients is used, which implies a fixed-size catchment area. It would be interesting to see whether the trench data allow determining which concept is more realistic. In addition, an estimate of the catchment area based on topographic analysis would be extremely helpful to compare rainfall and runoff volumes in millimeters, even if used only as an upper bound.

Author's reply:

Our analysis is primarily an exploratory data analysis where relationships between SSF metrics, rainfall and initial conditions are explored. The underlying qualitative, process-based description of SSF generation (i.e., our perceptual model of hillslope hydrology) is provided in the Introduction, based on findings from prior SSF hillslope studies. We interpret our results in the context of these published perceptual models and use them to guide the discussion of plausible mechanisms underlying the correlations.

We agree that an unconstrained quadratic increase of runoff with rainfall is not physically interpretable outside the observed range (see the more extensive reply to comment "L 333ff"). We now state this limitation explicitly in the text.

We agree that the terms "contributing area", and "subsurface runoff coefficient" (Csub) were poorly defined and could have been source of confusion. In our study, contributing area, refers to the area providing water to the trench over an SSF event. The Csub, *was* intended in this study as the subsurface runoff coefficient of the contributing area, and not as the classical catchment-scale runoff coefficient calculated considering a fixed catchment area. To avoid confusion we now define Csub referenced to the SSF catchment area (i.e. Csub = Volume in mm / rainfall in mm, where Volume in mm = Volume in liters/catchment area in m^2).

The minimum contributing area (MCA) represents the minimum area required to generate the observed SSF volume assuming that 100% of rainfall on that area becomes SSF (it is defined as Volume in liters / rainfall in mm). 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. Hence, an increasing/decreasing MCA is used as an indicator that the contributing area increases/decreases. The subsurface runoff coefficient (Csub) is directly related to MCA and can be calculated as MCA/ SSF catchment area, provided that the SSF catchment area is known.

We agree that providing the catchment area derived from topographic analysis would be beneficial, as it could be used instead of the unknown SSF catchment area, enabling an estimate of the subsurface runoff coefficient. However, it must be made clear that the catchment area of the subsurface stormflow may be different from the topographically derived one, as it depends on the hydrogeological setting, which is only partly known. We included the topographic catchment area in our manuscript clarifying that it may differ from the "real" SSF catchment area.

Changes in manuscript:

We revised the text discussing the validity and implications of the quadratic increase. Furthermore, we now clearly define the terms "contributing area", MCA and subsurface runoff coefficient, and added the topographically derived catchment area.

Another point is that the analysis is based on field data, yet measurement uncertainty is not addressed, and there is no discussion of how it might affect the results. The authors select data for the analysis, but the criteria for this selection are not made fully clear (only two of three trenches are used, "outliers" are removed without further explanation, selective data ranges in figures).

Author's reply:

We agree that measurement uncertainty was not addressed. Therefore, we added the uncertainty for the flowrate measurements carried out with the tipping buckets. We also revised the manuscript discussing how this uncertainty might affect the results.

Regarding the selected data for the analysis:

- as stated in L 118 only two trenches were analyzed as they had a more comprehensive dataset. We agree that in the future an analysis of more trenches would be beneficial.

- we filtered out outliers (L 201) in the trenchflow timeseries. The outliers were caused by sporadic malfunctions of the tipping-bucket logging system, where the number of recorded tips within a time interval was markedly lower than in the preceding and subsequent interval, producing isolated, unrealistic drops in discharge (e.g. from 30 l/h to 1 l/h and back to 30 l/h). These hydrologically non-explainable, isolated drops in discharge were manually removed during the visual analysis of the trenchflow timeseries.

- the only figures with selective data ranges are Fig. 11, 12 and 14, as is described in the text and/or included in the figure caption. In Figure 11 and 12 we did not include the largest events to focus on the dynamics of the small to medium events; plotting the full data range would make the interpretation of the low-medium events dynamics more difficult in the lin-lin plot (left plot in Fig. 11 and 12). The selective data range in Fig. 14 derives from the ranges selected to apply Eq. 4 and Eq. 5. As suggested in comment "L 333ff" we now only fit Eq. 5 and use the same range of rainfall (up to P = 50 mm) for both trenches (see details in the reply to L 333ff).

Changes in manuscript:

We revised the manuscript addressing measurement uncertainty and possible implications. Regarding the outliers we added the cause as well as how and why those were removed. Fig. 14 has been also revised (see also reply to L 333ff).

Major Comments

***L 117**: Only two of three trenches were used in the analysis. Please provide more reasoning why T2 was excluded. Even if not as comprehensive, it could provide further insight to the matter, and with your automatic analysis, it should not be too much work to analyze the data as well.*

Author's reply:

We thank the reviewer for this suggestion. The second trench was excluded because its record was less comprehensive; specifically, more frequent gaps in the trenchflow timeseries and a limited number of usable events reduce the robustness of the event-based analysis and increase uncertainty in relationships involving derived event metrics. However, we agree that the analysis of additional trenches would be beneficial and will be addressed in future work.

Changes in manuscript:

The manuscript was not modified in this regard.

***L 301 to 306**: This approach gives much more weight to sensors located far from others, and can exceed the actual soil volume that is sensed by the device. In your case, the sensor at 150 cm depth has more weight than the two sensors at the top combined. The soil moisture in the upper soil layer would typically be more dynamic, and perhaps also more influential on hydrological processes. Including soil moisture from a lower layer with an even higher weight might thus dampen the contrast in VWC across events. Integrating over an interpolated profile might be an alternative.*

Author's reply:

Indeed, Eq. 2 gives more weight to the more widely spaced sensors, effectively assuming that these sensors represent a thicker soil layer. Eq. 2 was applied to obtain a single profile-averaged VWC value that approximates the mean soil moisture of the instrumented profile. We fully agree that the volume sensed by the device is smaller than the volume implicitly represented by the weighting. However, given the lack of more closely spaced sensors in the deeper part of the soil profile, some assumptions (e.g. Eq. 2) are required to aggregate the discrete observations into a profile-scale metric. Integrating over an interpolated profile might indeed be a valid alternative approach. We also agree that the upper soil layer is typically more dynamic, and that consequently, including deeper sensors with relatively high weights will likely dampen the contrast in VWC across events. However, despite the greater weight of the deeper sensors, the distinction between wet and dry antecedent soil moisture conditions (represented by VWCi) is clear, and the observed VWCi ranges are substantial (T1: 23.3–34.0%; T3: 14.9–35.6%).

Changes in manuscript:

The manuscript was not modified with regard to any possible alternatives to weight sensors (e.g. the mentioned interpolation). However, the text was revised to clarify that the weights are applied to reflect the layer thicknesses attributed to the sensors.

***L 314f**: While in Eq. 2 sensors from six depths are averaged, yet here only three are used? Furthermore, the weighted averaging described above considered the midpoints between sensors. Following this approach, the total depth would be in the middle of 60 and 150 cm, and not 75 cm. This is not consistent.*

Author's reply:

Yes, to calculate the ASI in Eq. 3 only the upper three sensors are used. The decision to use only the upper sensors was made to focus on the influence of the shallow soil moisture conditions. We agree that, following the weighted approach described above, the midpoint between 60 and 150 cm would have to be the total depth. However, when calculating the ASI the "150 cm sensor" was not considered, so that the lowermost considered sensor is the 60 cm one, and for the lowermost (considered) sensor the distance from the adjacent (considered) sensor is used for the calculations. However, this was not properly described and rightfully led to confusion. Therefore, we revised the text decoupling Eq. 2 from Eq. 3 by reformulating Eq. 3 (the new formulation yields the same ASI values as the previous one). In the new formulation, ASI is computed in a more direct way. The water content measured by the sensors at depths of 10, 30 and 60 cm BGL are directly multiplied by the layer thickness (in mm) being attributed to each sensor (i.e. 200 mm, 250 mm and 300 mm, respectively). The total water equivalent represents therefore the shallow soil part from 0 to 75 cm BGL.

Changes in manuscript:

We revised the manuscript reformulating Eq. 3 and clarifying how ASI is calculated.

***L 317**: "if the ASI is too large, the effect of the initial conditions can be over-represented" – how is "too large" defined? The goal of this study is to investigate the effect of initial conditions, so how can you know that it is "over-represented"? This would only be possible if you knew about the relationship beforehand.*

Author's reply:

We agree that that we cannot know a priori whether the ASI is "over-represented" in the composite variable. In our study, ASI is an integrated water-storage proxy (units of mm water over a 0.75 m depth), and therefore its absolute values and event-to-event range can be hundreds of millimeters, whereas rainfall (P) is typically on the order of tens of millimeters. Here, 'too large' refers to cases where ASI's magnitude/variability is much larger than P, such that variability in P+ASI is dominated by ASI and the composite variable increasingly behaves like ASI alone.

Hence, to explore whether the correlation between an SSF metric and P+ASI would improve by reducing the contribution (weight) of ASI, we tested P + w·ASI with w ranging from 0 to 1. In nearly all cases (Figure 8), the weight to achieve the best (Spearman) correlation was lower than 1, suggesting that ASI was "over-represented" in P+ASI.

For example, based on the water balance, we expect the correlation between volume and P to improve (or at least not to worsen) when initial wetness is taken into account (i.e. P+ASI). However, for T1 this was not the case. But when ASI was adjusted (weighted down) the relationship did improve.

Furthermore, the weighting factor provides a diagnostic measure of how strongly antecedent wetness contributes 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 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 relationship between P+w·ASI and the analyzed SSF metric

Changes in manuscript:

We revised the text clarifying the reasons and the purpose of why a weighting factor was applied to ASI, and what we mean with "over-representing".

***L 333ff**, eq. 4 and 5: Eq 4 has P_tot on both sides, and thus results in spurious correlation. Eq 5 separates V_tot and P_tot, so this problem does not occur. This quadratic model could also be fitted to the data without taking the detour of eq. 4. But, more importantly, what is the physical interpretation of this? Is there a limit to this quadratic growth, or will there be a point where, say, 10 mm of rainfall produce 100 mm of flow (for a respective rainfall duration)?*

Author's reply:

We agree that Eq. 4 is prone to spurious self-correlation. Therefore, as suggested, we now 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 supports the interpretation that the contributing area likely increases with P.

We also now clarify that there is a limit to this quadratic growth. The hydrological limiting case V/(catchment area) = P yields P = (-b + catchment area)/a (for a>0 and CA > b). In practice, we restrict interpretation to the fitted domain and consider Eq. 5 only valid up to the max. P used for fitting. Moreover, to avoid comparing different precipitation ranges between trenches, Eq. 5 was fitted over the same range (0–50 mm) for both trenches.

Changes in manuscript:

We revised the manuscript accordingly: Eq. 5 was fitted instead of Eq. 4 using the same P range for both trenches. Table 3 was updated to report the revised coefficients *a* and *b* (with confidence intervals) and r^2 of the fitted Eq. 5. Figure 14 was updated and now shows the fit of Eq. 5 and the linear fit, both including confidence intervals. The text describing the relationship between Eq. 4 and Eq. 5 and the interpretation in terms of contributing area was revised. Furthermore, the limit of the quadratic growth was specified.

***L356ff**, Fig 5: Unclear what aggregation by association means. Please describe that clearly in the text. It appears that intensities were averaged over events - why not take the maximum? Please add to the caption that the axis in a) has a gap, and that the axes in a and b are not the same scale.*

Author's reply:

We agree that the figure caption was not fully clear. By "aggregation by association" we meant that for each complex SSF event, a single representative rainfall intensity was computed as the mean of the intensities of its associated triggering rainfall events (e.g. if two triggering events (A and B) are associated with the same complex SSF event, we calculate the representative intensity as ($I30_A$+$I30_B$)/2). By averaging the intensities we get a more inclusive/general picture of the intensities of the triggering rainfall events occurring during a complex SSF event, as they all contribute to the generated SSF.

Changes in manuscript:

We revised the text to clarify what we mean by “aggregation by association“. Furthermore, we added that the axis in panel a) has a gap, and that axes in a) and b) have not the same range/scale.

L 403, Fig. 8*: Have you also analyzed the correlation of ASI alone?*

Author’s reply:

Yes, we also analyzed the correlation of ASI alone. Nevertheless, due to its similarity with VWCi (ASI is based in part on the same raw soil moisture data from which VWCi is derived), we decided not to include it in Figure 8.

Changes in manuscript:

We revised the text noting that correlations computed with ASI were similar to those based on VWCi.

L 418-427, Figs. 9, 10*: The manually plotted trends can be dangerous, if they imply a different kind of relationship than actually supported by the data. Both variables involve measurement uncertainty, which will influence the locations of the data points. Conclusions about the difference of the two sites cannot be drawn from this. On the linear scale, the patterns do not look too different, and also on the log scale, the dashed line from T1 could be made fit to T3.*

Author’s reply:

We agree that manually drawn trend lines can be misleading and may imply relationships not supported by the data. Therefore, we removed the manually plotted trend lines from Fig. 9 and Fig. 10, and revised the figure captions accordingly. In the text related to Fig. 9 and Fig. 10, we avoid strong conclusions about differences between T1 and T3 based on these figures alone and now phrase any site-related interpretation cautiously (e.g., “appears”).

Changes in manuscript:

We revised Figure 9 and 10 removing the manually plotted trend. The corresponding text was revised to phrase any site-related interpretation cautiously.

L 443*: Why should the subsurface runoff coefficient not be “very small” in some cases? How are these events characterized in terms of rainfall, not SSF metrics? For small rainfall events, this would be plausible, because of interception, or soil storage.*

Author’s reply:

We agree that the subsurface runoff coefficient can be very small, particularly for small rainfall events. We therefore removed the statement.

Changes in manuscript:

The sentence in question was removed.

L 449f*: How many events were included in this selection (% of all events)?*

Author's reply:

Considering the old precipitation ranges (T1: P<40 mm; T3: 15<P<43 mm), at T1 90% and T3 40% of all events were included. We now carry out the analysis using the same range of P = 0–50 mm for both trenches (see reply to comment to L 333ff). Considering the new range, 92% at T1 and 86% at T3 of the events are now included.

Changes in manuscript:

We revised the manuscript adding the percentage of the events included in the analysis.

How do you justify analyzing only a selected subset of the data? Including other data would probably give very different results, which is evident when comparing Fig 14 against Figs 9 and 10. The full data show a plateau, although the log scale makes comparison difficult. The relationships in the data thus are essentially different when looking at the entire range, compared to the quadratic relationship this analysis is tailored to. This seems very questionable. Even more so, as a quadratic relationship suggests that the SSF volume increases faster with increasing rainfall. Without limitation, this implies that the SSF volume surpasses the rainfall volume at some point, which is not physically possible (compare comment on eqs. 4 and 5).

Author's reply:

We agree that the previously used subsets of the data (T1: P<40 mm and T3: 15<P<43 mm) were questionable. We therefore re-ran the analysis using the same broader range for both trenches P < 50 mm. We choose P = 50 mm as an upper limit because only a few events fall above this threshold (at T1 5 events represent the P range 50–130mm, and at T3 5 events represent the P range 50–90 mm) and these events show a large scatter. In addition, (as stated in L 585–590), very large events (P>50 mm) are characterized by very long durations and multiple peaks (subevents). Hence, during these events the subsurface has more time to drain/dry between peaks and more water is lost by the consecutive rainfall events to replenish the storage deficit and during the transmission to the trench. Therefore a weaker Volume–Precipitation relationship is somewhat expected for these events (this limitation was acknowledged in the manuscript).

We also agree that the quadratic relationship is only valid within a given range of P. Therefore, we now specify the range where the quadratic relationship is valid (see also reply to comment L 333ff).

Changes in manuscript:

We revised the manuscript and now analyse a greater percentage of events (P<50 mm). The text, Table 3 and Figure 14 concerning the analysis and fit of Eq. 5 was updated. Furthermore, the limit of the quadratic growth was specified.

L 494*: The role of rainfall intensity is not so evident for me in Fig. 16. Between 20 mm and 30 mm, the Q_dmax values cover a wide range. There is not a clearly discernible pattern with regard to VWC_i and I_30? For example, at T1, which has more events in this range, the Q_dmax for medium intensities and medium VWCs range between ~150 and 650 l/h.*

Author's reply:

We agree that the role of rainfall intensity was overemphasized in the original text and that the pattern at T1 was not so evident. We revised the text to reflect this and now describe the pattern more generally. Specifically, we observe how for P>20mm, high Qmax (>200 l/h) values were mostly associated with high I30 and/or with a rainfall amount preceding peak rainfall intensity exceeding 8 mm.

Changes in manuscript:

We revised the text concerning the influence of rainfall intensity on Qmax, reducing its emphasis as a controlling factor, and describe patterns more cautiously (e.g. higher Qmax tended to occur with higher I30 and/or higher rainfall amount preceding peak rainfall intensity).

***L 563f**: Interception storage does not seem to be subordinate in view of your statements regarding the non-triggering rainfall events. Would interception not be the main explanation for the triggering threshold and the seasonality in "Fig." 19?*

Author's reply:

We agree that interception, together with the soil moisture storage deficit, is the main explanation for the triggering threshold and the observed seasonality. In our statement we meant that the interception storage plays a subordinate role in controlling the SSF response (e.g. Volume, Qmax) in comparison to the soil moisture deficit. However, we acknowledge that it was not clear as written. To avoid confusion, we removed the statement that "interception plays a subordinate role".

Changes in manuscript:

We revised the text removing the statement that "interception plays a subordinate role".

***L 571 – 576**: Would considering the soil moisture deficit, i.e. the water volume required for saturation, be a more useful proxy here than ASI? The minimum water volume available for SSF (and overland flow) would then be P_tot - Interception - Soil Moisture Deficit.*

Author's reply:

In our study ASI is used to quantify the wetness conditions preceding an SSF event. Converting ASI into a soil moisture deficit term would require assuming an effective saturated water content (porosity). With a fixed porosity, this conversion is essentially a linear rescaling, so the event-to-event wet/dry dynamics would remain unchanged. Consequently, the relationship based on P+ASI and that based on P-Soil_Moisture_Deficit would be equivalent, except for a shift along the x-axis. Interception is expected to be comparatively small (ca. ca. 2 mm). Moreover, SSF does not require saturation of the entire soil profile; therefore, while the P-interceptio- Soil_Moisture_Deficit would give a *minimum* water volume available for SSF, the actual volume transformed into SSF could be much different (e.g. SSF could be generated despite P<Soil_Moisture_Deficit).

Changes in manuscript:

The text was not revised.

***L 576**: I do not agree that the goal of hydrological research should be limited to "obtaining the best fit". I would favor to include our physical understanding in the data analysis and the interpretation. The "optimum depth" is site dependent because of soil properties like thickness or porosity, which are known to vary across sites.*

Author's reply:

We agree that the goal of hydrological research is not to "obtain the best fit". In our study, ASI was weighted in the composite variable P+ASI to explore how changing its weight affects the correlation with SSF metrics. The weight that maximizes the correlation is used as a descriptive diagnostic of how strongly antecedent wetness contributes to improving the correlation with respect to the analyzed SSF metric (see also reply to comment L 317).

Changes in manuscript:

We revised the manuscript to clarify why a weighting factor was applied to ASI and how the results should be interpreted.

L 584f: *"P_tot –V_tot relationship is poor" - This statement is not clearly supported by the data. Figures 9 + 10 indicate a positive correlation of V_tot and P_tot also above 40 mm, at least at the linear scale, and possibly also a difference between antecedent soil moisture conditions. There is some scatter, and there are admittedly only a few data points in that rainfall range, but a trend is visible, and it is also what one would have expected (more rain, more flow).*

Author's reply:

The data in our V–P plots (Fig. 9 and 10) shows that at both trenches, for P>40 mm, V does not necessarily increase with P (e.g. at T1 the V associated with the ca. 80 mm P event is lower than that associated with the ca. 125 mm, and at T3 the V of the events associated with P ranging from ca. 70 to 80 mm actually decreases). Antecedent soil moisture conditions partly provide an explanation for the observed V–P relationships; however the small number of data points undermines the strength of this interpretation. Hence, the V–P relationship is not well defined for P>40 mm. Possible explanations for the "poorer" trend are presented in L585–589.

Changes in manuscript:

We did not revise the text.

L 597: *Is the 5 to 12 mm threshold really backed up by these citations? Quoting from Tromp- van Meerveld and McDonell 2006a: "Our analyses suggest that there is a clear threshold for significant (>1 mm) subsurface stormflow to occur; significant subsurface stormflow occurred only during rainstorms larger than 55 mm". Fu et al. 2013, Fig. 7 shows a threshold of 5 mm per 0.5 h, not total volume. Noguchi et al. 2001, Table 2 show that the smallest storm event analysed had 15.6 mm total precipitation.*

Author's reply:

We describe how in other studies <5–12 mm of rainfall was enough to activate SSF (i.e. to generate SSF). Figure 6 of Tromp-van Meerveld and McDonnell (2006) clearly shows how SSF was also generated for rainfall amounts much smaller than 55 mm. Only for "significant" SSF to occur rainfall had to exceed 55 mm. Fu et al. (2013) state that "the generation thresholds of total event precipitation for measurable interflow at plots A and B were found to be 12.0 (event 10 in Table 1) and 13.7 (event 3 in Table 2) mm". In Figure 8 of Noguchi et al. (2001), which illustrates the relationship between precipitation and total outflow (i.e. SSF), it is visible how for P<10 mm some SSF is generated. Table 2 does not include all the storm events that were used to analyze the relationship between SSF and P, and that are plotted in Fig. 8.

Changes in manuscript:

The text was not revised.

***L 611f**: This raises the question of what the physical meaning of this would be. Do you consider the soil to drain below the initial soil moisture level during the SSF event?*

Author's reply:

We agree that the previous text did not elaborate on the meaning of the 40 mm threshold and that the sentence was not that clear. We therefore revised the text and now describe how the fact that the P+ASIadj threshold was ca. 40 mm indicates that significant SSF volumes can be generated by a relatively small rainfall event if the antecedent conditions are very wet (i.e. ASIadj is close to 40 mm), whereas if the antecedent conditions are dry (ASIadj << 40 mm) higher rainfall amounts are required to generate comparable SSF volumes. For example, a similar Vtot of 0.63–0.71 m^3 resulted from a relatively small Ptot of 6 mm under wet initial conditions (ASIadj ca. 37 mm), as well as from a large Ptot of 28 mm under dry initial conditions (ASIadj ca. 17 mm).

We do not consider the soil to drain below the initial soil moisture during an SSF event.

Changes in manuscript:

The text was revised to include the interpretation of the 40 mm threshold and to add an example to further clarify it.

***L 616ff** : Either the contributing area changes, or the runoff coefficient of the catchment area changes. They are mutually exclusive. Which kind of field observations of contributing area are you referring to?*

Author's reply:

We agree that the text was confusing because the terms contributing area and subsurface runoff coefficient (Csub) were not well defined. Previously, Csub was intended to represent the runoff coefficient of the contributing area. However, since this was not specified, this rightfully led to confusion. We now clearly define the contributing area as the area that provides water to the trench during an SSF event, and we now define Csub with respect to the catchment area (see also our reply to the general comment).

Regarding the field observations, we refer to the water level dynamics of the uppermost instrumented well. As described in L618–621, the water level in this well did not show any reaction (i.e. an increase in water level) during some of the smaller SSF events, suggesting that for these events the contributing area was likely smaller than the maximum MCA (700 m^2 and 2000 m^2) observed at the trenches. The uppermost well is located ca. 20–25 m from the trench; considering a trench width of 14 m, an area of ca. 350 m^2 is "monitored".

Changes in manuscript:

We revised the text and now clearly define the terms contributing area and subsurface runoff coefficient. Moreover, we now clarify that the water level dynamics of the uppermost instrumented well are the field observation to which we refer.

***L 619ff**: Are there alternative possible explanations? For small events, could there have been increasing accumulation of water in downslope direction, resulting in a small SSF volume at the trench, but not enough for a detectable rise in groundwater level? Or overland flow that infiltrated downslope? Or spatially variable rainfall?*

Author's reply:

We thank you for considering possible alternative explanations. Although we cannot fully exclude that the contributing area extended further uphill than the uppermost well during events when its water level did not increase, the strong relationship between trenchflow rate and water levels in the wells—particularly those close to the trench—led us to interpret the absence of a water-level response in the uppermost well as an indication that the area around the well was not contributing to the observed SSF. However, it is indeed possible that a small amount of water could have laterally "bypassed" the (fully screened) well without a detectable increase of the water level in well. We now acknowledge this in the manuscript.

The occurrence of overland flow bypassing the well without any response of the water table is highly unlikely, as given the highly permeable upper soil layer we expect a high infiltration capacity. Regarding the spatially variable rainfall, we do not expect large differences over the small instrumented area of ca. 350 m² (trench width of 14 m times the 25 m distance to the uppermost well).

Changes in manuscript:

We revised the manuscript to acknowledge the possibility that water could have laterally bypassed the well and thus that the contributing area may have extended further upslope than the well.

L 623ff*: On the contrary, one could also argue that without knowledge about the contributing area and about specific discharge, the observed volumes are not interpretable in terms of the local water balance.*

Author's reply:

We agree that if the extent of the SSF catchment area can be confidently defined, it would be beneficial to compute the specific discharge. However, precisely delineating the catchment area of SSF is very challenging, as it depends not only on the topography but also on the hydrogeological setting, which over the entire catchment is not known in detail. Nevertheless, in order to provide an "upper bound", we calculated the topographically-derived catchment area and now report it in the manuscript. Hence, if desired, the volumes can now be converted to mm.

Changes in manuscript:

We added the topographically derived catchment areas to the site description in Section 2.1, we now direct the reader to these areas in the text.

L 635 - 643*: Although you might not be wrong, I do not fully get the point here. First, is the "barrier" you are referring visible in Fig. 10? Is it the part of the hand-drawn curve between 10 and 20 mm of P_tot? If so, it might be an artifact of the manually drawn trend line. If I understand your argumentation correctly, the increasing rainfall results in an increased contributing area, which in turn might add additional "barriers", or sinks, to the flow field. Why would then the volume not increase just due to the additional rainfall received by the former smaller contributing area?*

Author's reply:

The "barrier" we are referring to in the text manifests in the reduced increase of Vtot with Ptot between ca. 5–15 mm (wet antecedent conditions) and ca. 10–20 mm (dry antecedent conditions) (see Fig. 10). The volume is still expected to increase with Ptot but with a reduced rate. We agree that the hand-drawn trend may have overemphasized the "plateau" between 10–20 mm. We therefore removed the hand-drawn trends from Fig. 9 and 10.

We agree that the reduced increase of Vtot with Ptot in the ranges of Ptot indicated above, may also be due to chance, and additional SSF events in that range with higher Vtot than those observed might change our interpretation. We now acknowledge this limitation in the manuscript.

Changes in manuscript:

We revised the manuscript acknowledging that the small number of events within the reported Ptot ranges limits the robustness of this interpretation. Furthermore, the hand-drawn trend line in Figures 9 and 10 was deleted.

***L 725**: Please elaborate the discussion of this point, as this contradicts your finding of higher runoff coefficients in winter.*

Author's reply:

Thank you for your comment. As written, this discussion point would indeed contradict our findings. In the text we swapped "summer" with "spring" by mistake. We revised the text and now describe how the runoff coefficients were higher in spring than in summer, as observed by Penna et al. (2015).

Changes in manuscript:

The text was revised. We now report that higher runoff coefficients were observed in spring, as stated by Penna et al. (2015).

***L 739ff**: This conclusion is not supported by your data. The two sites may differ in the precipitation threshold, but they do not really show a different pattern in rainfall-runoff response, see Fig 14. An apparent difference in behavior could be artificially created by the manual trend lines in the semilogarithmic plots of Figs. 9 and 10, at least for wet conditions. The different scale of the x-axis for the linear-linear plots could also add to this misconception.*

Author's reply:

We agree that, when considering the big picture of all data points, the general pattern of the rainfall-SSFvolume response is indeed broadly similar at both trenches.

At T3, in the Ptot range between 0–50 mm, an increase in Vtot with Ptot can be observed (Figure 10). However, within the general increasing relationship, we see a number of data points in the range of Ptot = 5–15 mm (wet antecedent conditions) and ca. Ptot = 10–20 mm (dry antecedent conditions) being located somewhat lower than the general trend suggested by the surrounding observations. This region can be interpreted as a "plateau" characterized by a weaker increase of Vtot with Ptot, which interrupts the general increase of Vtot with Ptot. Although the physical cause cannot be pinned down, a certain retention volume that has to be overcome could explain the behavior. However, we acknowledge that the robustness of this interpretation is undermined by the scatter and the relatively small number of data points present within the considered ranges. Hence, we removed this interpretation from the conclusion.

The (new) fit of the quadratic Eq. 5 to the Vtot–Ptot data (Ptot range 0–50 mm for both trenches; see also our reply to L 333ff) shows that, for events with wet antecedent conditions, the rate at which Vtot increases with Ptot is greater at T3 than at T1. This difference between the trenches was added to the conclusion. Moreover, to visually facilitate the comparison between the two trenches, the x- and y-axis ranges in Figure 14 were adjusted and are now identical.

Changes in manuscript:

We revised the text to convey that the general rainfall–SSFvolume pattern does not substantially differ between the two trenches. Any apparent deviation at T3 is now described cautiously acknowledging the uncertainty. The statement on the precipitation threshold at T3 was removed from the conclusion. However, we added that under wet antecedent conditions, the rate at which Vtot increases with Ptot is greater at T3 than at T1. Figure 14 was updated so that the x- and y-axis ranges are now the same for both trenches.

***L 775f**: You probably should not simply adjust the lab-measured porosities because the Topp equation gives a different result. On the one hand, you measured porosities in the lab in order to use the CRIM because Topp cannot account for temperature effects, which you claim to be important. In cases where the two differ, you ignore the measured data, including the reported bulk densities, and take some arbitrary value? Also, the increase in porosity at greater depth at T1 is questionable. How would your results change if the actually measured porosities are used?*

Author's reply:

We thank the reviewer for the in-depth analysis and acknowledge that the paragraph was not well written and that crucial information on why the correction was necessary and, on the criteria, used to carry it out where missing.

The lab-measured porosities were only adjusted when the max. VWC resulting from the Topp and from the CRIM equation (computed using the lab-measured porosities) were substantially higher than the measured porosities (e.g. 41% vs 25%). For the CRIM equation this implies that the max. measured apparent dielectric permittivities were higher than the theoretical max. apparent dielectric permittivity that would result for a fully saturated medium assuming the lab-porosity; this was also true when, in a test, the permittivity of the solids was assumed to be 4 or 5. VWC values greater than the porosity do not make physically sense. Therefore, we adjusted the measured porosity so that it would be similar to the max. observed VWC (Topp) and the max. CRIM-calculated VWC approximated the adjusted porosity.

Different factors can provide a plausible explanation for the higher (required) porosity values. First, the soil moisture sensors were not installed at exactly the same location where the soil samples were collected, and small-scale heterogeneity can characterize these unconsolidated deposits. 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 cm^3 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.

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 soilds (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 principally used as an indicator for wet/dry conditions (relative temporal changes), such that VWC dynamics are more relevant than the absolute VWC values.

Changes in manuscript:

We revised the manuscript to clarify why the porosity correction was necessary and what criteria were used to carry it out. Moreover, we 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).

<u>Minor Comments</u>

L 67*: Maybe word it the other way round, making V_tot the dependent variable*

Author’s reply:

We agree, that it makes more sense the other way around.

Changes in manuscript:

The sentence was reformulated as suggested.

L 74*: I am not sure whether it is helpful to portray the findings as “complex”. We indeed know about the threshold behavior, which is partially depending on soil moisture conditions. Otherwise, SSF is proportional to the rainfall volume and intensity, which is logical considering their causal relationship.*

Author’s reply:

We agree.

Changes in manuscript:

The sentence was removed.

L 76*: Soil characteristics including macropores should be mentioned here as well.*

Author’s reply:

We agree and included soil characteristics and macropores in the sentence.

Changes in manuscript:

Soil characteristics and macropores were included in the sentence.

L 116ff*: Perhaps consider adding some photographs of the trenches?*

Author’s reply:

We added some pictures in the Appendix.

Changes in manuscript:

We added some pictures in Appendix A.

L 134*: " to the base of the trench" at xy m: mention the depth of the trench here*

Author's reply:

We agree. This information is useful in the sentence.

Changes in manuscript:

We added the depth of the trench base to the sentence.

L 144*: "silt (loam)" – what is it? Silt and loam are describing different particle size distributions*

Author's reply:

We agree that it was confusing as written. Loam was referring to the general classification of the soil. To avoid confusion we removed "(loam)".

Changes in manuscript:

We removed "(loam)" to avoid confusion.

L 175*: "the end of the study period (27 May 2024–25 June 2024)" – are the dates indicating the study period, or the time of installation of the weir? If the latter, you may just omit the parentheses and add "from.. to"*

Author's reply:

We meant the time of the installation of the weir.

Changes in manuscript:

We revised the sentence following the suggestion.

L 201*: Please provide your definition of outliers.*

Author's reply:

Outliers were caused by sporadic malfunctions of the tipping-bucket logging system, where the number of recorded tips within a time interval was markedly lower than in the preceding and subsequent interval, producing isolated, unrealistic drops in discharge (e.g. from 30 l/h to 1 l/h and back to 30 l/h). These hydrologically non-explainable, isolated drops in discharge were manually removed during the visual analysis of the trenchflow timeseries.

Changes in manuscript:

We revised the text clarifying the cause for the outliers.

***L 272f**: Instead of I_5, would computing I_10, i.e. maximum rainfall rate in 10 minutes, not help to increase the sample size? It would not touch the significance of the analysis in my opinion, since trench flow was also recorded in 10-minute intervals, but would add ~25% to the number of events at both sites.*

Author's reply:

We agree that the sample size would be larger for I10. However, we do not expect the results to change substantially if I10 is considered instead of I5.

Changes in manuscript:

The manuscript was not changed.

***L 274f**: Please explain this more clearly. Was I_m derived from the others, or were all averaged to get something else?*

Author's reply:

The rainfall intensities (Im, I5, I30 and I60) of multiple triggering events associated with the same complex SSF event were additionally averaged. In other words, for each complex SSF event, the associated Im, I5, I30 and I60 was obtained by computing the arithmetic mean of the Im, I5, I30 and I60, respectively, of the triggering events. For example, for two triggering events A and B associated with the same complex SSF event, the I30 intensity was calculated as the arithmetic mean of the triggering events: $(I30_A + I30_B)/2$.

Changes in manuscript:

We revised the text clarifying the averaging approach.

***L380, Fig 7**: More explanation needed, in caption and in text. Why are the numbers for subevents/simple events for the same site differing between variables?*

Author's reply:

We agree that more explanation was needed. For the same site, the sample size of the subevent/simple events metrics differed because:

1) for the rising rate calculation very small events having Qmax<3 l/h were not considered.
2) in a few cases the rising phase of a subevent was interrupted by a successive subevent, preventing the potential peakflow of the preceding subevent from being reached. As a result, for these subevents the peak-dependent metrics (Qmax, lag time, rising rate) were not considered.
3) occasionally simple events and subevents were characterized by a well-defined double peak with a higher second peak. In these cases the Qmax was used in the analysis but the time-dependent metrics (e.g. rising rate) were not, as the double peak nature of the event would lead to inconsistent results. These criteria are outlined at the end of Section 2.3.6.

Changes in manuscript:

We revised the manuscript clarifying why the sample number of different variables of the same site may differ.

***L 409**: "P_tot > 40–50 mm"– Choose one number that defines "very large" events.*

Author’s reply:

We choose 50 mm to define *"very large" events.*

Changes in manuscript:

We revised the manuscript and now define a very larger event as Ptot > 50 mm.

***L 428f**: Does this refer to the correlation of V_tot and P_tot? Above, the Spearman correlation was used. Uclear why you are showing Pearson here.*

Author’s reply:

We agree, this was confusing. This referred to the linear correlation (Pearson) between Vtot and Ptot. To avoid confusion, we removed the correlation value from the sentence.

Changes in manuscript:

We removed the correlation value from the sentence

***L 464, Fig. 14**: The x-axis of the right panel is cut off before 15. Please show the full range, or at least point to it in the caption.*

Author’s reply:

We updated the figure to include the Ptot range (x-axis) from 0 to 50 mm for both trenches.

Changes in manuscript:

We updated Figure 14 to include the Ptot range (x-axis) from 0 to 50 mm for both trenches.

***L 495**: What are the correlations of events > 20 mm?*

Author’s reply:

We did not specifically calculate the correlations of events > 20 mm. Hence, to avoid confusion, we removed the sentence.

Changes in manuscript:

The sentence was removed from the manuscript.

***L 496**: “were associated with fairly high rainfall intensities” – and high total rainfall volumes.*

Author’s reply:

Since the uncertainty in the interpretation was high, we removed the sentence from the manuscript.

Changes in manuscript:

The sentence was removed from the manuscript.

***L 539**: Results not shown? Consider putting these in supplementary material.*

Author's reply:

These specific results are not shown. We will consider putting these in supplementary material.

Changes in manuscript:

The manuscript was not changed.

***L 569-571**: Perhaps mention that this corroborates similar findings of many earlier studies, and cite them.*

Author's reply:

We agree that citing earlier studies that corroborate our findings would be beneficial.

Changes in manuscript:

We revised the text adding studies that corroborate our findings.

***L 603**: "In contrast…" - Is this an additional threshold to the ≈ 2 mm one, or is the threshold here at ≈ 15 mm? Please clarify.*

Author's reply:

Exactly. As specified in the following sentence L604–605, 15 mm represents a second threshold, beyond which significantly more SSF is generated. The 2 mm is the "activation threshold", i.e. rainfall amount required to generate any SSF (even very small volumes). We revised the text in order to clarify the distinction between the two thresholds.

Changes in manuscript:

We revised the text to clarify the distinction between the two thresholds.

***L 605**: Does this mean that the observations at T1 do not match these results?*

Author's reply:

Yes. At T1 the Vtot increases gradually with Ptot without showing any pattern that would suggest the presence of such threshold.

Changes in manuscript:

The manuscript was not changed.

***L 610**: T3 is in Fig 12. Would the 40 mm not apply to both T1 and T3?*

Author's reply:

No. The 40 mm refers to the second threshold identified analyzing the relationship with the combined variable $P+ASI_{adj}$. The second threshold was only identified for T3 (see also reply to L 603). No such threshold is observed for T1 (Figure 11).

Changes in manuscript:

The manuscript was not changed.

***L 706f**: This would imply that no fast SSF response can be expected in case of large and intense rainfall in wet conditions. What else is expected then - overland flow?*

Author’s reply:

On the contrary. Because under wet antecedent conditions a fast SSF response was observed *even* for rainfall events with low intensities and small rainfall amounts, events with higher intensities and/or larger amounts are likewise expected to—and did—produce fast (and in some cases faster) SSF responses under wet conditions. Our point was that *similarly* fast SSF responses can also occur under dry antecedent conditions when rainfall is large and intense—i.e., response time is not uniquely determined by rainfall amount or intensity. Our data showed that, comparable (short) response times occurred across contrasting combinations of rainfall characteristics and antecedent wetness, and short response times were occasionally observed even for events with small rainfall amounts under dry antecedent conditions.

Changes in manuscript:

We revised the text to clarify the sentence and to incorporate the explanation above.

***L 710 ff**: The different configurations of flow paths - across sites and across conditions - is the important factor in SSF dynamics. Thus, perhaps consider discussing the methodological issues first and then these 'complexities', rather than putting them 'beside'.*

Author’s reply:

Because, the difference in flow path configurations - across sites and across conditions - is the important factor in SSF dynamics, we choose to discuss them first and then later address the methodological issues. We chose this structure to guide the reader from process understanding to methodological considerations.

Changes in manuscript:

The manuscript was not changed.

***L 733**: Which are these “few” studies, and how do their findings relate to yours? Please elaborate.*

Author’s reply:

We agree the sentence was confusing. The “few” studies we refer to are those already cited in the Section. We wanted to emphasize that the number of studies that examine the seasonal variability of SSF is very small.

Changes in manuscript:

We clarified the sentence to avoid confusion.

Technical Corrections

L 69*: 2006a or b?*

Author's reply:

We referred to Tromp-van Meerveld and McDonnell (2006a), but the reference field was not updated.

Changes in manuscript:

We now specify that the correct reference is "2006a."

L 69-70*: Noguchi et al. … matrix flow: Check wording*

Author's reply:

Noguchi et al. analyzed the relationship with both macropore flow as well as matrix flow.

Changes in manuscript:

The manuscript was not changed.

L 71-72*: how … was present: Check Wording*

Author's reply:

We agree, "how" is not he best fit in this sentence. "How" was replaced with "that".

Changes in manuscript:

"How" was replaced with "that".

L 126*: temperature and precipitation ARE*

Author's reply:

We agree.

Changes in manuscript:

"is" was replaced with "are".

L 134*: Is the range 2 to 256 mm, or are all greater than 2-256 mm? Consider omitting the ">" and check style guide how to indicate range of numbers*

Author's reply:

We referred to the range 2 to 256 mm. Hence, the ">" was omitted.

Changes in manuscript:

We removed the “>”.

L 271 *and elsewhere: P_tot etc. - Check style guide if subscript should be italicized*

Author’s reply:

Indeed. Following the guidelines of HESS the subscript should not be italicized (“If super- or subscripts are abbreviations … it is typeset in roman”). Thus, in the variables P_{tot}, but also V_{tot}, I_{m}, I_{5}, I_{30}, I_{60}, Q_{i} and $Q_{\Delta\mathrm{max}}$ the subscript has to be roman. We changed the typeface styles of the subscripts from italics to roman in all these variables throughout the manuscript, including the Figures.

Changes in manuscript:

We changed the typeface styles of the subscripts from italics to roman in all affected variables throughout the manuscript, including the Figures.

L 405*, Fig. 8: row header reads “Time lag”, column header “lag time”*

Author’s reply:

We agree, it has to be “lag time” in both cases.

Changes in manuscript:

We revised Figure 8, substituting “time lag” with “lag time”.

L 418-427, Figs. 9, 10*: Consider making one figure with two panels (also applies to other figures)*

Author’s reply:

Keeping the figures separate preserves layout flexibility in Section 3.3.1, allowing them to be positioned closer to the relevant text, which is helpful given the large number of figures in that section.

Changes in manuscript:

The manuscript was not changed.

Figs 15, 16, 17, 18*: Circles in figures are larger than in the legend, which makes the figure unreadable.*

Author’s reply:

We agree. Hence, we revised the figures so that the circle sizes shown in the legend match those in the plots.

Changes in manuscript:

The figures in the manuscript were revised.

***L 499**: “QΔmax ranging from 130 to 638 mm/h” – check unit*

Author’s reply:

We agree the unit was wrong. The sentence was removed (for other reasons) from the manuscript, so this is no longer relevant.

Changes in manuscript:

The sentence was removed from the manuscript.

***Fig 17, 18**: Mention log scale in caption*

Author’s reply:

Ok.

Changes in manuscript:

We added that “The y-axis is logarithmic” to the figure caption.

***Fig 19**: Is this deemed to be a table? Give more details in the caption and/or the table: these are statistical values of selected data - "trimmed means". Also indicate the number of events in each season after trimming.*

Author’s reply:

We opted for a heatmap to visualize the data. We agree that the caption was not so informative. Furthermore, we now also indicate the number of the samples after trimming.

Changes in manuscript:

Figure 19 was revised and now also displays the sample number. Moreover, we revised the figure caption clarifying that the values (with coloured background) show the seasonal trimmed means and the values (with withe background) show the number of the samples after trimming.

***L 592**: Year missing in citation.*

Author’s reply:

We agree.

Changes in manuscript:

The year was added to the citation.

***L 598**: 2006 a or b?*

Author’s reply:

We meant 2006a.

Changes in manuscript:

We revised the citation.

L 601*: Consider replacing "= ca" with ≈, or with words " is approx."*

Author’s reply:

We replaced ca. with ≈*.*

Changes in manuscript:

We replaced ca. with ≈*.*

L 604*: Figs 10 and 11, should be 9 and 10?*

Author’s reply:

Yes, it should be 9 and 10.

Changes in manuscript:

We revised the figure number. (The original sentence was revised).

L 710*, and others: Consider combining parentheses to avoid “)(“*

Author’s reply:

We considered this. However, in this case “(McGuire et al., 2024)” is cited to support all the claims in the sentence, not just the SSF-producing mechanisms listed in (e.g. …) (McGuire et al., 2024). In some other parts of the manuscript we now avoid using “)(“.

Changes in manuscript:

In some other parts of the manuscript we now avoid using “)(“.

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, 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.

Fu, C., Chen, J., Jiang, H., and Dong, L.: Threshold behavior in a fissured granitic catchment in southern China: 1. Analysis of field monitoring results, Water Resour. Res., 49, 2519–2535, https://doi.org/10.1002/wrcr.20191, 2013.

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, 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.

Noguchi, S., Tsuboyama, Y., Sidle, R. C., and Hosoda, I.: Subsurface runoff characteristics from a forest hillslope soil profile including macropores, Hitachi Ohta, Japan, Hydrol. Process., 15, 2131–2149, https://doi.org/10.1002/hyp.278, 2001.

Penna, D., Van Meerveld, H. J., Oliviero, O., Zuecco, G., Assendelft, R. S., Dalla Fontana, G., and Borga, M.: Seasonal changes in runoff generation in a small forested mountain catchment, Hydrol. Process., 29, 2027–2042, https://doi.org/10.1002/hyp.10347, 2015.

Selby, M. J.: Earths´s 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.

Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope, Water Resour. Res., 42, W02410, https://doi.org/10.1029/2004WR003778, 2006.

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.