

EGUSPHERE-2025-5110

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

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

Point-by-point reply

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

Reviewer 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 will state this limitation explicitly in the text.

We agree that the terms "contributing area", and "subsurface runoff coefficient" (C_{sub}) 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 C_{sub} , 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 will define C_{sub} referenced to the SSF catchment area (i.e. $C_{sub} = \text{Volume in mm} / \text{rainfall in mm}$, where $\text{Volume in mm} = \text{Volume in liters/catchment area in m}^2$).

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 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 will include the topographic catchment area in our manuscript clarifying that it may differ from the “real” SSF catchment area.

We will revise the text discussing the validity and implications of the quadratic increase. Furthermore, we will clearly define the terms “contributing area”, MCA and subsurface runoff coefficient, and add 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 will add the uncertainty for the flowrate measurements carried out with the tipping buckets. We will also revise 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). A compromise could be plotting the full data range in the nearby semilogarithmic plot (right 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 will directly fit Eq. 5 in the future and use the same range of rainfall (up to $P = 50$ mm) for both trenches (see details in the 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.

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 VWC_i) is clear, and the observed VWC_i ranges are substantial (T1: 23.3–34.0%; T3: 14.9–35.6%). We will revise the text 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. This was not properly described and rightfully led to confusion. Therefore, we will revise the manuscript decoupling Eq. 2 from Eq. 3 by reformulating Eq. 3 (the new formulation will yield the same ASI values as the previous one). In the new formulation, ASI will be computed in a more direct way: the water content measured by the sensors at depths of 10, 30 and 60 cm BGL is 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.

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 $P + w \cdot ASI$ 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 the relationship between $P+w \cdot ASI$ and the analyzed SSF metric

We will revise 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 will fit Eq. 5 directly.

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

We will also clarify that there is a limit to this quadratic growth. The hydrological limiting case $V/(\text{catchment area}) = P$ yields $P = (-b + \text{catchment area})/a$ (for $a > 0$ and $CA > b$). In practice, we will 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 will be fitted over the same range (0–50 mm) for both trenches.

We will revise the manuscript fitting Eq. 5 instead of Eq. 4 using the same P range for both trenches. Table 3 will be updated to report the revised coefficients a and b (with confidence intervals) and r^2 of the fitted Eq. 5. Figure 14 will be revised to show 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 will be revised. Furthermore, the limit of the quadratic growth will be 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 $(I_{30A} + I_{30B})/2$). 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.

We will revise the text to clarify what we mean by “aggregation by association”. Furthermore, we will add 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.

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 will remove the manually plotted trend lines from Fig. 9 and Fig. 10, and revise the figure captions accordingly. In the text related to Fig. 9 and Fig. 10, we will avoid strong conclusions about differences between T1 and T3 based on these figures alone and will phrase any site-related interpretation cautiously (e.g., "appears").

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 will therefore remove the statement.

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. In the future, we will carry out the analysis using the same range of $P = 0\text{--}50$ 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 will be included. The percentages of included events will be added.

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 will therefore re-run 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 will specify the range where the quadratic relationship is valid (see also reply to comment L 333ff).

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 will revise the text to reflect this and will describe the pattern more generally. Specifically, we observe how for $P > 20$ mm, high Q_{max} (> 200 l/h) values were mostly associated with high I_{30} and/or with a rainfall amount preceding peak rainfall intensity exceeding 8 mm.

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, Q_{max}) in comparison to the soil moisture deficit. However, we acknowledge that it was not clear as written. To avoid confusion, we will remove 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 $P_{tot} - \text{Interception} - \text{Soil Moisture Deficit}$.*

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 $P - \text{Soil_Moisture_Deficit}$ 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 $P - \text{interceptio- Soil_Moisture_Deficit}$ 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 $P < \text{Soil_Moisture_Deficit}$).

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). We will revise 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.

L 597: *Is the 5 to 12 mm threshold really backed up by these citations? Quoting from Tromp- van Meerveld and McDonnell 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.

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+ASladj 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. ASladj is close to 40 mm), whereas if the antecedent conditions are dry (ASladj \ll 40 mm) higher rainfall amounts are required to generate comparable SSF volumes. For example, a similar Vtot of 0.63–0.71 m³ resulted from a relatively small Ptot of 6 mm under wet initial conditions (ASladj ca. 37 mm), as well as from a large Ptot of 28 mm under dry initial conditions (ASladj ca. 17 mm).

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

We will revise the text 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. In the future, we will clearly define the contributing area as the area that provides water to the trench during an SSF event, and we will define Csub with respect to the SSF 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 (we will clarify this). 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 m² and 2000 m²) 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 m² is “monitored”.

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 will 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).

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 will calculate the topographically-derived catchment area and will report it in the manuscript. Hence, if desired, the volumes can be converted to mm.

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 V_{tot} with P_{tot} 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 P_{tot} but with a reduced rate. We agree that the hand-drawn trend may have overemphasized the “plateau” between 10–20 mm. We will therefore remove the hand-drawn trends from Fig. 9 and 10.

We agree that the reduced increase of V_{tot} with P_{tot} in the ranges of P indicated above, may also be due to chance, and additional SSF events in that range with higher V_{tot} than those observed might change our interpretation. We will revise the manuscript to acknowledge that the small number of events within the reported P_{tot} ranges limits the robustness of this interpretation.

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 will revise the text to describe how the runoff coefficients were higher in spring than in summer, as observed 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 P_{tot} range between 0–50 mm, an increase in V_{tot} with P_{tot} can be observed (Figure 10). However, within the general increasing relationship, we see a number of data points in the range of $P_{tot} = 5–15$ mm (wet antecedent conditions) and ca. $P_{tot} = 10–20$ 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 V_{tot} with P_{tot} , which interrupts the general increase of V_{tot} with P_{tot} . 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 will remove this interpretation from the conclusion.

The (new) fit of the quadratic Eq. 5 to the V_{tot} – P_{tot} data (P_{tot} 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 V_{tot} increases with P_{tot} is greater at T3 than at T1. This difference between the trenches will be added to the conclusion. Moreover, to visually facilitate the comparison between the two trenches, the x- and y-axis ranges in Figure 14 will be adjusted to the same limits.

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 were 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 physical 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 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.

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 soils (K_s) and porosity. Nevertheless, we used the CRIM approach primarily to reduce temperature sensitivity in the permittivity–VWC conversion. In this study, VWC was solely used as an indicator for wet/dry conditions (relative temporal changes), such that VWC dynamics are more relevant than the absolute VWC values.

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

Minor Comments

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. We will reformulate the sentence 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. We will remove the sentence.

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

Author’s reply:

We agree and will include soil characteristics and macropores in the sentence.

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

Author's reply:

We initially considered adding some pictures, but since the paper is already very long, we chose not to. We leave the decision to the editor.

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. We will add 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 will remove “(loam)”.

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. We will revise 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. We will revise 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.

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 (I_m, 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 I_m, I5, I30 and I60 was obtained by computing the arithmetic mean of the I_m, 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: $(I_{30A} + I_{30B})/2$. We will revise the text to clarify 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 $Q_{max} < 3$ 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 (Q_{max} , 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 Q_{max} 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.

We will revise the manuscript to clarify 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. We will revise the manuscript accordingly.

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 V_{tot} and P_{tot} . To avoid confusion, we will remove 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 will update the figure to include the P_{tot} 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 will remove the sentence.

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 will remove the sentence 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.

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. We will revise 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:

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 will revise the text in order 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 V_{tot} increases gradually with P_{tot} without showing any pattern that would suggest the presence of such threshold.

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 $P+ASl_{adj}$. The second threshold was only identified for T3 (see also reply to L 603). No such threshold is observed for T1 (Figure 11).

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. We will revise 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.

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. We will clarify 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. We will revise the manuscript correcting the reference to “2006a”. This will be done also for other Tromp-van Meerveld and McDonnell citations where the “a” or “b” was missing.

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.

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

Author’s reply:

We agree, “how” is not the best fit in this sentence. “How” will be replaced with “that”.

L 126: *temperature and precipitation ARE*

Author’s reply:

We agree. “is” will be 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 “>” will be omitted in the revised manuscript.

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 $Q_{\Delta max}$ the subscript has to be roman. We will change the typeface styles of the subscripts from italics to roman in all these 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. We will revise 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:

Since trenches are sometimes described individually in the text, referencing directly the figure (of the individual trench) without having to specify the panel comes in handy. We leave the decision to the editor.

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

Author's reply:

The largest circle in the legend is indeed smaller than the largest circle in the plot. The legend is intended to be qualitative, indicating that smaller circles correspond to smaller values (e.g., Q_i) and larger circles to larger values. In MATLAB (R2021b), the marker size shown in the legend for scatter objects is limited by the legend rendering, so the legend symbol does not fully reproduce the maximum marker size used in the plot.

L 499: " $Q_{\Delta max}$ ranging from 130 to 638 mm/h" – check unit

Author's reply:

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

Fig 17, 18: Mention log scale in caption

Author's reply:

Ok. We will add 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 will add the number of samples after trimming to the Figure. We will also revise the figure caption clarifying that the values (e.g. with coloured background) show the seasonal trimmed means and the values (e.g. with white background) show the number of the samples after trimming.

L 592: *Year missing in citation.*

Author's reply:

We agree. We will add the year to the citation.

L 598: *2006 a or b?*

Author's reply:

We meant 2006a. We will revise the citation.

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

Author's reply:

We will replace ca. with \approx .

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

Author's reply:

Yes, it should be 9 and 10. We will revise the figure number.

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). We will revise other parts of the manuscript to avoid using "((" where possible.

References

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.

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.

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.