

Point-by-Point Response to Review Comments

Manuscript Title: Increased Nonstationarity of Stormflow Threshold Behaviors in a Forested Watershed Due to Abrupt Earthquake Disturbance

Authors: Guotao Zhang, Peng Cui*, Carlo Gualtieri, Nazir Ahmed Bazai, Xueqin Zhang, Zhengtao Zhang

Manuscript ID: hess-2022-315

Key Laboratory of Land Surface Pattern and Simulation,
Institute of Geographic Sciences and Natural Resources Research,
Chinese Academy of Sciences (CAS),
Beijing 100101, China

(**C** and **R** denote Comment and Reply, respectively)

Response to Editor (Genevieve Ali) Comments:

C: Two referees have evaluated your initial submission. As you saw from their comments, they both recommended that a range of edits be made, to improve the clarity of the abstract and methods, and to ensure that the interpretations and conclusions are supported by the data (among other things). I see, from the Interactive Discussion page, that you have already answered some review comments. I therefore return your manuscript for major revision. I look forward to receiving your revised manuscript, which will be sent out for review.

R: We are very grateful for valuable suggestions from the editor and the reviewers for their positive assessments of our work. We have completely addressed the comments point by point from two Reviewers as below. We also hope that this study with the interconnection of hydrological sciences and flash flood disasters could be considered for publication in the “*Hydrology and Earth System Sciences*”.

Response to Reviewer #1 Comments:

C1: Summary: Increased Nonstationarity of Stormflow Threshold Behaviors in a Forested Watershed Due to Abrupt Earthquake Disturbance assessed changes in hydrologic response of a forested experimental watershed in the eastern Tibet Plateau following an earthquake. The authors characterized longer-term changes in threshold behavior in the watershed and introduced a new metric to quantitatively express thresholds for watersheds with areas of disparate land use, ecology, and physiography. The authors found that lower threshold values were observed in disturbed landslide regions and that non-stationarity in thresholds was mainly controlled by changes to the dominant runoff generation mechanisms of subsurface stormflow and the variable source area.

Significance: This work is significant in several ways:

It contributes to our growing understanding of threshold-mediated hydrologic response. It contributes to the further advancement of a unified threshold-based hydrologic theory. It assesses longer-term trends in threshold behavior following an environmental disturbance. It introduced a new metric to quantify and compare thresholds.

R1: We are very grateful for having the summary of your positive assessments and appreciation of our work. The point-by-point comments have been addressed below. We also hope that this study with the interconnection of hydrological sciences and flash flood disasters could be considered for publication in the “*Hydrology and Earth System Sciences*”.

C2: I found the abstract difficult to digest. Multiple results are communicated, but there is little context for the reader, making it difficult to understand the methodology or the jargon used in the abstract. Consider revising the abstract to be more general to start and highlighting only key results.

R2: Thanks for your comment. The abstract **has been revised as** “Extreme earthquake disturbances to local and regional landscape vegetation could swiftly impair former hydrologic

functioning, significantly increasing the challenge of predicting threshold behaviors of rainfall-runoff processes as well as the hydrologic system's complexity over time. It is still unclear how alternating catchment hydrologic behaviors under an ongoing large earthquake disruption are mediated by long-term interactions of landslides and vegetation evolutions. In a famous Wenchuan earthquake-affected watershed, the nonlinear hydrologic behavior is examined as having two thresholds with intervening linear segments. A lower *rising threshold* (T_r) value (210.48) observed in post-earthquake local landslide regions exhibited a stormflow response faster than that in undisturbed forest and grass-shrub regions, easily triggering huge flash flood disasters. Additionally, an integrated response metric pair (integrated watershed average *generation threshold* T_{g-IWA} and *rising threshold* T_{r-IWA}) with areas of disparate land use, ecology, and physiography was proposed and efficiently applied to identify catchment hydrologic emergent behaviors. The interannual variations of two hydrologic thresholds pre- and post-earthquake were assessed to detect the temporal nonstationarity in hydrologic extremes and nonlinear runoff response. The year 2011 was a turning time in the unsteady recovery process, as post-earthquake landslides evolutions reached a state of extreme heterogeneity in space. At that time, the T_{r-IWA} value decreased by ~ 9 mm compared to the pre-earthquake level. This is closely related to the fast expansion of landslides leading to a larger extension of variable source area from channel to neighboring hillslopes and faster subsurface stormflow contributing to flash floods. Finally, we present a conceptual model interpreting how the short- and long-term interactions of earthquake-induced landslides and vegetation affect flood hydrographs at event timescale that generated an increased nonstationary hydrologic behavior. This study expands our current knowledge about threshold-based hydrological and nonstationary stormflow behaviors in response to abrupt earthquake disturbance for the prediction of future flood regimes.”

C3: Starting at the end of Line 50 the authors suggest that most threshold behavior in rainfall-runoff relationships reported in the literature has been of the hockey-stick diagnostic shape. I think it is notable that most of the listed studies had an identification procedure only compatible with this shape of a threshold. Otherwise, the wording is somewhat ambiguous and may lead to readers assuming that the

dominance of the hockey-stick shape is process-driven or a reflection of some common element in watershed behavior.

R3: Thanks for your serious comment. I agree with you, and the sentence **has been revised as** “In the literature, the runoff behaviors with Hockey stick shape were found at the hillslope (Tromp-Van Meerveld and McDonnell, 2006; Fu et al., 2013a; Wang et al., 2022) and watershed scales (Wei et al., 2020; Farrick and Branfireun, 2014; Scaife and Band, 2017; Buttle et al., 2019; Zhang et al., 2021b).” (Please see **lines 49-52** in the revised manuscript)

C4: In L58-59 I see that the authors have referenced Wei et al. (2020) and the proposed three-linear hydrologic behaviors. I find this wording hard to follow, which I also address in comments about the abstract. I think it might make more sense to describe this form of rainfall-runoff relationship as having multiple inflections/thresholds with intervening linear segments.

R4: Thanks for your valuable and logical suggestion. The three-linear hydrologic behaviors could be well expressed by the rainfall-runoff relationship as having multiple thresholds with intervening linear segments. In the revised manuscript, the sentence **has been revised as** “Wei et al. (2020) proposed a rainfall-runoff relationship as having multiple thresholds with intervening linear segments to reflect the initial streamflow activation and larger flood response.” (Please see **lines 56-58** in the revised manuscript)

C5: The paragraph spanning L62-89 was very clear and informative. It contrasted with the writing style of earlier paragraphs. I hope that a revised version of the manuscript more broadly applies the tone and writing quality of this section.

R5: Thanks for your serious comment and good suggestion. This Section has been modified and improved in the revised version of the manuscript.

C6: L109-112: I am unsure if this information is a study area description or is an early interpretation of results. Perhaps, it is just the wording, particularly “let to an unstable

trend of the disturbance-response-recovery trajectory....” that is confusing me.

R6: Thanks for you pointing it out. It is an early interpretation of our results in the study area, and has been marked by the reference of Zhang et al., 2021a. The sentence has been revised “After the earthquake, the forest land had a 19.9% shrinkage percentage (Zhang et al., 2021a)”. (Please see lines 110-111 in the revised manuscript)

Additionally, the unclear sentence **has been changed as** “The post-earthquake hydro-geohazards, such as landslides and debris flows, could lead to an unstable recovery trend of landscape vegetation (Figure 1), significantly influencing the stability of hydrologic function and stormflow behaviors of the watershed from rainfall to runoff (Zhang et al., 2021a)”. (Please see **lines 111-113** in the revised manuscript)

References:

Zhang, G., et al. (2021). "Changes in hydrological behaviours triggered by earthquake disturbance in a mountainous watershed." *Science of the Total Environment* 760: 143349.

C7: In L123-125, the authors mention the disturbance recovery process of vegetation and how analyzing this might help better understand runoff generation. I think that this information is critical, and a more detailed process-based description of these relationships would be a welcome addition to the introduction.

R7: Thanks for your suggestion. It is indeed critical for us to reasonably describe the effects of the disturbance-recovery process of vegetation on runoff generation.

The sentence **has been revised as** “After the abrupt disturbance, the exposed bedrock in the trailing edge of the landslides easily induced the Horton overland flow, and the generated loose deposition in the lower part of the landslides generally increased the subsurface stormflow with the microporous flow (Mirus et al., 2017a; Zhang et al., 2018). Such hydrological behaviors are related to the quick runoff generation mechanism with short lag time, resulting in higher runoff potential (Figure 1)” and has been added to the Introduction. (Please see **lines 76-80** in the revised manuscript)

The sentence has been revised as “During the recovery processes, the earthquake-derived

amounts of geohazards affected by large rainstorms led to unstable forest shrinkages and landslide expansions (Figure 1) at long-term timescales in a forest-dominated mountainous watershed. The unstable disturbances from endogenous (earthquake) and exogenous (rainstorms and concomitant hydro-geohazards) origins remarkably increased the uncertainty in the assessment of the hydrological regime from disturbance to recovery and flood risk management (Seidl et al., 2017)” (Please see lines 80-85 in the revised manuscript)

C8: In section 2.4, I was hoping for more details rationalizing the proposed integrated watershed average index for the thresholds. In the discussion, I think that a section should be added to further elaborate on the efficacy of this metric and some introspection about how this metric may or may not be well suited for other environments/conditions where the control factors on the threshold behaviors differ.

R8: Thanks for your valuable suggestions. The *IWA* index mainly considers the processes of runoff generation in the watershed's underlying surface based on the principle and framework of runoff potential for curve numbers (Deshmukh et al., 2013). The underlying surface mainly encompasses the land use types, the shallow storage capacity as well as the physical properties of soils and bedrock at different locations. These factors play vital roles in the runoff generation processes. Another dominant water source of event precipitation amounts in the atmosphere was also taken into account. Therefore, the index is mainly calculated from the area contribution ratio of different land use types (R_i), the shallow water storage capacity at different locations ($DASI_i$), and event precipitation amounts (P). The corresponding sentences have been added.

Additionally, we also presented the applicability and limitations of the proposed detailed metric in the results and discussion Sections. The index was presented and verified through the applications of the magnitudes in two threshold values during the flood hydrograph with 5-min intervals (Figure 5). Meanwhile, in this experimental watershed (Figure 4) a significant negative correlation relationship

($p < 0.05$) between hydrologic thresholds and peak discharges derived from runoff potential was found. Their hydrologic signatures were observed simultaneously, providing an efficient verification of the application of the index in the interannual hydrological variations. Of course, we also acknowledged as a limitation that only the dominant hydrological process of runoff generation was considered while the important confluence flow was mostly ignored. In a future study, such a metric will be involved in the runoff generation and confluence flow to more efficiently reflect the watershed's hydrologic behavior.

Therefore, the sentence “The IWA index mainly considers the processes of runoff generation in the watershed's underlying surface based on the principle and framework of runoff potential for curve numbers (Deshmukh et al., 2013). The underlying surface mainly comprises the land use types, the shallow watershed storage capacity and physical properties in soils at different locations, and bedrock types. These factors play vital roles in the processes of runoff generation. Another dominant water source of event precipitation amounts in the atmosphere is also taken into account.” has been added. (Please see **lines 152-159** in the revised manuscript)

Therefore, the sentence “The threshold index was efficiently verified through the applications of the magnitudes in two threshold values during the flood hydrograph (Figure 5) and their concomitant hydrological variations with the discharges derived from runoff potential (Figure 4). However, we also acknowledge as a limitation that only the dominant hydrological process of runoff generation was considered while the important confluence flow was mostly ignored. In a future study, the such metric will be involved in the runoff generation and confluence flow to more efficiently reflect the watershed's hydrologic behavior.” have been added. (Please see **lines 237-241** in the revised manuscript)

C9: In the discussion, I think some attention should be given to uncertainty in the actual threshold values. I understand that the use of PRA in this context is to characterize the relationship shape rather than to be used in prediction. Still, how robust was the PRA, and are there any concerns about the unequal distribution of

events and leverage from particularly large events?

R9: Thanks for your good comments. It is very important to exploit the uncertainty in the actual threshold values at the watershed scale. For example, the threshold values with large uncertainty could be affected by seasonal and interannual forest transpiration (Scaife et al., 2017). In the revised manuscript, a total of 47 large events in this experimental watershed were identified during periods of June ~ August from 2018 to 2020, possibly filtering out the uncertainty in assessing hydrological behaviors from seasonal variations of the vegetation forest canopy (Hwang et al., 2018). We mainly used the potential mean values of our measured hydrological thresholds from 2018 to 2020 to identify their past long-term changes before and after the Wenchuan earthquake. The changes or uncertainty in thresholds triggered by the earthquake disturbance were mainly considered rather than other factors.

The catchment threshold behaviors were quantitatively assessed using piecewise regression analysis (*PRA*) combined with the Levenberg-Marquardt method and global search optimization algorithm. The standard error of estimate (*SEE*) in multiple regression was estimated considering different locations of the forest, grass-shrub, and landslide lands. These *SEE* values are listed in Table 2. We acknowledge the existence of nonlinear and complex stormflow generation, but automatic searching and calculating the breakpoints and slope parameters using *PRA* with top-down approaches and maximum likelihood approach (Muggeo, 2003) could relatively efficiently determine the emergent hydrological behavior.

The Sentence “A total of 47 events in this experimental watershed were identified during periods of June ~ August from 2018 to 2020, possibly filtering out the uncertainty in assessing hydrological behaviors from seasonal variations of the vegetation forest canopy (Hwang et al., 2018)” was revised. (Please see **lines 138-140** in the revised manuscript)

The Sentence “Uncertainty in visually assessing hydrological thresholds is typically increased by nonlinear and complex stormflow behaviors (Detty and Mcguire, 2010). However, the automatic identification of thresholds and linear slope parameters with a maximum likelihood approach (Muggeo, 2003) could be effective.” was revised. (Please see **lines 145-148** in

the revised manuscript)

Table 2: Comparison of parameters in assessing the three-linear threshold behaviors of $DASI+P$ and Q_q relationships at the confidence level of 95%

Location	Period	Parameters						
		T_g (mm)	T_r (mm)	m_{i1}	m_{i2}	m_{i3}	r^2	SEE
Forest land	Pre-earthquake [#]	111.2	260.7	0.28	0.33	2.36	0.88**	17.17
Grass-shrub land		130.4	247.9	0.21	0.49	1.12	0.84**	15.65
Landslide land	Post-earthquake	91.98	210.48	0.24	0.36	1.04	0.87**	16.54

Note:

m_{ij} indicates the values in the slope parameter of PRA equations from the j_{th} phase at the i land (i =forest, grass-shrub, and landslide lands, j =1, 2, 3 shown in Figure 3).

[#] denotes the collected data in a row, reported by Zhang et al. (2021b).

** indicates that correlation is significant at the 0.01 level (two-tailed).

SEE is the standard error of estimate in multiple regressions.

C10: In Section 3.2, especially in later parts, interpretation and discussion begin to creep in a bit.

R10: Thanks for your pointing it out. Some text “Two thresholds had the opposite trends for these simulated interannual variations in peak discharge and flood volume (Figure 4b), as reported by observed data combined with a hydrological model from Zhang et al. (2021a) in this experimental watershed”, and “This revealed that the lower values in generation and rising thresholds after the earthquake require a lower watershed storage capacity (rainfall and antecedent soil content) input to readily trigger the observed huge flash flood .” has been removed and revised.

C11: I think that a stronger definition of threshold is needed to maintain clarity throughout the manuscript. On the first introduction of the tipping points, I also feel

that a clear distinction should be made so that the reader can more readily determine that different patterns are being assessed.

R11: Thanks for your valuable and logical suggestions. The meanings of the words “threshold” and “tipping point” are very similar, and are not easy for the reader to recognize. We refer to the interpretation from Ali et al., 2013 and some hydrologists you suggest, and characterized the hydrologic threshold as follows.

The observed hillslope- or catchment-scale threshold runoff response (Zehe and Sivapalan, 2009; Fu et al., 2013b; Ross et al., 2021; Wang et al., 2022) shows a hydrologic emergent pattern, which could be used to identify key hydrologic signatures across spatiotemporal scales. The hydrologic threshold is the critical point in time or space at which abrupt changes in stormflow response are observed (Ali et al., 2013). Below the hydrologic threshold, small generated stormflow enters the adjacent channel, but significantly higher runoff magnitudes generally occur above the threshold (Tromp-Van Meerveld and McDonnell, 2006; Zehe et al., 2007; Wei et al., 2020). The corresponding sentences have been added in the introduction part.

In this manuscript, the word “tipping point” is ambiguous, and should be a critical or turning point in time (2011) from 2008 to 2018. It is completely different from the meaning of “threshold” from the words “stormflow threshold behaviors”. We redefined the term “tipping point”, and changed it to “turning time” in the revised manuscript. Therefore, some details in the revised manuscript were revised.

C12:L237-230: The authors describe bedrock depression storage and soil moisture deficit as the main factors controlling a runoff initiation threshold. How do the environments of the referenced studies compare to that of the area in the current study? Are there common characteristics that make this process-based interpretation transferable to this study environment? Figures 6 and 7: I like figures 6 and 7! They were a nice conceptual addition to the manuscript.

R12: Thanks for your appreciation of our work and some good suggestions. Some common characteristics allow transferring such process-based interpretation to

different future studies (Tromp-Van Meerveld and McDonnell, 2006; Fu et al., 2013b; Farrick and Branfireun, 2014; Scaife and Band, 2017; Zhang et al., 2021b; Ross 2021; McDonnell et al., 2021). Firstly, the low permeability of bedrock such as granite and basalt. Secondly, the hillslope is characterized by steep slopes and highly permeable soils. These properties generally lead to a significant soil-rock interface, readily triggering the subsurface stormflow on the interface under heavy rainfall conditions. In the future, it is expected to prepare a reasonably generalized ecohydrological zoning based upon some organizing principles. Such ecohydrological zoning is closely related to climate zones, vegetation types, lithology types, and soil-to-rock depths in the subsurface, topography, landforms, etc. By extending the physical processes which affect the formation and development of flash floods, several dimensionless parameters associated with ecohydrological processes as new metrics could be proposed to characterize the ecohydrological zoning.

The corresponding sentences” Some common characteristics associated with process-based interpretation are highly permeable soils and low permeability of bedrock at the hillslope with steep slopes (Tromp-Van Meerveld and McDonnell, 2006; Farrick and Branfireun, 2014; Scaife and Band, 2017; Ross et al., 2021). These properties generally lead to a significant soil-rock interface, readily triggering the subsurface stormflow on the interface under heavy rainfall conditions.” have been added. (Please see **lines 251-255** in the revised manuscript)

C13: For Section 4.1, controls on threshold behaviors, I found that the author's rationalization of the controls was detailed. With that said, it did read as a mere explanation of different runoff generation mechanisms, and I found there to be a lack of synthesis connecting the experimental observations and analysis results to these more processed-based interpretations. It would be nice if the authors could add some checkpoints in the theoretical explanations to better articulate how their interpretations are supported by their data and how these observations differ from or parallel other studies.

R13: Thanks for your valuable suggestions. It is very important to closely connect

the experimental observations and analysis results to those more processed-based interpretations. In Section 4.1, we analyzed and examined whether the thresholds could separate the initial runoff generation and the flood response using the observed data, including rainfall intensity ($I_{5\text{min}}$), event accumulative precipitation (EAP), and discharge (Q) at a 5-min interval. Additionally, below and above the two thresholds, we attempted to estimate and analyze the changes in minimum contributing area (MCA), stormflow discharge (Q_g), and soil water in different depths at hillslopes based on our collected data (Zhang et al., 2021b). The corresponding sentences were revised as “Below and above the generation and rising thresholds, the changes in minimum contributing area (MCA , with the mean value of 13.79 km², 22.52 km², and 34.43 km², respectively) and stormflow discharge (Q_g , with the mean value of 3.14 mm, 22.5 mm, and 138.3 mm, respectively) are significant (Zhang et al., 2021b; Dickinson and Whiteley, 1970). Higher values of MCA above the rising threshold exceeded 60% of the watershed area (Zhang et al., 2021b), significantly increasing the hydrological connectivity of hillslope riparian-stream and readily triggering catastrophic flash floods (Figure 6e-f).” (Please see **lines 261-266** in the revised manuscript)

In a future study, the indoor potential scaled model test or runoff plots with bedrock depression could be applied to further parse the potential constitutive relationship of rainfall-runoff in different scenarios.

References:

Zhang, G., Cui, P., Gualtieri, C., Zhang, J., Ahmed Bazai, N., Zhang, Z., Wang, J., Tang, J., Chen, R., and Lei, M.: Stormflow generation in a humid forest watershed controlled by antecedent wetness and rainfall amounts, *Journal of Hydrology*, 603, 127107, <https://doi.org/10.1016/j.jhydrol.2021.127107>, 2021b.

Dickinson, W. and H. Whiteley (1970). "Watershed areas contributing to runoff." *IAHS publ 96*: 12-26.

C14: For Section 4.3 point 3. I think that this is an interesting recommendation. Can the authors provide an example of how this could be done? It is a little ambiguous, but I think that this could be a potentially appealing avenue for future work.

R14: Thanks for your good comment and appreciation for future recommendations.

We think that a reasonably generalized ecohydrological zoning with some organizing principles is very necessary. The eco-hydrological zoning is closely related to climate zones, vegetation types, lithology types, and soil-to-rock depths in the subsurface, topography, landforms, etc. By extending the physical processes which affect the formation and development of flash floods, it is expected to propose several dimensionless parameters associated with ecohydrological processes as new metrics to characterize such ecohydrological zoning. If possible, we could better describe region heterogeneity in runoff processes related to flash floods.

C15: Abstract L12: Consider “former hydrologic functioning” rather than “original hydrologic functioning”.

R15: Thanks for your suggestion. The words “original hydrologic functioning” have been changed into “former hydrologic functioning”.

C16: Abstract L16: I am confused by “three-linear stormflow threshold behaviors are examined”, as graphical representations of threshold behaviors are nonlinear. Also, the following segment refers to “both thresholds”, which I also find confusing.

R16: Thanks for your comment. The term “three-linear stormflow threshold behaviors” denotes a nonlinear hydrologic behavior as having two thresholds with intervening linear segments. According to your suggestion from **C3**, the corresponding sentence has been revised as “the nonlinear hydrologic behavior is examined as having two thresholds with intervening linear segments”. (Please see **lines 16-17** in the revised manuscript)

C17: L38-40: This sentence was confusing to me. I interpret these thresholds as emergent patterns or hydrologic signatures that are an integrated representation of processes spanning spatiotemporal scales. If I have correctly interpreted what the authors were aiming for, I do not believe that this is conveyed in their writing.

R17: Thanks for your comment. The sentence has been revised as “The observed hillslope- or catchment-scale threshold runoff response (Zehe and Sivapalan, 2009; Fu et al., 2013b; Ross et al., 2021; Wang et al., 2022) shows a hydrologic emergent pattern, and could be used to identify key hydrologic signatures across spatiotemporal scales (Ali et al., 2013). The hydrologic threshold behavior is the critical point in time or space at which abrupt changes in stormflow response are observed (Ali et al., 2013). Below the hydrologic threshold, a small generated stormflow enters the adjacent channel, but significantly higher runoff magnitudes generally are observed above the threshold (Tromp-Van Meerveld and McDonnell, 2006; Zehe et al., 2007; Wei et al., 2020).”. (Please see **lines 37-43** in the revised manuscript)

C18:L40-44: I understand the intent of this sentence, but I found the wording unusual. Consider revising for clarity.

R18: Thanks for your suggestion. The sentence has **been revised to** “A unified threshold-based hydrological theory that possibly advanced catchment hydrology was extensively discussed during the AGU 2011 Fall Meeting (Ali et al., 2013), and later was continuously developed (Ross et al., 2021; Ross, 2021; Ali et al., 2015; Scaife et al., 2020). Theoretical advancements in hydrology can support the development of appropriate algorithms for more efficient predictive models.”. (Please see **lines 42-45** in the revised manuscript)

References:

- Ross, C. A. (2021). *Moving towards a unified threshold-based hydrological theory through inter-comparison and modelling.*
- Ross, C. A., et al. (2021). "Evaluating the Ubiquity of Thresholds in Rainfall-Runoff Response Across Contrasting Environments." *Water Resources Research* 57(1).
- Scaife, C. I., et al. (2020). "Non-linear quickflow response as indicators of runoff generation mechanisms." *Hydrological Processes* 34(13): 2949-2964.
- Ali, G., et al. (2015). "Comparison of threshold hydrologic response across northern catchments." *Hydrological Processes* 29(16): 3575-3591.
- Ali, G., et al. (2013). "Towards a unified threshold-based hydrological theory: necessary components and recurring challenges." *Hydrological Processes* 27(2): 313-318.

C19:L44-46: Ambiguous wording. I suggest providing a concise definition of the

threshold behavior in the runoff response being referred to. The Ali et al., 2013 reference provided, offers one such definition.

R19: Thanks for your logical suggestion. According to the research from Ali et al., 2013 and some hydrologists, we have analyzed and defined the threshold behavior in the runoff response. The following sentences **have been revised and added** as “The observed hillslope- or catchment-scale threshold runoff response (Zehe and Sivapalan, 2009; Fu et al., 2013b; Ross et al., 2021; Wang et al., 2022) shows a hydrologic emergent pattern, and could be used to identify key hydrologic signatures across different spatiotemporal scales (Ali et al., 2013). The hydrologic threshold behavior is the critical point in time or space at which abrupt changes in stormflow response occur (Ali et al., 2013). Below the hydrologic threshold, a small stormflow enters the adjacent channel, but significantly higher runoff magnitudes generally are observed above the threshold (Tromp-Van Meerveld and McDonnell, 2006; Zehe et al., 2007; Wei et al., 2020).”. (Please see **lines 37-43** in the revised manuscript)

References:

- Zehe, E. and M. Sivapalan (2009). "Threshold behaviour in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications." *Hydrology and Earth System Sciences* 13(7): 1273-1297.
- Fu, C., et al. (2013). "Threshold behavior in a fissured granitic catchment in southern China: 1. Analysis of field monitoring results." *Water Resources Research* 49(5): 2519-2535.
- Ross, C. A., et al. (2021). "Evaluating the Ubiquity of Thresholds in Rainfall-Runoff Response Across Contrasting Environments." *Water Resources Research* 57(1).
- Wang, S., et al. (2022). "Rainfall-runoff characteristics and their threshold behaviors on a karst hillslope in a peak-cluster depression region." *Journal of Hydrology* 605.
- Ali, G., et al. (2013). "Towards a unified threshold-based hydrological theory: necessary components and recurring challenges." *Hydrological Processes* 27(2): 313-318.
- Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope, *Water Resources Research*, 42, 336-336, 10.1029/2004WR003778, 2006.
- Zehe, E., Elsenbeer, H., Lindenmaier, F., Schulz, K., and Blöschl, G.: Patterns of predictability in hydrological threshold systems, *Water Resources Research*, 43, 10.1029/2006wr005589, 2007.
- Wei, L., Qiu, Z., Zhou, G., Kinouchi, T., and Liu, Y.: Stormflow threshold behaviour in a subtropical mountainous headwater catchment during forest recovery period, *Hydrological Processes*, 34, 1728-1740, 10.1002/hyp.13658, 2020.

C20:L46-48: It is unclear to me why threshold is plural in this sentence – I also think that this information can be incorporated into the former sentence where I have suggested clearly defining the author’s operational definition of threshold behavior.

R20: Thanks for your serious comment. The word “threshold” has been revised to be singular. Additionally, the clear interpretation and definition of threshold behavior were illustrated in the fore-mentioned **R19**. Additionally, the information can be incorporated into the former sentence (Please see R19).

C21:L50: “They might indicate...” is vague. Are the authors referring to the different diagnostic shapes or the transition from pre-threshold to post-threshold behavior?

R21: Thanks for your serious comment. The word “They” is the transition from below-threshold to above-threshold behavior for different diagnostic shapes. Therefore, the sentence **has been revised** as “The transition from below-threshold to above-threshold behavior for different diagnostic shapes suggests several mechanisms of water retention and release in the watershed”. (Please see **lines 48-49** in the revised manuscript)

C22:L50: “The runoff behaviors....”. Are the authors referring to the thresholds in the cited literature or in the current study?

R22: Thanks for your logical comment. These thresholds are from the cited literature. The sentence **has been revised** as “In the literature, the runoff behaviors with Hockey stick shape were found at the hillslope (Tromp-Van Meerveld and McDonnell, 2006; Fu et al., 2013a; Wang et al., 2022) and watershed scales (Wei et al., 2020; Farrick and Branfireun, 2014; Scaife and Band, 2017; Buttle et al., 2019; Zhang et al., 2021b).”. (Please see **lines 49-52** in the revised manuscript)

C23:L78-L79: I found this hard to follow and after reading it multiple times am not sure about the intended messaging.

R23: Actually, the exposed bedrock in the trailing edge of the landslides easily

stimulated the occurrence of the Horton overland flow, and the generated loose deposition in the lower part of the landslides generally motivated more subsurface stormflow with the microporous flow. The two hydrological behaviors are related to the quick runoff generation mechanism with short confluence time, resulting in higher runoff potential. Therefore, the corresponding sentences have been revised as “After the abrupt disturbance, the exposed bedrock in the trailing edge of the landslides easily induced the Horton overland flow, and the generated loose deposition in the lower part of the landslides generally increased the subsurface stormflow with the microporous flow (Mirus et al., 2017a; Zhang et al., 2018). Such two hydrological behaviors are related to the quick runoff generation mechanism with short confluence time, resulting in higher runoff potential”. (Please see **lines 76-80** in the revised manuscript)

References:

Mirus, B. B., et al. (2017). "Hydrologic Impacts of Landslide Disturbances: Implications for Remobilization and Hazard Persistence." *Water Resources Research* 53(10): 8250-8265.

Zhang, J., van Meerveld, H. J., Tripoli, R., and Bruijnzeel, L. A.: Runoff response and sediment yield of a landslide-affected fire-climax grassland micro-catchment (Leyte, the Philippines) before and after passage of typhoon Haiyan, *Journal of Hydrology*, 565, 524-537, 10.1016/j.jhydrol.2018.08.016, 2018.

C24:L119-L120: Please clarify.

R24: Thanks for your suggestion. The unclear sentence has been revised to “Subsurface stormflow generated on the soil-bedrock interface under heavy rainfall conditions is one of the dominant runoff sources contributing to flash flooding”. (Please see **lines 120-122** in the revised manuscript)

C25:L129: why is volumetric soil moisture content italicized?

R25: Thanks for your suggestion. The font “volumetric soil moisture content” is not italicized, and has been revised.

C26:L129-131: this sentence is unclear/ hard to follow.

R26: Thanks for your suggestion. The sentence has been revised to “Each probe equipped with eight sensors at a 10 cm depth interval was installed 80 cm in soil profiles below the surface (Figure 2a).”. (Please see **lines 130-131** in the revised manuscript)

C27:L132-134: tense changes from rest of paragraph.

R27: Thanks for your pointing it out. It is revised to be simple past tense and is consistent with the tense from the above sentences.

C28:L139-140: please include in-text which method of baseflow-stormflow separation was used.

R28: Thanks for your serious comment. The two-parameter recursive digital filter method was used to separate the quick flow (i.e. stormflow) and delayed flow (i.e. base flow) from total discharge for the storm runoff events (Eckhardt, 2005). The method has been illustrated in the revised manuscript. (Please see **lines 140-141** in the revised manuscript)

References:

Eckhardt, K., 2005. How to construct recursive digital filters for baseflow separation. Hydrol. Process. 19 (2), 507–515

C29:L193: “non-stationary” rather than “non-stationarity”.

R29: The word “non-stationarity” has **been revised as** “non-stationary”.

C30:L203-205: I found the first half of this sentence difficult to understand/follow.

R30: The unclear sentence has been removed, and changed to “A significant negative correlation relationship ($p < 0.05$) between hydrologic thresholds and peak discharges was observed in this experimental watershed (Figure 4b).”. (Please see **lines 212-213** in the revised manuscript)

C31:L271: “severally” rather than “several”?

R31: The word “several” has been revised to “severally”.

C32:L274-277: I am not sure what the authors are saying in this sentence.

R32: It was really unclear. The sentence mainly introduced that earthquake disturbance induced spatially uneven distribution and dynamic nonstationarity at timescales of landslide patches at watershed scales, such as the back-slope effect, hanging wall effect, etc. These processes generally have a large limitation in accurately assessing runoff generation and the dynamic evolution of catastrophic flash flood disasters. In the revised manuscript, the paragraph has been revised to be “This phenomenon highlighted the importance of spatially uneven distribution and dynamic nonstationarity at timescales of earthquake-induced landslide patches for an accurate assessment of the runoff generation and the dynamic evolution of catastrophic flash flood.” (Please see **lines 294-296** in the revised manuscript)

C33:L305: “from2011” missing space.

R33: Thanks for your reminder. It has been revised.

C34:L306: “It triggered...” what are the authors referring to as “it” in this context?

R34: Thanks for your comments. The paragraph “It triggered...” was unclear, and **has been revised to be** “During the period, the *RP* and *VSA* zones rapidly expanded while the hydrological threshold behaviors were quickly recovered and improved”. (Please see **lines 325-326** in the revised manuscript)

C35:L310: I do not understand the messaging of the Section 4.3 title.

R35: The title in **Section 4.3** has been revised to “Challenges in Identifying Nonunique Threshold Behaviors disturbed by Large Disaster Events”.

C36: L320: The spatial patchiness of which characteristics?

R36: Herein, the spatial patchiness within the watershed is mainly triggered by sudden disaster events. It possesses these characteristics. The landscape vegetation could be impaired. The forest canopy and vegetation-soil system generally could be destroyed, facilitating the reduction of the canopy interception and shallow soil water storage capacity. With the expansion of the broken patchiness, the structural hydrological connectivity is rapidly enhanced, accelerating the confluence process and stormflow generation. The words “spatial patchiness” have been revised into “spatially broken patchiness”. (Please see **line 340** in the revised manuscript)

Response to Reviewer #2 (Band in University of Virginia) Comments:

C1: Zhang et al. present an interesting study of stormflow runoff threshold non-stationarity over a time-line before and following a major earthquake in the eastern periphery of the Tibetan Plateau. The earthquake resulted in a massive disturbance of the dominant forest cover due to extensive landsliding which subsequently expanded with monsoon-initiated landslide growth, then slowly began to recover with revegetation, and presumably, renewed colluvial infilling of scars.

The paper provides a good illustration of specific controls of non-stationary threshold behavior in response to geomorphic disturbance and a chronology of ecosystem recovery. This adds to our knowledge of storm event-based threshold behavior with good evidence of the watershed system dynamics and time scales of adjustment. It may be argued that this is an end-member in terms of magnitude of disturbance, but may be increasingly applicable to other cases of large, sudden land use change and slow recovery due to devastating storms, fires, or other disasters.

R1: We are very grateful for having the summary of your positive assessments of our work. Each comment has been addressed below point-by-point. We also hope that this study about the relationship between hydrological sciences and flash flood disasters could be considered for publication in the “*Hydrology and Earth System Sciences*”.

C2: The documentation of the threshold stormflow behavior is interesting, but there are a set of areas in the text that are unclear. Specifically, the methods need to be clarified. Otherwise, some of the interpretation and conclusions may appear to be more qualitative and speculative, and not specifically supported by the data.

R2: Thanks for your serious comments. the methods in the revised manuscript have been clarified in the revised manuscript and the following **R3** in the revised manuscript.

The sentences “To clearly understand long-term threshold evolutions and emergent behaviors variations pre- and post-earthquake at the watershed scale, an integrated watershed average (IWA) index for the thresholds was proposed to characterize the watershed stormflow emergent behaviors. The IWA index mainly considers the processes of runoff generation in the watershed's underlying surface based on the principle and framework of runoff potential for curve numbers (Deshmukh et al., 2013). The underlying surface mainly comprises the land use types, the shallow watershed storage capacity and physical properties in soils at different locations, and bedrock types. These factors play vital roles in the processes of runoff generation. Another dominant water source of event precipitation amounts in the atmosphere is also taken into account.” have been revised. (Please see Lines 150-156 in the revised manuscript)

The sentence “The threshold index was efficiently verified through the applications of the magnitudes in two threshold values during the flood hydrograph (Figure 5) and their concomitant hydrological variations with the discharges derived from runoff potential (Figure 4). However, we also acknowledge as a limitation that only the dominant hydrological process of runoff generation was considered while the important confluence flow was mostly ignored. In a future study, the such metric will be involved in the runoff generation and confluence flow to more efficiently reflect the watershed's hydrologic behavior.” have been added. (Please see Lines 237-241 in the revised manuscript)

C3: Figure 3 is a major result and contributes prominently to the conclusions. However, there do not appear to be sufficient observations to separate out the highest thresholds and trend with statistical significance as it appears this is determined by a single, large event. It is also not clear from the methods whether discharge was separately measured or determined for grass shrub, forest, and landslide areas. The position of the gauges suggests each drainage area is a mixture of all three land covers, and more information is required to see how each land cover contribution is deconvolved. Add more detail to this discussion. If separate measurements were not made it is not clear how these piecewise regressions were made. If this is done by

modeling using curve numbers or HEC-HMS this should be clear.

R3: Thanks for your serious comments and valuable suggestions. Some results in Figure 3 indeed are important contributions to our conclusions in the revised manuscript. A sufficient amount of observed hydrological data is very significant to identify the stormflow threshold behaviors with good statistical significance. In the revised manuscript, 47 rainfall-runoff events ($P > 4\text{mm}$) in our study area were collected and used to analyze the hydrological behaviors at watershed scales. However, in a future study, a larger amount of meteorological and hydrological data will be collected to improve the statistical significance of the data and the accuracy of data analysis.

The best option would be to simultaneously collect the event precipitation amounts (P), $DASI$, and discharge at a separate forest, grass-shrub, and landslide land. Actually, it is difficult to do that. In the revised manuscript, in the forest, grass-shrub, and landslide lands, P and $DASI$ were obtained for different land use while the discharges were measured at the gauging station S6. We mainly consider the $P + DASI$ contributions of each land use to flow discharge in each storm event using observed field data. The potential scaled model test or runoff plots at different locations of land use could be applied in the future. Additionally, the hydrological outputs derived from the HEC-HMS model were presented by Zhang et al., (2021b). Some results were compared with our observed flood events and discussed in the revised manuscript

Reference:

Zhang, G., P. Cui, W. Jin, Z. Zhang, H. Wang, N. A. Bazai, Y. Li, D. Liu, and A. Pasuto (2021), *Changes in hydrological behaviours triggered by earthquake disturbance in a mountainous watershed*, *Science of The Total Environment*, 760, 143349.

C4: Finally Figure 3 is difficult to interpret as all data points have the same symbol and color, over all land uses. Either color code or use a different symbol so the reader can assess the degree of separation between the trends. Clarifying the statistics provided would also help. There is a composite r^2 provided in table 2 for each of

three distinct land uses, including two thresholds and three slopes. While the overall correlation is significant. What is the confidence in each of these parameters? Is it possible to provide SEE for each? I presume this may not be possible for the highest flow slope values if they were supported by a single large storm observation.

R4: Thanks for your serious comments and valuable suggestions. Figure 3 has been modified as follows. Each color code and symbol point is presented at different locations of land uses. Each of these parameters was analyzed at the confidence level of 95%, and the Standard Error of Estimate (*SEE*) for each in multiple regressions is listed in Table 2. In the future, a larger amount of data will be collected and analyzed to better identify the threshold behaviors at hillslope and watershed scales in our experimental watershed.

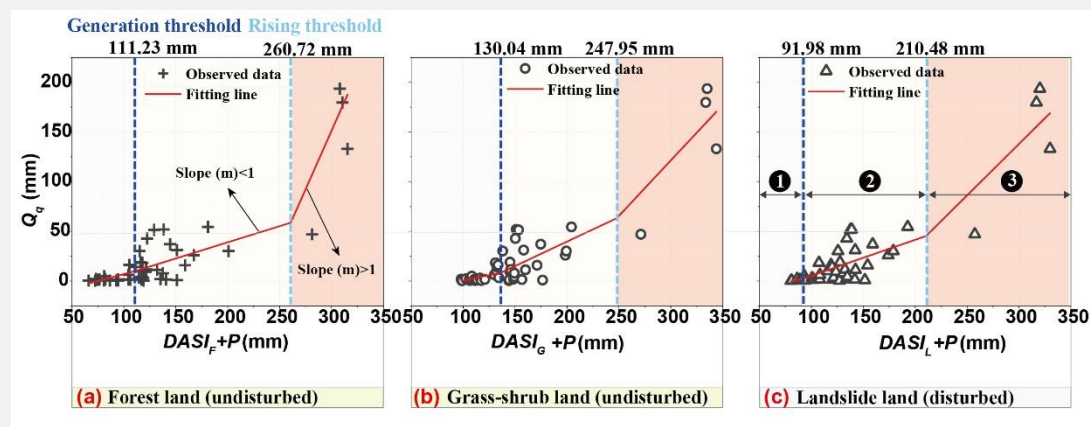


Figure 3: The piecewise regression analysis of event stormflow amount (Q_q) plotted against the sum of event precipitation amounts (P) and $DASI$ at the forest (a), grass-shrub (b), and landslide (c) lands. The undisturbed forest and grass-shrub lands represent the pre-earthquake period, as reported by Zhang et al. (2021a), and the disturbed landslide land represents the post-earthquake period. Red lines indicate the linear fitting for the piecewise regression for the variable of $P + DASI$ at the confidence level of 95%.

Table 2: Comparison for parameters in assessing the three-linear threshold behaviors of $DASI + P$ and Q_q relationships at the confidence level of 95%

Location	Period	Parameters					r^2	<i>SEE</i>
		T_g (mm)	T_r (mm)	m_{i1}	m_{i2}	m_{i3}		
Forest land	Pre-earthquake [#]	111.2	260.7	0.28	0.33	2.36	0.88**	17.17
Grass-shrub land		130.4	247.9	0.21	0.49	1.12	0.84**	15.65

Landslide land	Post-earthquake	91.98	210.48	0.24	0.36	1.04	0.87**	16.54
----------------	-----------------	-------	--------	------	------	------	--------	-------

Note:
 m_{ij} indicates the values in the slope parameter of *PRA* equations from the j_{th} phase at the i land (i =forest, grass-shrub, and landslide lands, $j=1, 2, 3$ shown in Figure 3).
denotes the collected data in a row, reported by Zhang et al. (2021b).
** indicates that correlation is significant at the 0.01 level (two-tailed).
SEE is the standard error of estimate in multiple regressions.

C5: The authors cite the scarcity of measurements pre-earthquake, and the logistical difficulty of accessing areas post-earthquake as limiting the information available to assess stormflow threshold behavior through this time. Some information is derived from simulation modeling, developed in a previous paper. More information on the number of actual measurements, the information provided by the HEC-HMS model, and its reliability should be provided. The authors point to a specific “tipping point,” after which the stormflow thresholds begin to increase again. Are these based on land cover change derived curve numbers within the model, and are there discharge measurements sufficient to verify these changes? In figure 4b we see peak discharge for a set of events first increase and then decrease as the forest ecosystem begins to recover. How are these peak discharges adjusted for the size of the storm, or are they averaged from a larger number of events?

R5: Thanks for your serious comments and valuable suggestions. In our previous study (Zhang et al., 2021b), based on 5 min time-series data in rainfalls (9 rainfall stations) and streamflow (2018–2019), the HEC-HMS model was calibrated and validated to predict the historical (2007–2018) hydrological behaviors. The mean Nash–Sutcliffe efficiency was 0.76, showing good model performance and reliability. But some hydrological data pre- and post-earthquake are indeed scarce. This might be a limitation.

Herein, the stormflow threshold behaviors, including integrated watershed generation threshold (T_{g-IWA}) and rising threshold (T_{r-IWA}), were observed and calculated using equations (1) rather than the curve numbers within the model. The historical (2007–2018) flood response was predicted via the curve numbers within

the model. The scarcity of continuous runoff measurements from 2007-2018 pre- and post-earthquake might be a limitation.

The trends of values in peak discharges and flood volumes under different sizes of storm were consistently based on our previous study from Zhang et al., 2021b). In the revised manuscript, we selected the long-term changes in a high-magnitude flood event ($>10^2$) to analyze and compare the changes in T_{g-IWA} and T_{r-IWA} . It is efficient for us to compare the changes between stormflow threshold behaviors and flood response. The corresponding text has been revised as “Changes in (a) observed stormflow threshold behaviors, including the integrated watershed generation threshold (T_{g-IWA}) and the rising threshold (T_{r-IWA}), and (b) a large flood event response selected from Zhang et al. (2021a) during the periods of 2007 ~ 2018 before and after the earthquake, including peak discharge (Q_p) and flood volume (V).” (Please see Lines 220-222 in the revised manuscript)

Reference:

Zhang, G., P. Cui, W. Jin, Z. Zhang, H. Wang, N. A. Bazai, Y. Li, D. Liu, and A. Pasuto (2021), *Changes in hydrological behaviours triggered by earthquake disturbance in a mountainous watershed*, *Science of The Total Environment*, 760, 143349.

C6: The analysis of threshold behavior shown in figure 5 is presented in the discussion section. I think this should be in the results section, then discussed/interpreted in the discussion section.

R6: Thanks for your suggestion. The analysis of threshold behavior shown in figure 5 mainly examines and verifies the occurrence of initial streamflow at T_g and large flood response at T_r . It is an important interpretation and discussion of the values of obtained hydrological thresholds from Section 3 Results.

C7: Reword so it is clear that it was estimates there were roughly 2×10^5 landslides initiated (which is amazing)

R7: Such number of roughly 2.0×10^5 landslides following the Wenchuan earthquake was identified by Xu et al., (2014) and Fan et al. (2018). The references are as

follows:

Xu, C., et al. (2013). "Three (nearly) complete inventories of landslides triggered by the May 12, 2008 Wenchuan Mw 7.9 earthquake of China and their spatial distribution statistical analysis." Landslides 11(3): 441-461.

Fan, X., et al. (2018). "What we have learned from the 2008 Wenchuan Earthquake and its aftermath: A decade of research and challenges." Engineering Geology 241: 25-32.

C8: Line 107, the term indeciduous is not clear. Remove and simply call the canopy conifer.

R8: Thanks for your comments. The term indeciduous has been removed, and revised to “*canopy conifer*”. (Please see Line 110 in the revised manuscript)

C9: Line 168-170, sentence may be better placed either in the introduction or discussion. It is not a specific result of the analysis done here.

R9: Thanks for your good suggestion. The sentence has been placed in the introduction. (Please see Lines 103-105 in the revised manuscript)

C10: Line 261 – I presume you mean the time to peak decreased not increased following the earthquake?

R10: Thanks for your logical comments. You are right. The time to peak is decreased by 25min following the earthquake. The sentence has been revised to be “Peak discharges increased by 22.58%~367.42% and the time to peak was advanced by 25 min”. (Please see Lines 280-281 in the revised manuscript)