

While the topic of RoS is certainly interesting and of global relevance, I find that this analysis lacks firm roots in compelling, generalizable research questions and interpretation of the results in the context of hydrologic processes. Furthermore, while the identification of criteria for selecting RoS events is compelling, the approach applied herein appears very arbitrary, without justification or validation. Additional weaknesses relate to inadequate description and defense of the methods.

General comments

- Study questions are not terribly compelling or generalizable, and the Results don't clearly map to the study questions. It appears that some interesting study questions might be embedded in the Results, but the authors need to identify them explicitly and then reorganize the Results and Discussion around those student questions.
- The authors regularly reference flood risk (e.g. page 17, lines 15-16), but their analysis does not actually consider flood risk. Flood risk is characterized by a probability of an event occurs AND the consequence of that event (e.g. loss of life, economic damages). I believe this seriously weakens the paper to use flood risk as a justification and then to not actually consider flood risk. While I am not suggesting that evaluating flood risk is necessary for this paper as it can require a large effort, I am suggesting that it is not a valid basis on which to justify this study. Instead, the figures could contribute more by relating the variability in RoS events to hydrologic processes, which could be transferrable to other areas and time periods.
- Results: This section reads as a blow-by-blow (or day-by-day) account of precipitation depths, which makes it dull to read and difficult to identify the key findings. Why should the reader retain any of this content? The authors should rewrite the section, pulling out only the key findings so the reader doesn't have to wade through all of the non-essential details to understand the key points. Unfortunately, because the Methods and Results were so difficult to read, I could not easily assess the validity of the interpretations.
- The interpretation of the Results lacks any deep consideration of hydrologic processes. Aside from elevation, what explains the variability that the authors are reporting between sub-basins? For example, there is no discussion on how geology, land use, or connectivity to groundwater varies between the basins. To say (e.g. page 14, line 20) that a decreasing trend in RoS events for a single basin tells the reader very little that can be applied elsewhere.
- A major element of this paper is proposing criteria for identifying a RoS event, but I find the criteria to be arbitrary and their evaluation to be missing. For example, in Section 2.4: The authors give very vague set of thresholds (bottom of page 8), with a reference to Kohn et al. 2013, for defining RoS events. Despite the fact that these may be published elsewhere, more details, even briefly, are needed regarding how those thresholds were developed. Rather than simply saying "significantly" or "substantial," as was given (but not justified) in the abstract. The authors come back to this on page 12 (lines 5-7) and in the

Discussion (page 17), but there is no justification for using 10mm SWE, 20% snowmelt, and 3 mm rainfall. Instead, it appears very arbitrary. For example, couldn't a RoS event occur with 8mm of SWE, 40% snowmelt, and 4 mm rainfall? Just because the Jan. 2011 event had those characteristics doesn't mean that all RoS events should have those same characteristics. To say that they are the characteristics of this one event doesn't make them "good indicators for RoS driven flood events" (page 17, line 23). They also state (page 12, lines 14-15) that historic RoS events were well captured by the given criteria, but provide no evidence of what it means for these events to be "well caught." The idea that these can be transferred to other basins (page 17, line 27) is totally unfounded on any hydrologic processes. How might these thresholds change for areas with different vegetation conditions, connectivity to groundwater, albedo, etc.? Simply adding a paragraph into the Discussion (page 20, lines 17-24) acknowledging that physical processes may be important is not enough – these processes should be discussed in the interpretation of the results. And I disagree with the authors that a model is needed. Even some basic background information on the sub-basin hydrogeology and land use, which is used in interpreting the variable responses between the basins, would be helpful. Thus, these criteria need to be evaluated with some fitness measures. So say they have been validated (page 21, line 1) is totally inappropriate.

- The Introduction reads like a literature review, not a statement of the status quo and justification for this study. Also, consistent with the rest of the text, there is no discussion of mechanisms or processes in the Introduction.

Technical issues

- Page 3, line 7 : It is no longer accepted to use global warming. Please change to "climate change."
- Page 3, line 11 (and throughout text) : I believe rainfall is easier to understand than "liquid precipitation."
- Page 4, line 18 : I don't follow what "at least one of the four daily precipitation measurements" means. What are the four measurements?
- Figure 1 – it would be helpful to add a delineation of the sub-basins. The location of the gauges isn't really enough to show the size and shape of the basins. Also, if there is any way to show topography, maybe as an inset figure, this would also help understand the basins' drainage patterns. Also, please give the elevation ranges for the three HAD zones, either in the figure caption or in the text. The elevations referenced within the text (bottom of page 5, top of page 6) do not seem to map to Figure 1 at all.
- Page 5, line 22 – "only a small part of Southern Germany" – please be specific regarding how much is a "small part" by giving the area. This is important because it may influence the reliability of the results for the alpine sub-basin type.

- Page 5, line 24 – the reference to Zugspitze seems out of place since it isn't on the map anywhere. Suggest deleting.
- Page 6, line 14: For brevity... add reference to Figure 2 after "300 gauging station (Figure 2)" then delete the last sentence of this paragraph.
- Page 6, lines 17-19: this sentence on quality control has grammatical issues and is confusing. What criteria were used for suitability in this study? And what kind of post-flood corrections were made? Since this study emphasizes floods, substantially more information is needed on how "minor" the corrections were, how many corrections were made, and how they were made.
- Please be sure to define all variables (e.g. Ta, Mf) within the text.
- The temperature index model is quite simple and there has been a pretty extensive body of work describing the issues with it. The authors should at least acknowledge this body of literature and some of the limitations of the model, then justify why those limitations are not important to answering the questions posed here. Furthermore, the justification given, based on Ohmura (2011) that the model is "sufficiently accurate for most practical purposes" is far too vague.
- Applying the constant value of $M_f = 3\text{mmC}^{-1}\text{day}^{-1}$ seems like a pretty substantial assumption that is only qualitatively justified with a statement that snow melts quickly in Germany. Can the authors think of a more compelling (perhaps quantitative?) way to verify that this is a valid assumption? Did the authors perform the sensitivity analysis for M_f as they did for T_b ?
- Page 7, Line 20: When and who did the "before" work to demonstrate the lack of sensitivity to T_b ?
- Page 7, line 26: What does "from 2 to 1 August of the following year" mean? Also, does your hydrologic year start on August 1? If so, why? I thought hydrologic year always started on Oct. 01.
- It isn't clear to me how the 300 gauges were used, since only 12 gauges were used to define the relationship between discharge and precipitation depth.
- Could table 1 be more quantitative? That is, what does "flooding all over the river" mean? What is a "large floodings" or "large scale floodings" relative to "small floodings?" Ideally, the authors would give area or return year intervals to qualify what large or small is. Also note that flooding should be singular, not plural. Floods can be plural, but flooding cannot.
- Page 9, line 20: please be more specific about what "trends were calculated." The relationships between what? And why were these calculated? I don't see how they fit into the study objectives. Also, why do the authors conclude that the relationships are linear? This assumptions needs to be justified. Finally, what is the expected value based on?
- Page 9, line 24: "...with the others" what? Other basins or other parameters?
- Figures 2a-f: the inset figures are too small to read. The figures either need to be larger or simpler to make them easier to interpret.
- Page 11, lines 7-9: The topic sentence for this paragraph is a run on and has grammatical errors.

- Page 11, lines 9-13: Also a run on sentence with grammatical errors.
- Page 12, lines 7-9: I don't follow how the 6 day duration of the Jan 2011 storm leads to "the basin response time for the selection of RoS events." How does this help select RoS events? And how is this used in the analysis?
- Figure 3 needs further interpretation. What is the difference between selected and historic events? What does a negative correlation (3j) indicate? What can the authors say more *mechanistically* about why discharge from the alpine and upland sites would be better correlated with equiv. precip. depth?
- Figure 4: in the text, the authors need an explanation of the relevance of the quantile values. What does a value of 1.0 or 0.7 mean? And what is the direct relationship between quantiles and "flood-causing potential?" I find this assertion to be unsupported by the analysis and the text.
- Page 13, lines 12-13: "amplifies the observation in Fig. 3 that RoS events do not necessarily cause the highest floods." I don't see this in Figure 3, and don't see where it is explained in the text. Can the authors more explicitly point out where they see evidence of this in Figure 3?
- Page 14, line 3: "decreasing trend" of what?
- Figure 6: The red star indicating significant trends is not very visible in black and white.
- Figure 7: It is generally not acceptable statistically to evaluate discharge records for <30 years. Please justify the selection of this short period of time. It is particularly important since the authors are finding such radically different trends for the long-term vs. short-term periods.
- Page 16, lines 18-20 : typo/grammar issues
- Page 20, lines 4-16 : I don't find this justification compelling at all, especially given the very brief (1 sentence) explanation of the sensitivity analysis (which was performed only for Tb, right?). I am thus not sure that any of the trend analysis is terribly insightful.