Response to editor and reviewers

We would like to thank the editor and the reviewers for the re-review of our paper and provide below our answers to the overall comment of the editor and to the additional comments of reviewer #1. Our response is in normal, orange font, the comments are in italic.

Editor comment

Dear Authors,

The two referees who reviewed the revised manuscript are generally happy with the improvements made. Referee 1 raises a couple of comments that remain to be addressed.

A specific comment for the following points:

- title: the reviewer makes a nice title suggestion, consider changing accordingly.

Please see our title suggestion below in the response to the reviewer.

- referee makes a comment about the radar plot and alternative figures you provided. I agree with the reviewer that radar plots are difficult to interpret and a different format would be preferred. Yet I had trouble locating the alternative figures A and B you are referring to in your response. Please check and consider an alternative format to the radar plots.

The figures A and B are referring to the scatter plots from our answer to the point 1b in our merged author-reply document from the public discussion. Thanks for pointing this out. We decided to adapt those figures. Please see below the Figure 4 in our response to the reviewer.

After reading the manuscript, your first discussion point keeps puzzling me. Only 1 of the streamflow-generating events is asymmetric, yet asymmetry comes out as an important metric to explain variation in RC and lag time. I wonder how such regression results can be valid if the explanatory variable shows so little variability?

Thanks for pointing out the slightly misleading formulations. This comment refers to this the part of the manuscript in "5.1 Spatial heterogeneity of rainfall" copied hereafter (L536-542):

"Among the records showing a strong rainfall asymmetry, 7 out of the 8 events are too small to cause a detectable streamflow response (Figure 5), but one does create a streamflow response although it only rains over half of the 12 rain gauge stations. Despite of this absence of a strong asymmetry in the 14 rainfall events that cause a streamflow response, the regression analysis suggests that the spatial distribution might play an important role for the explanation of the lag time. The importance of this asymmetry predictor can be related to the fact that it captures the key feature of the spatial catchment organisation in terms of distance to the outlet, drainage density and subsurface storage potential."

What we called "asymmetric" is based on the threshold of $|I_{ASYM}| > 0.33$ which indicates that one part of the catchment received at least 2 times more rainfall than the other part; but I_{ASYM} is a value between -1 and 1 that describes the spatial rainfall distribution even when $|I_{ASYM}| < 0.33$. The Figure 1 shows the distribution of I_{ASYM} for the 48 rainfall events. Among 8 events associated with a streamflow reaction, 7 have an I_{ASYM} value between -0.33 and 0.33, but yet not equal to 0.

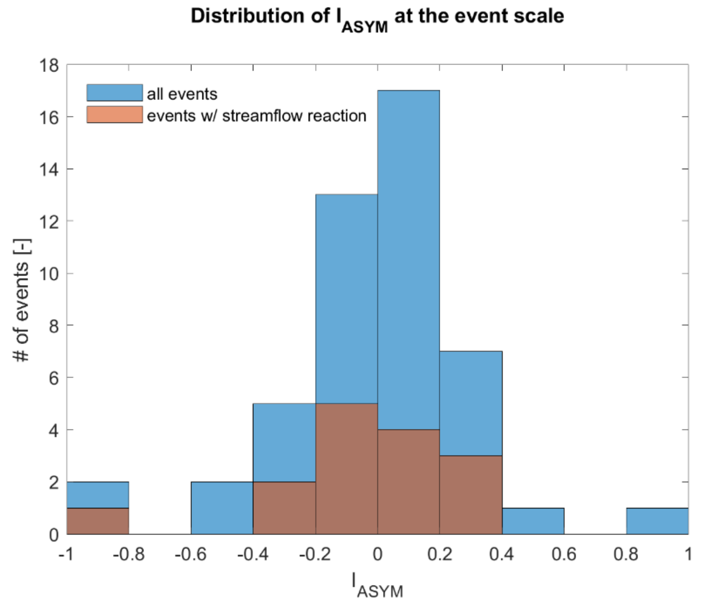

Figure 1. Distribution of I_{ASYM} for the 48 rainfall events (blue) and the subset of rainfall events associated with a streamflow reaction (brown).

We adapted slightly the formulation (L200-203) in "3.2.2 Spatial rainfall pattern metrics" by changing *"We consider a rainfall event as asymmetric when at least 2 times more rain has precipitated over one part of the catchment than over the other, i.e. when I_{ASYM} is below -0.33 or above +0.33"*. It now reads as: *"A value of -0.33 or 0.33 indicates that the catchment received at least 2 times more rain over one part of the catchment than the other."*

And in "4.1.1 Areal rainfall and asymmetry" by changing (L347-348) "*Interestingly, strong spatial asymmetry mainly affects events with low rainfall amounts, with 7 out of 8 asymmetric events (when $|I_{ASYM}| > 0.33$) receiving below 5 mm (Figure 5). "* into "*Interestingly, strong spatial asymmetry mainly affects events with low rainfall amounts, with 7 out of 8 events with $|I_{ASYM}| > 0.33$ receiving below 5 mm (Figure 5)".*

Quite a lot of papers have looked at this relation between rainfall spatial variability (storm position, movement, relative to the catchment and streamflow network), some of which you cite in the Introduction, yet it's worth putting your results in relation to the literature also in the Discussion part. The one paper you cite here is a modeling study, while there are several data-driven studies on the topic of "storm position and hydrologic response" that are more closely related to your study (see for instance work by J.A. Smith and his group).

Thank you for pointing this out. We added one sentence introducing 2 references in the discussion part at the end of the "5.1 Spatial heterogeneity of rainfall" (L568-571):

"Finally, we would like to point out here that this result on the prominent role of travel time along the stream network opens interesting new analogies with urban hydrology, where introduction times to the network are typically short (Smith et al., 2013). Future work might show what methods from urban hydrology (Cristiano et al., 2017) could be transposed to the analysis of spatial rainfall variability in small alpine catchments."

Reviewer 1: Major comment:

The authors provided in the last review round no uniform reply to the reviewers comments. Instead, the replies were scattered among the official reply and author comments in the public discussion of the manuscript.

We decided to refer to the public discussion to emphasize which parts of the revision were exactly as discussed in the Online Discussion. We did not consider that this might be impractical for an offline review and would like to apologize for this.

Indeed, stimulating the public discussion is important and I encourage the authors to do so in the future as well. However, for the non-public review the authors should provide a point-by-point reply to the reviewers in one document to avoid that the reviewers waste time 'searching' for the correct replies. I contacted the handling editor after the first submission of the scattered reply asking for a 'complete' reply from the authors. However, there are still replies linking to other documents.

The handling editor contacted us indeed, but we misunderstood her comment and did not understand that the links to the public discussion were the problem. Instead we tried to solve some discrepancies that could exist between our author-reply in the Public Discussion and the author-reply document we uploaded with the revised manuscript. As mentioned above, in our answers we referred to the public discussion to underline that the implemented solution was exactly the one discussed previously, but this was apparently not a good choice.

I did not take them into account the replies due to a very practical reason: I'm in a train with the printed version and cannot access other documents. Hence all issues which are not addressed directly in this reply remain unsolved for me. I strongly recommend to always provide a 'complete' reply, also for upcoming articles of the first author.

The two instances where we referred to the Online Discussion are i) the point concerning peak flow (point 1c for reviewer #1, point 8a for reviewer #2) and ii) the comment on modelling (point 2 for reviewer #1, point 8b for reviewer #2). The answers to these points are reported in Appendix 1 and 2 of this document, respectively.

Furthermore, we realize just now that some citations from the track change manuscript in our answer include several broken links (to sections) appearing when converting our document into PDF (i.e. *"3.3.1Erreur ! Source du renvoi introuvable."*). We apologize for having missed this during the revision submission. Those are just references to sections, i.e. do not contain further information.

Reviewer 1: Detailed comments:

1 - L333 The authors state the r-r model is calibrated using "the mean of the 20 stochastic rainfall realizations" (see also L535). If the "mean" is used, all rainfall peak intensities between the stations are smoothed out. The authors should comment why they think this smoothed rainfall time series is an appropriate input time series. Why were not simply all 20 realisations been taken?

Thanks for this important comment; it is important to clarify that we used the temporal mean (no specified in the paper) and furthermore, the 20 stochastic realizations are all conditioned on the precipitation observed at the stations and they accordingly do not smooth out the observed peaks as illustrated in the Figure 2 below.

We updated the manuscript as follows (L324-330):

Old: “For calibration, the model is run using the mean of the 20 stochastic rainfall realizations as reference input; it is then calibrated against observed runoff (i.e. discharge - baseflow) through likelihood maximization assuming that the model residuals are normally distributed (e.g. Schaefli et al., 2007).”

New: “The model is calibrated against observed runoff (i.e. discharge - baseflow) through likelihood maximization assuming that the model residuals are normally distributed (e.g. Schaefli et al., 2007). The reference input field for model calibration is the mean of the 20 stochastic rainfall realizations at each time step (note since all realizations are conditioned on the observed precipitation events, this mean preserves the individual observed peaks of precipitation).”

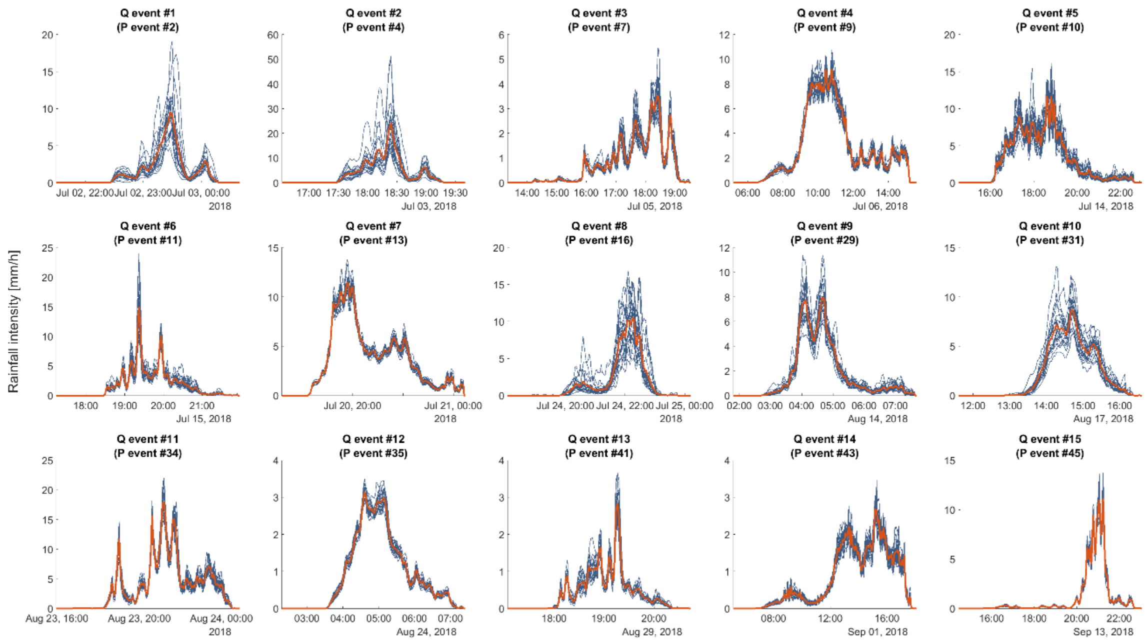

Figure 2. Areal rainfall for 15 rainfall events having a streamflow reaction. For each event is represented the 20 areal rainfall realizations (blue) and the mean value of the 20 realizations (orange).

2 - Fig. 13. Why not using a simple barplot here? The information would be much easier to catch. Except the stochastic rainfall model I do not see any clear network setup choice here. It is not as clear as authors state in l544-555.

Thanks for pointing this out, we decided to use a simple plot to present this result. The Figure 13 of the manuscript has been updated by the Figure 3 (and figure caption) below.

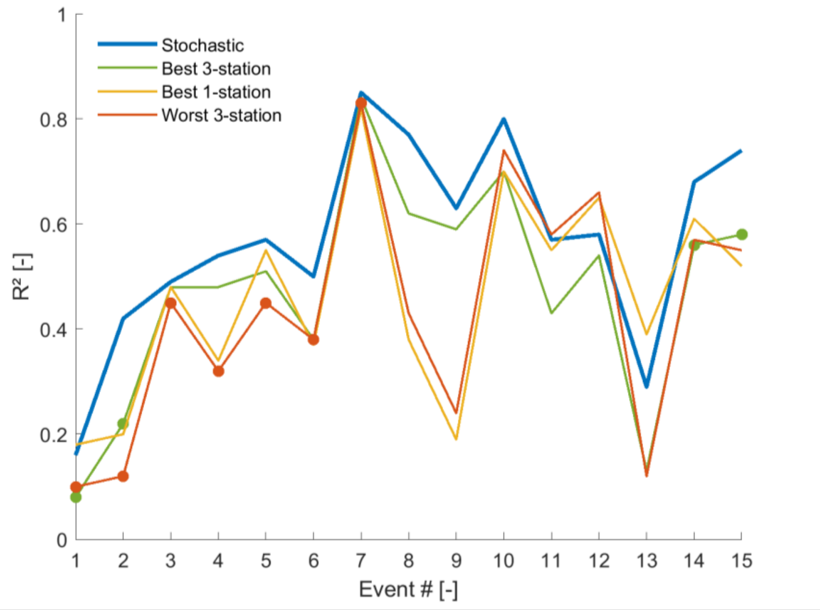

Figure 3. Analysis of 15 rainfall-runoff model events (subset #3, Table 2) with the correlation coefficient between simulated and observed streamflow for different rainfall fields inputs: the stochastic generation of rainfall fields based on all available rain gauge stations, the best 3-stations and the best 1-station network, and the worst 3-stations network. Larger dots highlight events where events where only 2 of 3 stations were operational (see Section 4.1.1). The lines connect the events to improve readability.

3 - Title: „Even event-scale hydrological response benefits from high density rain gauge observations" – Well, the rephrased title can be questioned, since the high network density is of course espially for events important. Why not „Identification of required rain gauge density for hydrological response analysis in small mountainous catchment " (or similar)?

We added "event-scale" because one of the reviewers insisted on emphasizing the event-scale character of the analysis; we suggest to change to: "Benefits from high density rain gauge observations for hydrological response analysis in a small alpine catchment".

4 - For me Fig. A and B are much better to interpret and it is easier to „catch" the relevant information in comparison to Fig. 12 and 13 (although it would be useful to have the full network in both figures on the same axis, not once on y (Fig. A) and once on x (Fig. B). Especially when lines cross each other it takes minutes to intepret which line shows the better fit. I also have in mind it it not recommended to use these polar plots for more than three datasets, because they are hard to read then (rule-of-thumb, of course). I'm wondering why ‚worst-3-stations are not implemented in Fig. A and B?

Thank you for this comment. We choose to replace the polar plots (Figure 12 in the manuscript) by the normal plots using ratio between partial network values and full network values (so we just have 3 sets of points). The Figure 12 of the manuscript has been updated by the Figure 4 (and figure caption) below.

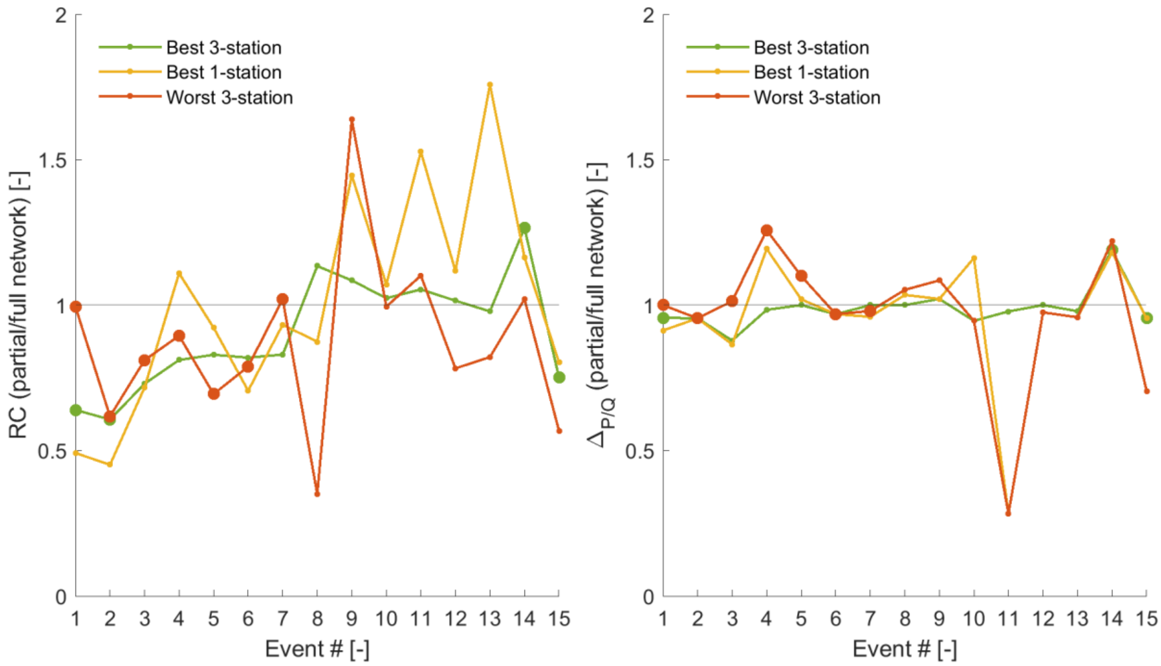

Figure 4. Comparison of streamflow response metrics ratios between a partial network (best 3-station, best 1-station and worst 3-station networks) and the full rain gauge network, using the RC (left) and lag time ΔP/Q (right). The dataset is subset#4 of Table 2. Larger dots highlight events where events where only 2 of 3 stations were operational (see Section 4.1.1). The lines connect the events to improve readability.

5 - The „lower dispersion" is hard to see/interpret. From Fig. A and B it seems both station sets lead to similar good results. For Fig. A and B, should there not always be one point for 1-station and 3-station network realted to one value from the x-axis? So the x-axis value of ~250 min has only one point fo the 3-station network, but none for the 1-station network. IF there are points missing, it is hard to judge on the dispersion.

Thank you for pointing out this mistake. Indeed, the axis limits were set too narrow and one point was out of boundaries. Please see below the same figure with the extended axis limits. The figure is reproduced with the same configuration (partial network on Y-axis, full network on X-axis) for comparison, but please consider the new Figure 12 above: using the ratio between the partial and full networks gives a better visualization of the station sets "performances".

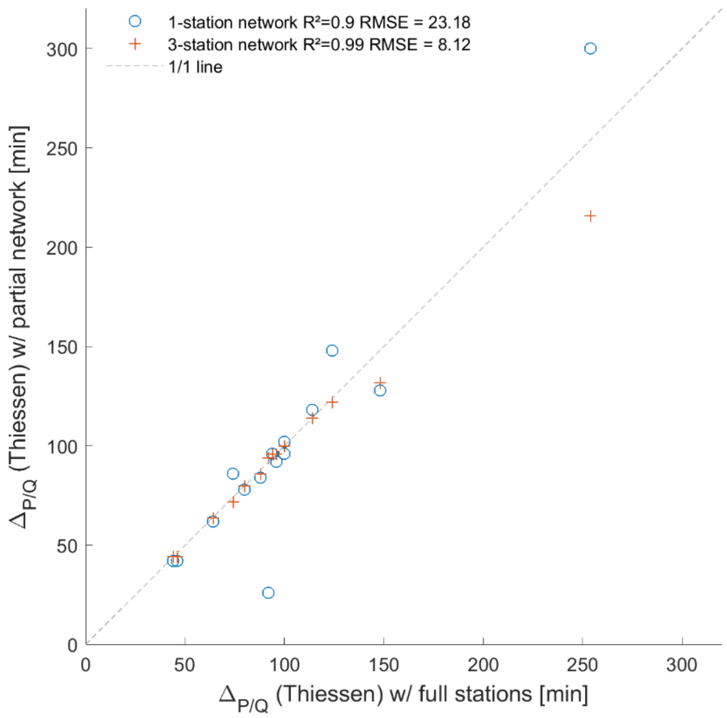

Figure 5. Difference of lag time ΔP/Q obtained from a partial network (1-station and 3-station network) and the full network.

6 - Former comment 1c - In general, I'm missing the runoff peak as important characteristic in the manuscript. Maybe the authors can involve it/comment on it why it was not considered.
-> Issue remains unsolved.

Please see our answer in the "Appendix 1: answer to point 1c for reviewer #1, 8a for reviewer #2" at the end of this document.

7 - I'm struggling where to find the new supplementary. It was not uploaded with the revised version, I'm afraid I cannot review that part.

We are sorry to read that. The supplementary material has been uploaded with our answer on November 20th (validated on November 28th). On our side the supplementary material is accessible from this link:

https://editor.copernicus.org/index.php?_mdl=msover_md&_jrl=13&_lcm=oc73lcm74a&_acm=get_supplement_file&_ms=87052&id=1619782&salt=17807967821329759507

Reviewer 1: Technical corrections/Minor comments:

- General: Please check all brackets with numerous references, spaces are missing everywhere)
  => Thank you for pointing out. It has been corrected throughout the whole manuscript.
- L24 Please add at the end of the sentence: "…for the studied catchment (0.22 rain gauges/km²)."
  => Ok, it has been modified accordingly.
- L207 "precipitated"
  => Ok, it has been changed.
- L226 "line" -> "straight line"
  => Ok, it has been changed.

- L320 “(Section 4.4 and 0)” <- What does “0” refer to?
  => This is due to a broken reference link, it has been corrected.
- Eq. 7 Please correct the formula: “0” instead of =:=”, root over whole term, providing proper limits for the sum operator. ??
  => The “:=” (defined as) sign is used in purpose as RMSE is an abbreviation. The root square is well defined over the whole term, but it looks like the PDF format conversion has shifted the top bar. The sum operator limits are now properly defined from t = 1 to n.
- L329 year for reference is missing
  => Thanks, the missing year (2018a) has been added.
- L343-344 repetition (see L226)
  => Referring to the last track change version available, we did not found any repetition between the L226 and 343-344.
- L549 space missing.
  => Thanks, it has been corrected.
- L552 What kind of reference is Beria 2020b? Is it a technical report, a doctoral thesis or an institute publication?
  => It is a PhD thesis, the reference has been updated.

References

Smith, B. K., Smith, J. A., Baeck, M. L., Villarini, G., and Wright, D. B., Spectrum of storm event hydrologic response in urban watersheds, *Water Resour. Res.*, 49, 2649– 2663, doi:10.1002/wrcr.20223, 2013.

Cristiano, E., ten Veldhuis, M. C., and van De Giesen, N.: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas - a review, Hydrol Earth Syst Sc, 21, 3859-3878, 10.5194/hess-21-3859-2017, 2017.

Appendix 1: answer to point 1c for reviewer #1, 8a for reviewer #2

The Figure 6 shows the hydrograph of the 15 rainfall events generating a river reaction. The runoff peak identification is straightforward for 5 of them (Q event #1, #2, #6, #14 and #15), but for 8 of them (Q event #3, #4, #5, #7, #9, #10, #12 and #13) the flatty shape makes the exercise very uncertain. As well for the Q events #8 and #11 showing a double peak, the shape of the hydrograph itself is then explained more by the fluctuations of the rainfall amounts than by the dynamic of the hydrological processes. This statement led us to use event-scale metrics for the hydrological response. We added the Figure 6 to the Supplementary Material (Figure S5).

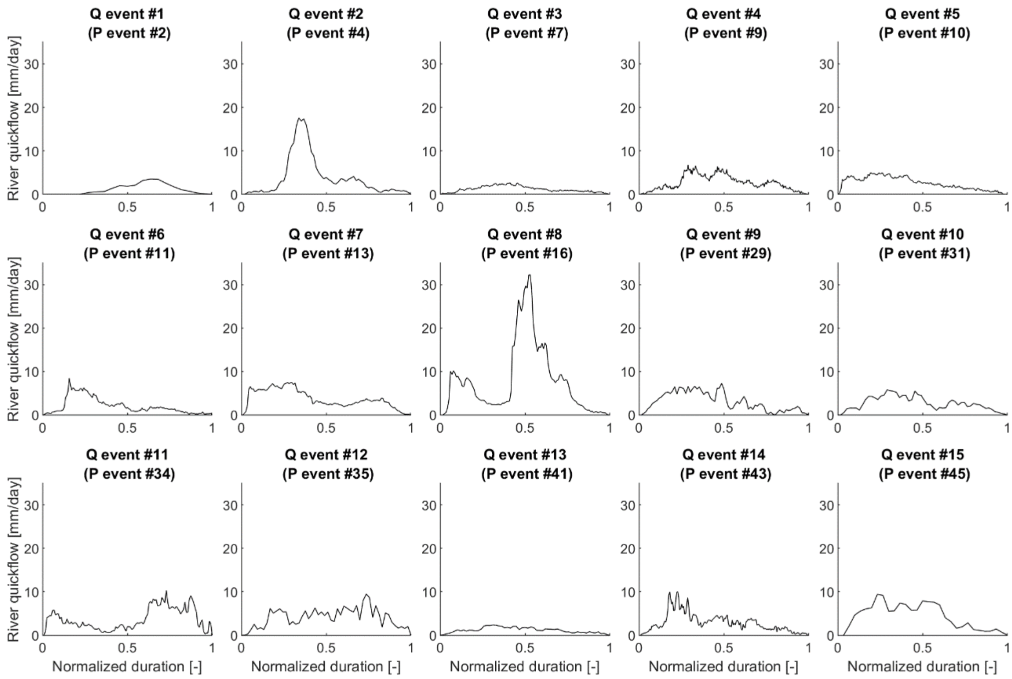

Figure 6. River quickflow for 15 rainfall events causing a noticeable river reaction. The length of events is normalized.

Appendix 2: Answer to point 2 for reviewer #1, 8b for reviewer #2

We decided to add a modelling component to this paper; the model is discussed in the public discussion:

"Following this suggestion and the suggestion of the referee #2, we implemented an event-based modelling approach (note: the following answer is identical to the answer 8b given to the referee #2). The runoff response of Vallon de Nant to rainfall forcing is modeled by a semi-distributed model. This model first simulates the mobilization of water at the sub-catchment scale (here 25 sub-catchments are defined over the Vallon de Nant) using a SCS runoff curve number approach. Next, stream discharge is obtained by convoluting the resulting hillslope responses with a travel path distribution derived from the stream network geometry (Schaefli et al., 2013). In the current version (to be refined) the subcatchments and the stream network geometry are identified using TopoToolbox (https://topotoolbox.wrrdpress.com) (Figure 7)1, in which travel paths correspond to the distance between the bottom part of each sub-catchment and the catchment outlet. In this model we focus on the fast response (i.e. runoff) of the catchment, and baseflow (defined here as the average discharge during the 30 min preceding event start) is subtracted from the actual discharge prior to runoff modeling. For calibration, the model is run using the mean of the 20 stochastic rainfall realizations as reference input; it is then calibrated against observed runoff (i.e. discharge - baseflow) through likelihood maximization assuming that the model residuals are normally distributed (e.g. Schaefli et al., 2007). After calibration the event-based runoff model is applied to the different network configurations to test how rain gauge network geometry influences the simulated runoff response. As the stochastic rainfall interpolation cannot be performed with a number of observation points as low as 3 stations (or less), we use the Thiessen polygons method to interpolate the rainfall fields from the 1 to 3-station raingauge network. The results of this model are all shown in the Appendix at the end of this document.

What we can say at this stage is that this kind of typical conceptual event-based hydrological model cannot reproduce all observed events equally well (Appendix 1). This would require in-depth analysis of different subsurface flow mechanisms related also to snow melt and shallow-groundwater recharge, work that is ongoing in this catchment. What is clear is that the simulations with the worst 1 station network are completely off. In exchange, the simulations with the best 1-station, 2-station or 3-stations network is always close to the simulations obtained with the stochastic rainfall fields, which underlines the value of the station network selection methodology in the submitted paper. The analysis furthermore shows that an ill-placed weather station can result in completely erroneous runoff simulation, whereas a network of at least 3 stations results in much better runoff simulations. This conclusion would not have been possible without the high density network observations. However, this model experiment cannot shed further light on the value of the high density networks as the ability of the model to reproduce streamflow responses is not good enough for clear conclusions. This cannot be easily solved with another conceptual model (we tried already other conceptual model structures, e.g. Benoit, 2020) nor with any "out-of-the-shelf" model, which do not exist for high alpine headwater catchments. The development of a fully distributed high resolution (e.g. 10 m x 10m) physical model with the inference of distributed model parameter fields is beyond the reach of this study.
In any case, we can try to include some key results from the modelling study in the revised version."

Corresponding modifications of the paper are (please note lines numbers refer to the previous track change version of the manuscript)

i) at the end of the introduction (L92-94):
"We conclude the analysis with an event-scale modelling of all recorded runoff response events with a semi-distributed model to evaluate the identified rain gauge network configuration."

ii) presenting the model used in the method part “3.6 Rainfall-runoff model” (L335-350)

“To further validate the obtained optimal rain gauge network configuration, we set up a a semi-distributed, event-based rainfall-runoff model. This model first simulates the mobilization of water at the sub-catchment scale (25 sub-catchments) using a Soil Conservation Service Curve Number (SCS-CN) approach (SCS, 1972). Next, the streamflow response is obtained by convolving the resulting hillslope responses with a travel path distribution derived from the stream network geometry (Schaefli et al., 2014). The subcatchments and the stream network geometry are identified using *TopoToolbox* (https://topotoolbox.wordpress.com), in which travel paths correspond to the distance between the bottom part of each sub-catchment and the catchment outlet. In this model we focus on the fast response (i.e. runoff) of the catchment, and baseflow (defined here as the average discharge during the 30 min preceding event start) is subtracted from the actual discharge prior to runoff modeling. For calibration, the model is run using the mean of the 20 stochastic rainfall realizations as reference input; it is then calibrated against observed runoff (i.e. discharge - baseflow) through likelihood maximization assuming that the model residuals are normally distributed (e.g. Schaefli et al., 2007). After calibration the event-based runoff model is applied to the different network configurations to test how rain gauge network geometry influences the simulated runoff response. As the stochastic rainfall interpolation cannot be performed with a number of observation points as low as 3 stations (or less), we use the Thiessen polygons method to interpolate the rainfall fields from the 1 to 3-station rain gauge network obtained during optimal network analysis.”

iii) in the results section in “4.4.3 Optimum network evaluation” (L544-555)

“In a second evaluation step of the identified optimum rain gauge network, we simulated the event-based streamflow response for the best 1-station network and the best and the worst 3-station network, to compare the result to the simulation with the original rainfall field and thereby obtain a validation on the entire streamflow dynamics rather than on RC or lag only (all simulations are available in Supplementary Material part 1). It is important to point out here that the semi-distributed hydrological model cannot reproduce all observed events equally well as shown by low correlation coefficients between observed and simulated streamflow in Figure 13. Even with the stochastic generation of rainfall fields, fast streamflow tends to be underestimated with the model; improving the simulation quality for all events would require an in-depth analysis of different subsurface flow mechanisms related also to snow melt and shallow-groundwater recharge, work that is ongoing in this catchment (Beria, 2020b).

Despite of this, we clearly see that the best 1-station network and the worst 3-station network considerably underperform with respect to the full network and that the best 3-station network yields a simulation performance close to the original rainfall field, confirming the results obtained for the summary streamflow response metrics RC and lag.”

iv) with the Figure 13 summarizing the results of the different simulations:

Please see above our answer to detail comments – point 2.

v) in the Supplementary Material part 1 with the Figure S9 (map of subcatchments, Figure 7 in this document), Figure S10 (the results of all simulations per event, Figure 8 in this document) and Figure S11 (the results of simulations per event, cumulated over time, Figure 9 in this document).

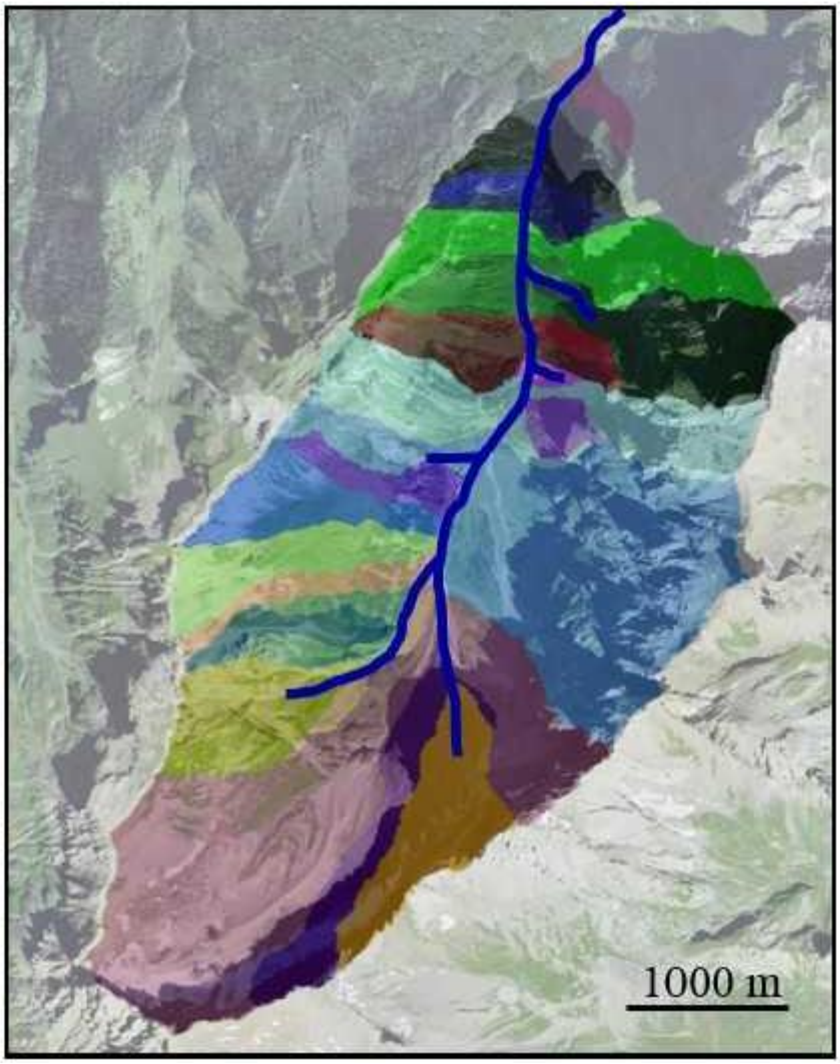

Figure 7. Map of the Vallon de Nant showing the 25 subcatchments and the stream network geometry used for the modelization.

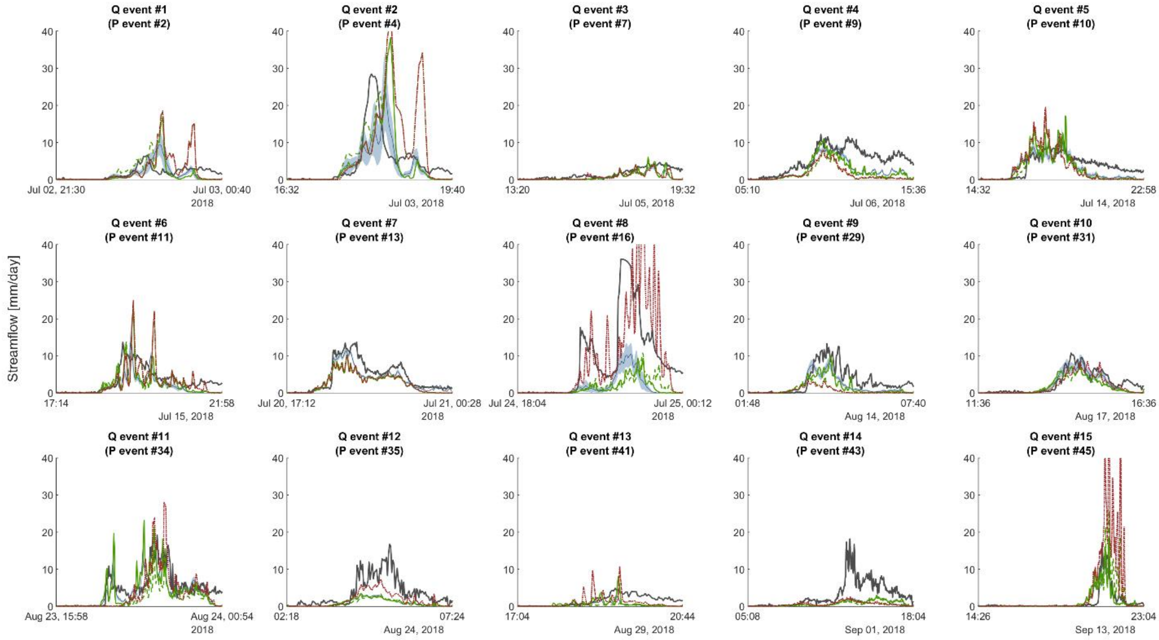

Figure 8. Results of simulation. For each of the 15 events having a streamflow reaction is plotted the observed streamflow (black curve), the simulated streamflow based on stochastic rainfall fields (blue curve and band), the simulated streamflow based on the best 3-station and 1-station network (respectively with plain and dashed green curves), and the simulated streamflow based on the worst 3-station network (dotted red curve).

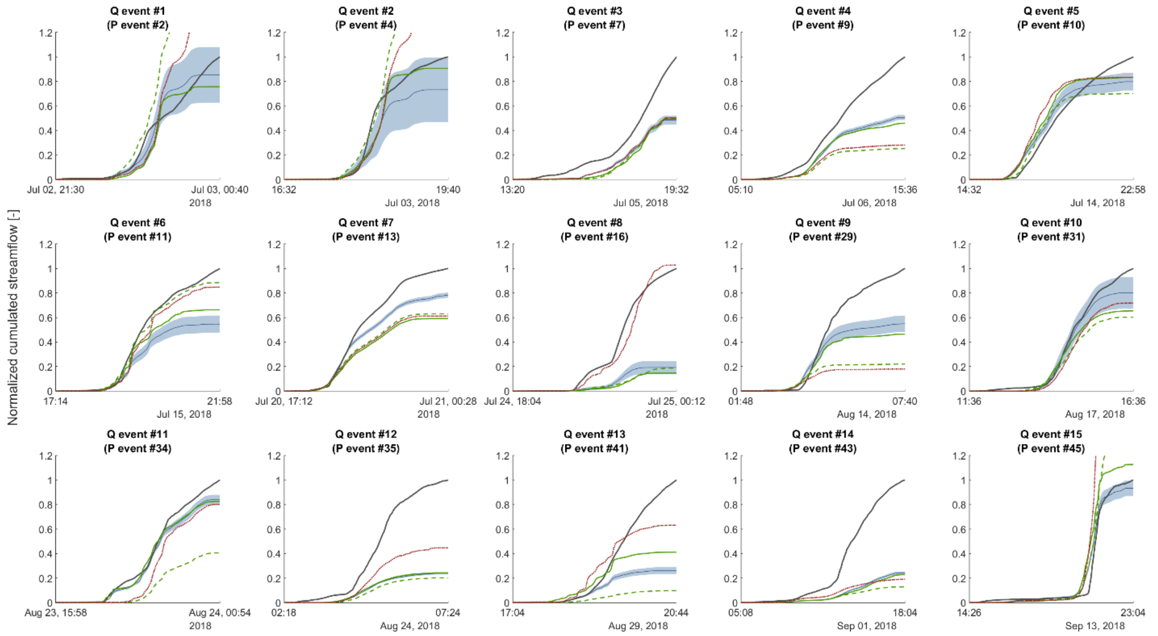

Figure 9. Cumulated results of simulation. For each of the 15 events having a streamflow reaction is plotted the cumulated observed streamflow (black curve), the cumulated simulated streamflow based on stochastic rainfall fields (blue curve and band), the cumulated simulated streamflow based on the best 3-station and 1-station network (respectively with plain and dashed green curves), and the cumulated simulated streamflow based on the worst 3-station network (dotted red curve). The amounts of streamflow are normalized by the cumulated amount of observed streamflow.