Answer to Referee #1

This was the first time I was involved as a reviewer for this manuscript. The aim is to identify the best rain gauge network setup for runoff predictions. The topic is suitable for the journal and of interest for the community; the manuscript is well-written. How-ever, I do not recommend a publication at its current stage. There are a few major comments listed below, which have to be addressed before the manuscript can be recommended for publication. More specific comments and some technical corrections follow afterwards. My overall recommendation of the manuscript is major revision.

Thank you very much for agreeing to review our paper and for the time spent to formulate all the constructive comments and corrections below.

1a - The title states "value of high density rain gauge observations for… hydrology". I'm struggling with this holistic formulation.

We propose to change the title from "*On the value of high density rain gauge observations for small Alpine headwater catchment hydrology*" to "*Even event-scale hydrological response characterization benefits from high density rain gauge observations*".

1b - Indeed, the value is "only" (please don't get me wrong here) based on prediction of RC and deltaP/Q. While a realistic estimate of these characteristics is valuable, the uncertainties resulting from the final network with 3 rain gauges for these two criteria is not shown and should be added in a later version of the manuscript.

It is true that the uncertainties resulting from the final network of 3-station raingauges is not shown. We propose to add in "4.4 Measurement network analysis" two figures showing i) the RC (Figure 1) and ii) the lag time $\Delta_{P/Q}$ (Figure 2) by comparing the values obtained from the best 1-station or 3-station raingauge network vs. the reference value calculated from the full raingauge network.

As the stochastic method for generating rainfall fields cannot be used with a number of points as low as 1 or 3 stations, we performed the computations using the Thiessen polygons methods and consequently no error bars are associated to these plots. Nevertheless, the Figure 3 compares the two methods (stochastic vs Thiessen polygons) when the RC and the lag time $\Delta_{P/Q}$ are computed from the full raingauge network.

We observe for both the RC (Figure 1) and $\Delta_{P/Q}$ (Figure 2) a lower dispersion of values while increasing the density of the raingauge network. With a 3-station raingauge network the error on the RC (RMSE = 0.186) drops below the error obtained by comparing 2 different interpolation methods (RMSE = 0.256), giving a good confidence to the Thiessen polygons method used for this calculation. In the same way, for $\Delta_{P/Q}$ the error with a 3-station network (RMSE = 8.12) is lower than the error obtain with the model comparison (RMSE = 13.22).

On Figure 2, the dispersion of $\Delta_{P/Q}$ is originally low. Even with a 1-station network the lag can be reproduced correctly for most of the events but can also be completely wrong for one of them. Outliers are still observed with the 3-station raingauge network though, even if the error gets lower (RMSE reduced from 23.18 to 8.12).

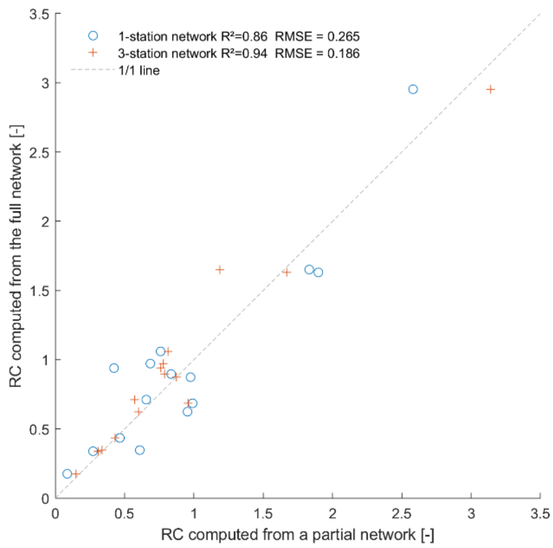


Figure 1. Comparison of RC whether it is calculated from the full raingauge network or from a partial, considering the best 1-station and 3-station network. The dataset is based on the 15 rain events associated to a river reaction.

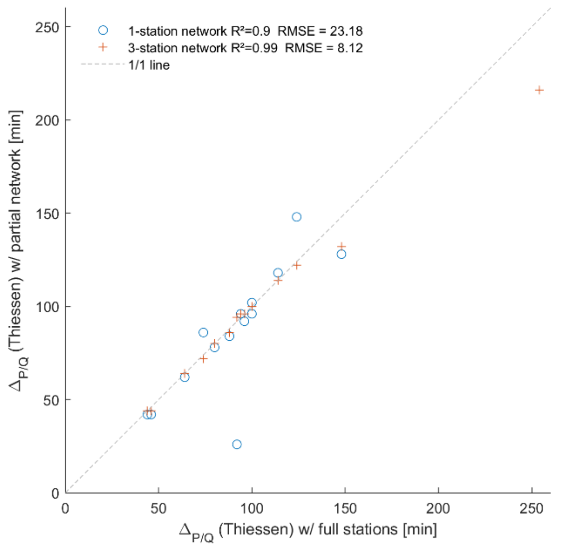
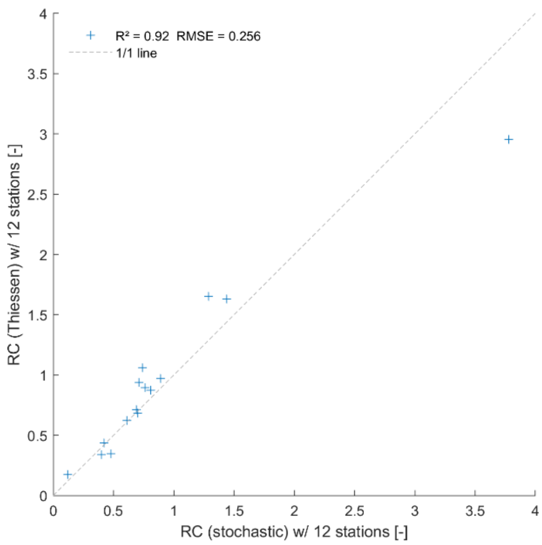


Figure 2. Difference of lag time $\Delta_{P/Q}$ obtained from a partial network (1-station and 3-station network) and the full network.



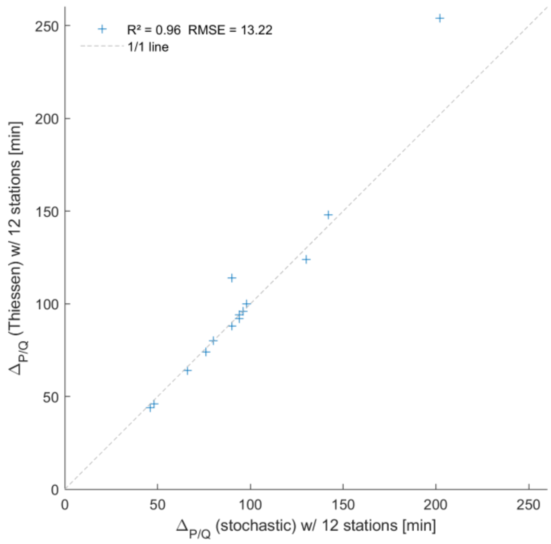


Figure 3. Comparison of the RC (left) and lag time $\Delta_{P/Q}$ (right) calculated using the full raingauge network, but with a different rainfall field interpolation method (Thiessen polygons vs. stochastic).

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.

The Figure 4 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 will add the Figure 4 to the supplementary material to illustrate our comment.

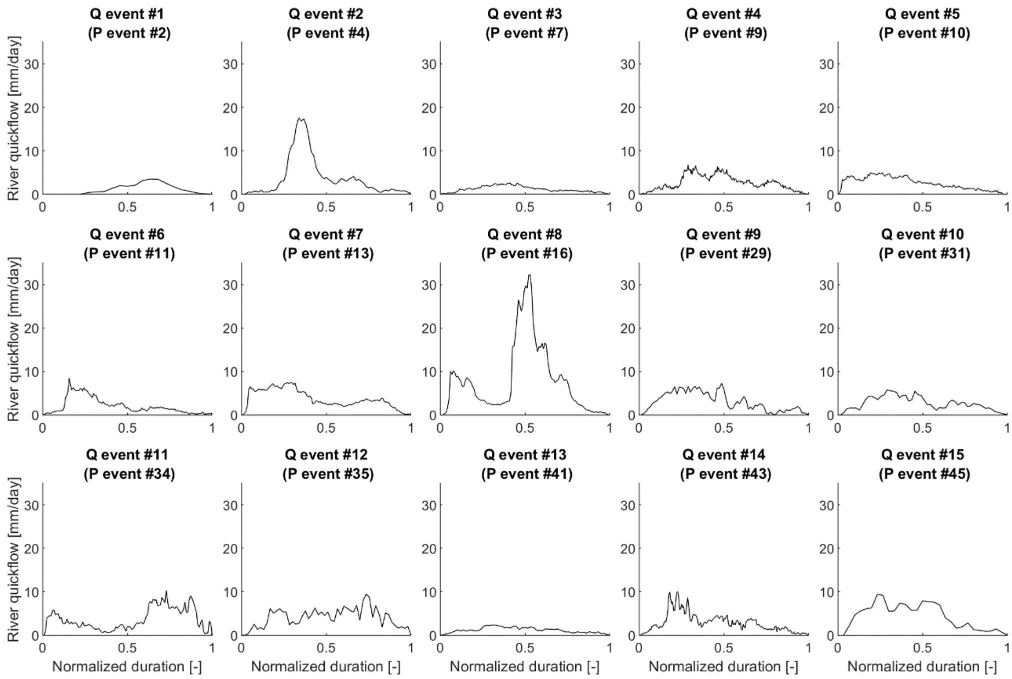


Figure 4. River quickflow for 15 rainfall events causing a noticeable river reaction. The length of events is normalized.

1d - Also, although the analysis is designed mainly for discharge estimation, results should be also interpreted in terms of rainfall (e.g. resulting areal rainfall (extremes) for different rain gauge network densities, spatial rainfall characteristics…).

We agree that these metrics are finally not fully exploited. We propose to add a short analysis on the impact of the raingauge density over i) the number of misestimated events and ii) the maximum rainfall intensities. The figures and text below will follow the existing paragraph of section 4.1.1 after changing its title from "Amounts and asymmetry" to the more general formulation of "Rainfall characteristics" (please read our answer to the point 3 of this review for further details concerning subsets). The references here are pointing to the table and figures of this document:

Relying on the rainfall events subset #2 composed of 23 rainfall events recorded by the full raingauge network (see Table 1), we tested what a partial raingauge network (all possible combinations of networks composed with less than 12 stations) would record, compared to the full raingauge network of 12 stations taken as a reference. The Figure 5 shows, in term of raingauge density, the number of events having the total amount of rainfall P_{TOTAL} overestimated or underestimated by a factor 2. We globally observe a misestimation inversely proportional to the raingauge density, with up to 3 events overestimated and 8 events underestimated with the lowest raingauge density of 0.07 raingauge per km² (1 raingauge). It is necessary to reach 0.82 raingauges per km² (11 stations) to no longer have events misestimated by a factor 2. We also observe, with few raingauges, a strongest trend to underestimate than overestimate events. The invoked reason is that facing a heterogeneous event for which a good spatial resolution of the rainfall field is needed, it is statistically more probable to miss the localized important part of the rainfall field than capture it.

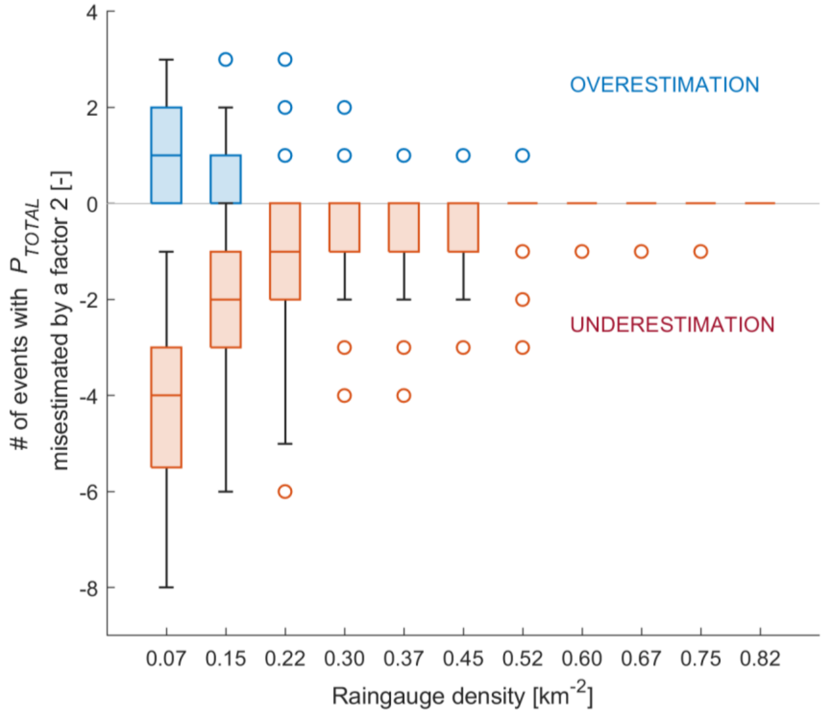


Figure 5. Number of rainfall events for which the total amount of rainfall is overestimated or underestimated by a factor 2, according to the raingauge density, going from 0.07 raingauges per km² (1 raingauge within the catchment) to 0.82 raingauges per km² (11 raingauges). For each raingauge density, all possible combination of raingauge network is tested. The reference value is estimated from the full 12-raingauge network. The bottom and top of each boxes are respectively the 25th and 75th percentiles of the sample. The line in the middle of each box is the sample median. The whiskers go up to 1.5 times the interquartile range away from the bottom or top of the boxes and values beyond are marked as outliers with circles.

The Figure 6 presents in the same way the maximum error encountered on the maximum rainfall intensity over 10 minutes P_{MAX}*(10 min).* We logically notice an inversely proportional trend, minimizing the error while the raingauge density increases. The figure also shows that in general a low raingauge density tend to overestimate more than underestimate the P_{MAX}*(10 min).* This bias originates the large footprint associated to a low raingauge density, increasing the disparities between the measuring points while interpolating the rainfall fields.

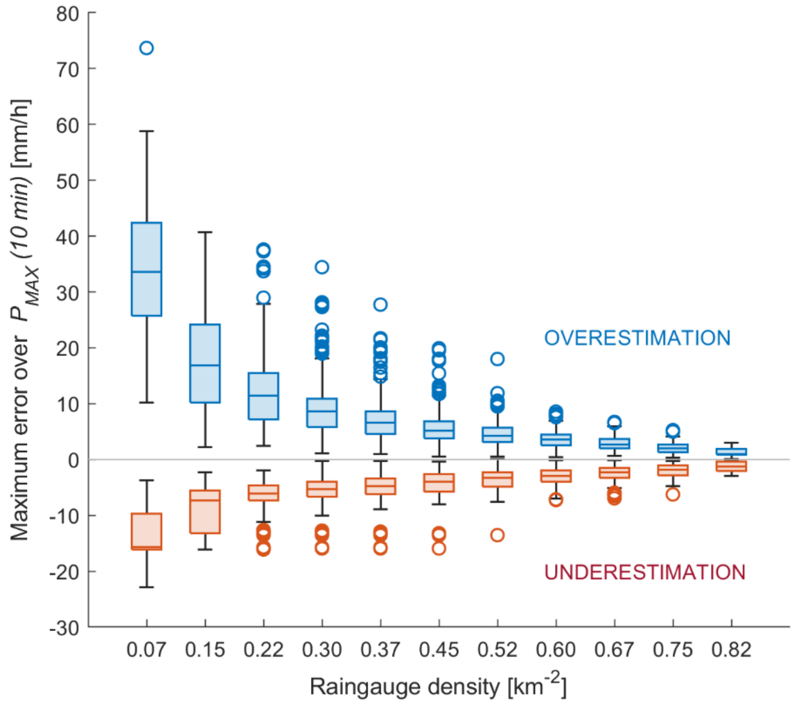


Figure 6. Error on the maximum rainfall over 10 minutes according to the raingauge density. For each raingauge density, all possible combinations of raingauge network is tested. The reference value is estimated from the full 12-raingauge network.

2 - Based on the comment before, the impact of the rain gauge network densities (and rain gauge locations) on the runoff is not analysed. In the additionally uploaded comment the main author states a rainfall-runoff modelling would go beyond the scope of the study. I do not agree with that and recommend this modelling approach to analyse the impact on the resulting runoff itself instead on single runoff statistics. To attribute the spatial rainfall variability, a distributed rainfall-runoff model would be the best solution.

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.

[1] An automatic identification of subcatchments corresponding to a manually identified stream network (i.e. identified in the field) is non trivial; solution to be found.

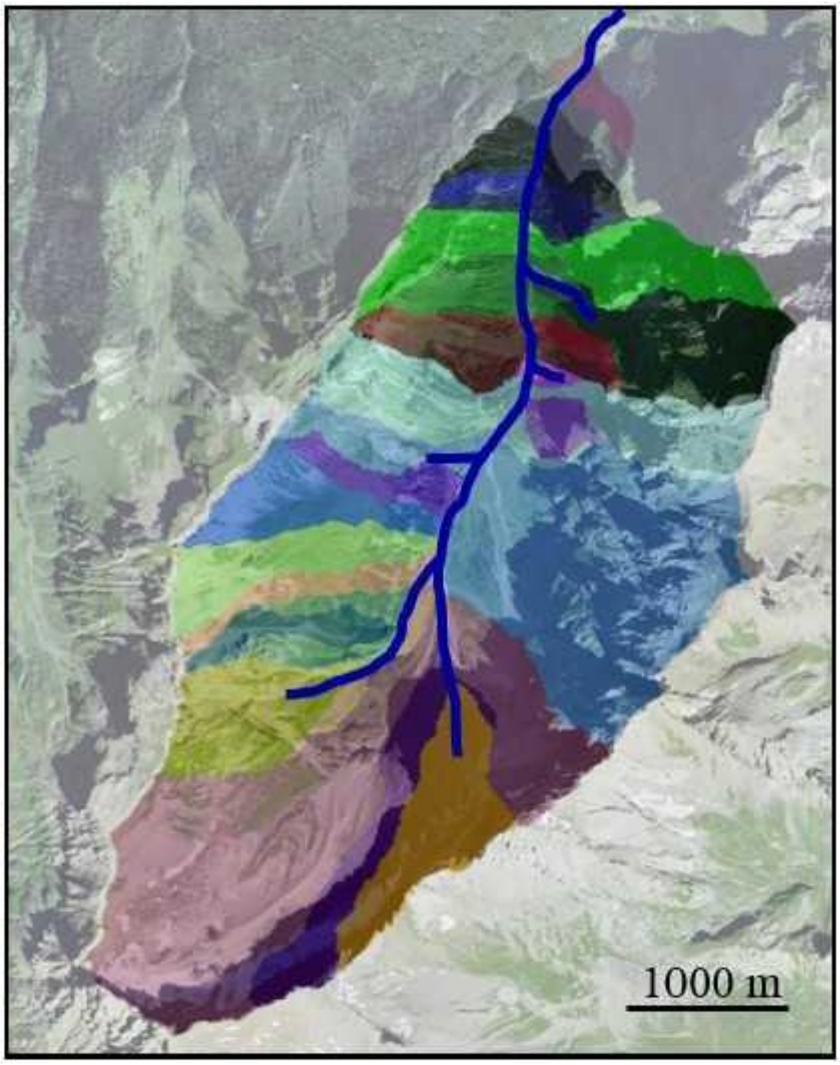


Figure 7. Map of the Vallon the Nant showing the 25 subcatchments and the stream network geometry used for the modelization.

3 - Also, I was wondering why is there not a consistent number of events analysed throughout the manuscript. I understand that there are always measuring issues and maybe some observations are questionable, but then please remove them at the beginning. There could be one number of rainfall events considered and one subset of them for discharge analysis, but at the current state results from different subsections cannot be compared with each other due to the different populations of considered events.

Indeed, we have different subsets of events. Among i) the initial 48 rainfall events that were recorded by at least 7 raingauges, which is enough for computing the spatial rainfall pattern metrics (Section 3.2.2), we choose ii) the subset of 23 rainfall events recorded by the full network of 12 raingauges to evaluate the best partial network (Section 3.5); the iii) 15 rainfall events associated with a streamflow reaction was reduced to iv) 14 events after discarding the July 24th event identified as an outlier. This last subset is used for computations involving the hydrological response. Although we are often referring to the July 24th outlier event and recall in the figure captions that this particular event is out of the axis limits, it is not included in our computations. We will make this clearer in the article by adding the above sentences and the synthetic Table 1 that summarizes and numbers the different subsets used within this study. However, we stick to the idea of keeping datasets as large as possible in order to increase the statistical robustness of our results.

Table 1. Summary of the different subsets of rainfall events (P_{event}) used within this study.

Subset	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
#1: P_{event} recorded by at least 7 stations	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
#2: P_{event} recorded by 12 over 12 stations																x	x	x		x	x		x	x
#3: P_{event} w/ river reaction		x		x			x		x	x	x		x			x	no discharge measure							
#4: P_{event} w/ river reaction (w/o outlier)		x		x			x		x	x	x		x											

Subset	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	# of events
#1: P_{event} recorded by at least 7 stations	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	48
#2: P_{event} recorded by 12 over 12 stations	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x								23
#3: P_{event} w/ river reaction					x		x			x	x						x		x		x				15
#4: P_{event} w/ river reaction (w/o outlier)					x		x			x	x						x		x		x				14

4 - L25-27 It should be mentioned here again that this issue is related to mountainous areas and is not a problem in general.

Indeed, it is useful to precise it in this sentence too in order to avoid confusion. We will mention it.

5 - Fig. 2 I don't see the additional worth of showing Fig. 2 and recommend to leave it out, especially since it is included in the supplement as Fig S2 as well.

We agree that the weir picture of the Figure 2 of the paper rather has an illustrative role and is not essential in a hydrology paper. As the Figure S2 in the supplementary material also fulfills this aim, the Figure 2 will be removed from the main part of the article.

6 - L90 "average elevation" Please change to mean or median, depending on how you determined the "average" value.

We agree to use "mean" rather than "average", it will be corrected in the revised version of the paper.

7 - L117-118 The construction of the rating curve is not interesting for the manuscript and can be left out, also the elements regarding its construction in the supplement.

We agree. The details concerning the rating curve construction and error estimate will be move to an appendix.

8a - L154-155 The term interpolation is not suitable in my opinion due to the rainfall generation mechanisms behind. I suggest "areal rainfall is generated after Benoit et al. (2018a) by constraining actual observations at rain gauge locations". The authors should give a less brief explanation, since in the cited manuscript different versions are applied for rainfall generation (three versions due to different covariance models) and it remains unclear for the reader, which model is used for the current study.

Thanks, we will add some more details in an appendix.

8b - Why did the authors choose this rainfall generation instead of a regionalization approach as kriging (maybe with altitude as additional information), inverse distance weighting or Thiessen polygons. The latter is chosen later in the manuscript nevertheless due to computational efforts, so why not for the whole study? Was it the authors intention to add an uncertainty analysis.

Indeed, we choose to use the stochastic approach for the valuable estimation of the errors it provides (in response to the first round of reviews during the first submission; the original manuscript used Thiessen only). The Thiessen method also used throughout this paper fills the weak points of the first method, namely i) the computation time, which is very short using the Thiessen method and allows to explore within a reasonable amount of time all the possible combinations of raingauge networks for their optimization, and ii) to calculate rainfall fields with a low number of raingauges. For this last point the stochastic method require at least 5 stations to capture correctly the spatial and temporal rainfall characteristics.

Concerning the altitude effect on rainfall, we do not observe any trend ($R^2 = 0.06$) between the cumulated rainfall per station and the altitude. This information will be added in "4.1.1 Rainfall characteristics" (please see 1d concerning the title change of this section). The Figure 8 below is shown as illustration purpose for this document.

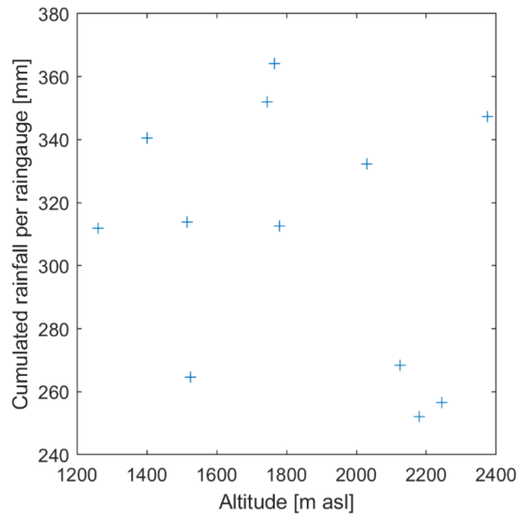


Figure 8. Cumulated rainfall at each raingauge vs. altitude of each raingauge.

9 - L154-163 The authors should bring this argument in context with the catchment concentration time.

This statement probably refers to "Using the interpolated rainfall fields, rainfall events were identified as rainy periods separated by at least 90 minutes without rain. This inter-event duration was selected based on the observed delay between rainfall onset and streamflow response for the large event recorded on August 23rd (detailed in the part 2 of supplementary material); the streamflow reaction to the first half-hour of this rainfall event was caused only by rainfall in the southern half of the catchment (stations 8 to 12)."

Following e.g. Dingman (2002), "the time of concentration Tc [is] defined as the time it takes for water to travel from the hydraulically most distant part of the contributing area to the outlet". It is difficult to determine in practice due to unknown flow paths. That is exactly why we choose the event on August 23rd, which happened far away from the outlet as an indicator for this travel time to the outlet. We will make this explanation clearer and add a reference to the concentration time.

10 - L165-166 The location of the line chosen for the splitting of the catchment seems to be chosen arbitrary. Would a line constructed perpendicular to the main flow direction of the river (or even better, not a straight line but following the lines perpendicular to the isohypses to separate flows exactly) lead to more representable results, since the catchment is then split into a real upper and lower part? Or (thinking the other way around) does it not matter at all and the splitting line could be also drawn from South to North as long as both parts have the same area?

Indeed, the choice of a west-east line splitting the catchment into 2 parts of equal area is arbitrary. We agree that a line splitting the catchment by crossing perpendicularly the river (or the close solution of following a line perpendicular to the isohypses) are better solutions, as they are a more general answer to this problem of splitting the catchment into two areas, whatever is its orientation. We believe though that the easy geometric solution we choose for the computations would not give, in this particular case, fundamentally different results.

We tried the solution of splitting the area following a given isohypse (around 2023 m asl, to split the catchment into two parts of equal area) but we finally discard this option later while introducing D_{HILLS}, which is a redundant metrics ($R^2 = 0.76$) that gives a more accurate description of the localization of the rainfall within the catchment.

However, we did not consider splitting the catchment following a north/south (or some titled line to follow the general orientation of the catchment) line. We would not do it in purpose of testing the relevance of the orientation of the splitting line, but because we think it could actually be a good choice. Such a line would roughly separate the steep slopes on the east side from the grassy slopes on the west side (see Figure 2 of the article). For this case study this choice would interestingly isolate geomorphological distinct areas. We unfortunately did not have time to explore this solution, but a sentence will be added in "3.2.2 Spatial rainfall pattern metrics" to criticize our choice of the north/south catchment splitting and the potential solutions to be considered.

11 - L211-215 I suggest to move this paragraph to the beginning of section 3.3.2

This comment refers to the statement on how baseflow is separated ("The beginning and the end of the streamflow response determine the initial and final baseflow, respectively; the streamflow volume above the line connecting these two points is considered here as fast runoff.") and we think that it is an integral part of event identification. We will rename the subsection "3.3.1 Event identification" to "3.3.1 Event and quickflow identification".

12 - L217-218 The authors declare volume and lag time as "the two key characteristics of streamflow reaction". I do not agree with that. The most important characteristic is peak flow, followed by volume and then lag time and flatting behaviour. Even if all characteristics are considered equal important, the authors should state why peak is not considered in the study. If there were attempts to include peaks which did not work, the authors should state so as "lessons learned" in the manuscript.

In agreement with our answer to the point 1c of this document, we will give more details about the reasons that led us to discard the peak flow.

13 - L219-221 Is this criterion developed by the authors or should a reference be cited in this context? How was 1/3 chosen as threshold? This value should be catchment-dependent in my opinion, or not? Please clarify.

This comment refers to the runoff coefficient. We will add a reference to a classic textbook (Dingman, 2002).

Classical lag time definitions are e.g. the lag between the start of the effective rainfall (the one that creates a reaction) and peak flow. As discussed above, the concept of peak flow is difficult to apply to our observed events. A classically used alternative is the centroid-lag (lag between the centroids of rainfall and of streamflow), which is a useful to characterize the response time (e.g. Dingman, 2002). Given the varying shape of our hydrographs, we empirically tested different lag formulations; the lag between 1/3 of rainfall and 1/3 of streamflow gives the best results in the regression analysis. We will clarify this in the revised version.

14 - L222 Why is this criterion "1/3 of the rainfall amount" more robust than "start of the rainfall event", although both starting points are linear correlated?

Thanks for pointing this out. The formulation was not well chosen; the start of the rainfall event will not contain information about the actual start of effective rainfall (the rainfall that creates a reaction), which depends on antecedent storage conditions. We will reformulate (see also comment 13).

15 - L275 Same differences lead to higher asymmetry values for smaller values. To avoid a misinterpretation ("Interestingly…") Pnorth and Psouth could be normalized by the mean event rainfall amount. This would provide deeper insights, especially since larger differences between both parts cannot be seen in the current approach if they occur for events with high rainfall amounts.

We agree. We will modify the Table 2 of the paper (and the supplementary material) by adding two columns with P_{NORTH}/P_{ALL} and P_{SOUTH}/P_{ALL}. The Table 2 of the paper will be modified as the Table 2 below.

Table 2. Modified table of the "List of recorded precipitation events with streamflow reaction (in 2018)".

Date	$P_{DURATION}$ [min]	$Q_{DURATION}$ [min]	$\Delta_{P/Q}$ [min]	P_{ALL} [mm]	P_{NORTH} [mm]	P_{SOUTH} [mm]	P_{NORTH}/P_{ALL} [-]	P_{SOUTH}/P_{ALL} [-]	$W_{3\ days}$ [mm]	Q_{INIT} [mm]	Q_{FAST} [mm]	RC [-]	I_{ASYM} [-]	D_{HILLS} [m]	D_{STREAM} [m]	*HAND* [m]
2-Jul	42	44	24	7.7	4.1	3.6	0.53	0.47	3.2	7.9	0.9	0.12	-0.06	1521	4008	611
3-Jul	40	135	23	12.1	7.4	4.6	0.62	0.38	12.7	7.5	8.5	0.71	-0.24	1336	3842	550
5-Jul	224	309	71	8.2	4.0	4.2	0.49	0.51	29.8	6.0	6.0	0.74	0.03	755	4374	350
6-Jul	478	587	65	20.2	8.6	11.6	0.43	0.57	40.3	5.8	25.9	1.29	0.15	874	4450	355
14-Jul	358	302	49	18.7	10.5	8.2	0.56	0.44	0.0	4.5	12.9	0.69	-0.12	1263	3574	554
15-Jul	136	281	33	10.7	6.0	4.7	0.56	0.44	18.9	5.5	9.5	0.89	-0.13	1122	3377	528
20-Jul	288	228	49	18.8	8.6	10.2	0.46	0.54	3.4	4.8	14.2	0.76	0.09	1282	3823	541
24-Jul	220	229	45	8.0	7.5	0.5	0.94	0.06	12.2	3.1	30.4	3.78	0.02	740	2184	419
14-Aug	204	152	47	11.1	4	7.1	0.37	0.64	10.2	4.0	7.8	0.70	0.27	1286	4305	540
17-Aug	152	109	38	11.9	6.2	5.7	0.52	0.48	17.5	3.2	4.9	0.42	-0.04	1122	3780	490
23-Aug	388	237	47	22.1	8.8	13.3	0.40	0.60	5.4	2.4	13.5	0.61	0.20	1371	3756	563
24-Aug	158	107	40	8.1	4.4	3.7	0.54	0.46	29.5	4.1	6.5	0.81	-0.08	692	4114	320
29-Aug	72	116	48	4.8	2.2	2.6	0.46	0.54	12.4	3.0	2.3	0.48	0.07	1207	3526	524
01-sept	628	341	101	11.4	4.3	7.2	0.38	0.63	20.4	3.4	16.4	1.44	0.25	725	4487	331
13-sept	370	59	45	10.9	7.0	3.8	0.65	0.35	0.0	2.6	4.4	0.40	-0.29	1291	3594	556

16 - L323-327. I cannot follow the argumentation here. Please explain in detail how you achieve this conclusion and consider at least one or two sentences for each argument.

The corresponding original text will be complemented as follows:

"The strong correlation between rainfall amounts and Q_{FAST} (0.77, Table 3) suggests that streamflow reactions are triggered by saturation-excess, rather than by infiltration capacity-excess: [NEW] If saturation is exceeded, every unit of rainfall will lead to a corresponding unit increase of streamflow, hence a strong linear correlation to rainfall amounts. Furthermore, saturation-excess also implies that a longer rainfall event leads to more streamflow reaction (once the saturation threshold is reached, all rainfall contributes to streamflow). If, on the contrary, the driving process was the exceedance of infiltration capacity, then only rainfall intensities above the capacity threshold would trigger a corresponding streamflow increase, small rainfall amounts would trigger almost no reaction. In this case (infiltration-excess), there would be no linear correlation between rainfall amounts or rainfall duration and streamflow amounts, but a strong correlation between streamflow amounts and high or maximum precipitation intensity. [end NEW] [REFORMULATE] Saturation-excess as a main driver of the fast streamflow response is confirmed by i) the absence of correlation between maximum rainfall intensity over 10 minutes and the RC (Table 3) [end REFORMULATE], ii) the strong correlation between rainfall duration and Q_{FAST} (0.73) and iii) by the clear threshold effect for the generation of streamflow as a function of total event rainfall (Figure 9); a streamflow reaction only occurs for total rainfall higher than 5 mm".

17 - L330 "to reach a higher "RC" Please rephrase, the manuscript is about observations, not modelling.

We will rephrase "to reach a higher RC, we need a higher level of saturation [...]" by "we observe a higher RC when the level of saturation increases [...]".

18 - L341 composites: If there is a differentiation into wet and dry state, how do the authors achieve only one value for each criterion? Are two values estimated (for wet and dry) and then the arithmetic mean is mentioned? Please clarify!

Instead of analyzing all the events with an identical "dry" or "wet" network extent all along, for the composite network the "dry" or "wet" state of the network is chosen each time at the beginning of each event. The network extent is based on the initial wetness conditions, by looking at the total amount of precipitations that fell during the 3 previous days before the beginning of the event. If this amount is over or equal to the threshold of 20 mm of rainfall, we use the "wet" network; below this threshold we pick the dry network. Thus, the estimated value is calculated once, using one or the other of the networks. The process will be more detailed into the revised version of the article.

19 - L351-355 It would be nice to have a table with all criteria, where it is stated which one was removed (and why) and which ones were kept. Maybe the information can be added to Table 5 or 6?!

Few criteria are retained at the end of the regression analysis. We believe it will be clearer to detail more the criteria which are kept and those which are removed, and the associated reasons, directly in the text. We will improve this part accordingly.

20 - L354 Again, it feels as the number of considered events changes among all subsections.

Please refer to our answer to the point 3 of this review for details.

21 - L380 What is the reason for IASYM preference in the Southern part? Due to the steeper areas? I would have estimated Northern part, since the hydrograph would have already been smoothed when originated in the South. Please try to find physical explanations to your results.

This refers to the statement "And for a single station network, the metric I_{ASYM} prefers a station location in the southern part rather than in the northern part."

This statement is wrong, thank you for pointing out. It must be a legacy effect and we apologize. We will remove i) this sentence and ii) the plot of the best 1-station network for I_{ASYM} in the Figure 11 of the article, as of course I_{ASYM} cannot be defined for a single station.

22 - General: Please double-check the abbreviation for "meter above sea level"; I have only seen "m a.s.l." and "m asl" so far, but not "m asl."

Thank you for this observation, it will be corrected using "m asl" throughout the paper.

23 - L155 Benoit et al. 2018 <- a or b? I assume a.

It is 2018a indeed, it will be corrected.

24 - Eq 2, 3, 4 I'm a bit confused what rainfall characteristic is used as input for these equations. Is every raster cell with rainfall used (so I understood it from the text) or only the centre of the rainfall events (as mentioned in Table 1)?

Thank you for pointing out that this part is not so clear. Within these 3 equations we use $P(i,j,t)$, the rainfall amount previously calculated using the stochastic method (section 3.2.1) for each of the 10 x 10 meters grid cell (referenced by i and j) at each 2-minute time step t. The rainfall characteristic and space-time resolution will be specified in the text introducing the first equation.

This remark also reveals that the descriptions of D_{HILLS}, D_{STREAM} and *HAND* in the Table 1 of the paper must be corrected by removing the "mean" at the beginning of each definition (e.g. for D_{HILLS}: "Mean distance of rainfall spatial center of mass to stream network (along hillslopes)" becomes "Distance of rainfall spatial center of mass to stream network (along hillslopes)").

25 - L163 "overlooked" -> ignored

Thanks, it will be corrected.

26 - Eq. 2, 3, 4 The term in the numerator should be put in brackets (Eq. 2: "P(..)dHills" -> "(P(..)dHills)")

Thank you for pointing out this oversight, it will be corrected.

27 - L195 DHAND is not a distance as indicated by the D, and in the text the variable is introduced with HAND. I suggest to stick to HAND throughout the manuscript to avoid confusions with the other two "real" distances".

We agree that the D is confusing. We will stick to the HAND abbreviation throughout the article.

28 - L202 Section 3.5 includes no network extent description. Is it missing in the manuscript?

The network extent is briefly introduced in the section 2, but a description of the composite network is missing. A description will be added in the section 3.4 Rainfall-streamflow response characterization, and the reference L202 will be corrected accordingly.

29 - L268 317.8 mm – Is it areal rainfall amount sum or sum over all stations?

We will specify that the value of 317.8 mm is the areal rainfall amount.

30 - L268-269 please provide also the mean values, not only the highest and lowest values, so that the reader get a "feeling" for the rainfall events.

We agree. The mean values (6.6 mm for the rainfall amount and 2h47 for the rainfall event duration) will be added to the revised version of the paper.

31 - L275 again, please don't use the term average, use mean or median to be more concise. Since Iasym can be positive and negative, the median of its absolute values would be worth to show instead of just the mean, since positive and negative values are levelling out each other.

We agree again that the "average" word is not adapted, and it will be outlawed from the manuscript. Indeed, in this case using the median value of I_{ASYM} (0.025) is better than using the mean. It will be corrected.

32 - Fig. 5 and 6 For a logical order the figures should show the rainfall events first, followed by the discharge plot.

We agree with this point. The figures will be corrected accordingly.

33 - L279 "One strongly asymmetric and high intensity event" -> "One strong asymmetric and very intense event"

Thank you for the suggestion, it will be modified.

34 - L283 A volume can't be fast (check also for later occurrences…)

Thank you, this occurrence and the others will be corrected.

35 - L288 In the sentence before authors mention that the number of events under consideration are reduced by "1", but here again 48 events are studied (also in the following subsections).

Indeed, it is confusing. The line it is referred to at the end of 4.1.1 "This event and its streamflow reaction are excluded from further analysis" will be replaces by "This event and its streamflow reaction are excluded from further analysis involving the hydrological response".

36 - L289 The authors should state what wet and dry networks are. I found it later in the caption of Table 1 in S1, but it would lead to clarifications here. Also, the Table 1 in S1 should be shown in the manuscript, since the written part in Section 4.1.2 is more confusing than explaining for me.

We will include a reference to the states (shown in Figure 1) at the beginning of 4.1.2 Stream network distance metrics. Their extent is first introduced in Section 2.

37a - Fig9 "events without reaction are not shown" belongs to part b), not a). Please correct the caption.

Thanks, it will be corrected.

37b - General: Maybe I missed it, but which temporal resolution was used to calculate the correlation (and other criteria)? 2min as this is the resolution of the rain gauge? Or are values aggregated up to e.g. 1h? This has a high impact on the values of the correlation coefficient.

The temporal resolution of times series used for correlation calculations is 2 minutes. The correlation between events is done at the event-scale. Occurrences will be checked and detailed throughout the article.

38 - L339 “absence of correlation”. Correlation cannot be absent. Better to speak of low correlation or provide absolute values.

Thanks, we will provide values and improve the formulation.

39 - L384-386 “is assessed”, “is evaluated” – two verbs, please rephrase the sentence.

Thanks. The sentence will be corrected by “Considering the small dataset underlying this analysis (23 events), the robustness of the best networks is assessed for two selected metrics (for the P_{ALL} and I_{ASYM}) by re-computing the optimal network when between 1 and 3 events are removed from the dataset”.

40 - L402 “what we previously thought”? What was the hypothesis of the authors before?

This statement refers to “The fact that D_{STREAM} outperforms here D_{HILLS} for the prediction of RC and lag time is an interesting result: it underlines that even in steep environments, with a priori fast instream processes and limited storage, the riparian area and related subsurface exchange processes could play a more prominent role than what we previously thought”.

We will reformulate. The classic hypothesis (e.g. Nicotina et al., 2008) is that in steep environments, the travel time in hillslopes strongly dominates over travel times in the stream network, because instream velocities are very fast compared to travel times in hillslopes. The fact that the travel distance in the stream network explains nevertheless more of the RC variation than D_{HILLS} might be an indirect effect: the longer the travel distance in the stream network, the more likely are delays due to exchange with groundwater in the riparian area. This will be explained in the revised version.

41 - L421 “three station network” It would be nice to provide the resulting density here as well as “(general) recommendation”.

A 3-station network corresponds to 0.22 raingauges per km². This value will be added, and we will check throughout the article if such corresponding density values must be added as well.

REFERENCES

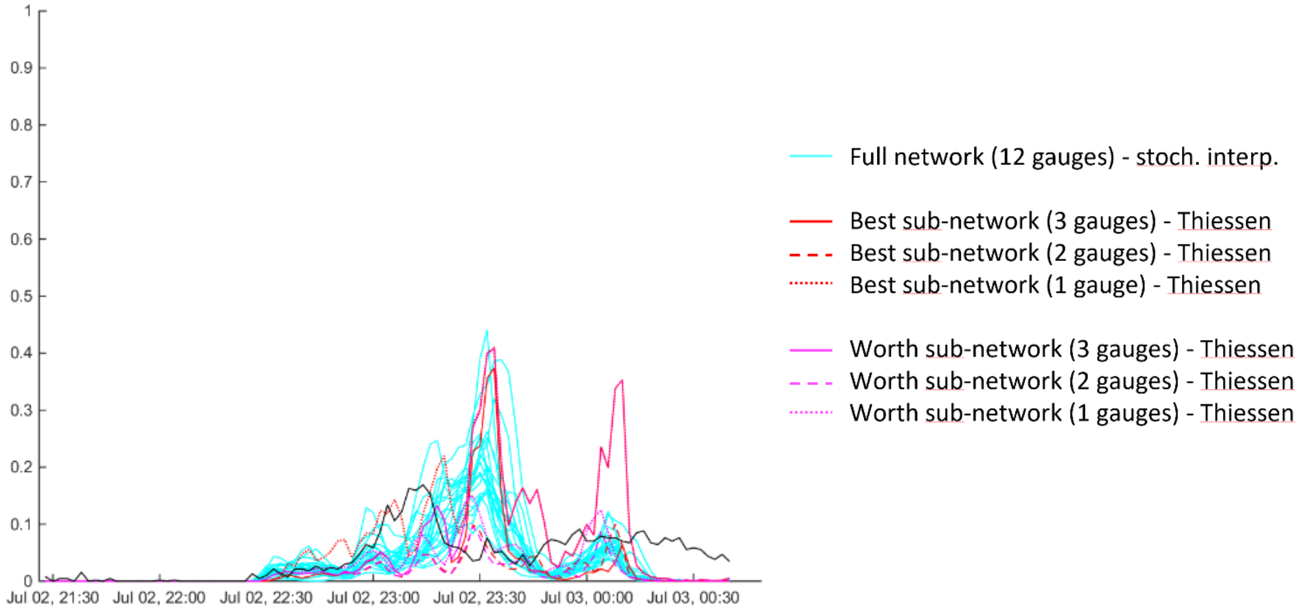
Benoit, L. (2020). High Resolution Stochastic Modelling of Local Rain Fields (PhD Thesis). University of Lausanne, Faculty of Geosciences and Environment, Institute of Earth Surface Dynamics, Lausanne, Switzerland.

Dingman, S.L. Physical Hydrology. 2nd Edition, Prentice Hall, Upper Saddle River, 2002.

Nicótina, L., Alessi Celegon, E., Rinaldo, A., and Marani, M.: On the impact of rainfall patterns on the hydrologic response, Water Resour Res, 44, W12401, 10.1029/2007WR006654, 2008.

Schaefli, B., Talamba, D.B., Musy, A., 2007. Quantifying hydrological modeling errors through a mixture of normal distributions. J. Hydrol 332, 303–315. https://doi.org/ 10.1016/j.jhydrol.2006.07.005.

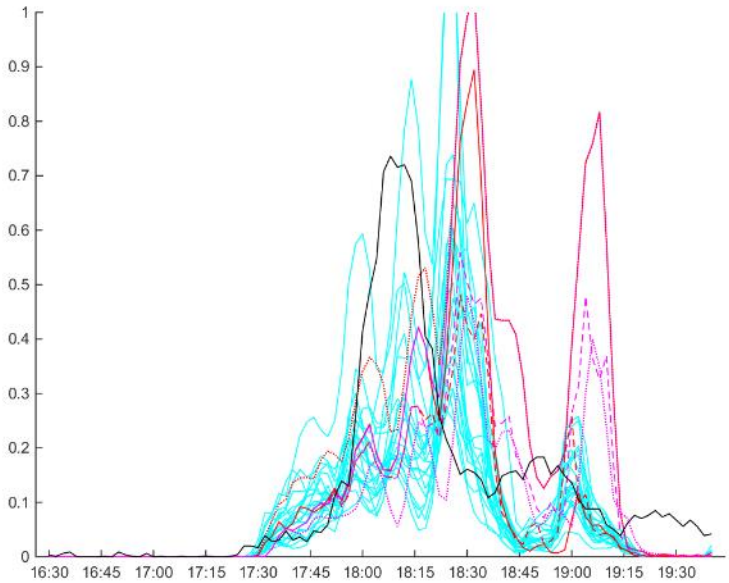
APPENDIX 1: Model results for all of the 15 events



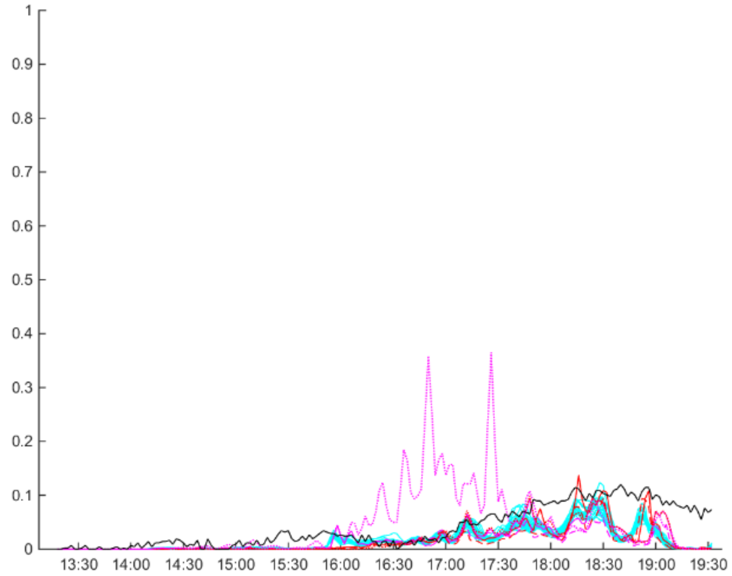
Q event #1 (July 2nd)

On each figure the Y-axis of each hydrograph is in m^3/s.

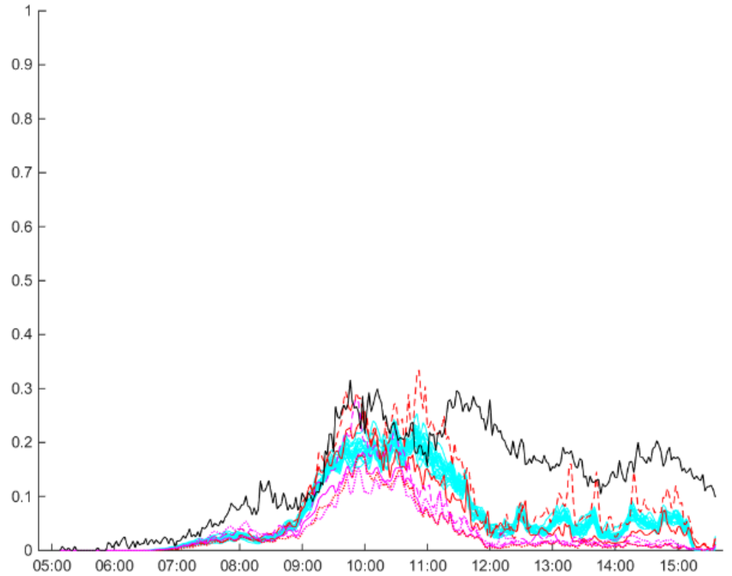
- The black curve is the observed streamflow.
- The 20 blue curves correspond the simulated streamflow based on the 20 possible rainfall fields from the stochastic interpolation method (12-station raingauge network).
- The plain, dashed and dotted red lines are resp. the simulated streamflow using the best 1-station (station #5), 2-station (stations #2 and #9) and 3-station (stations #2, #5 and #11) raingauge network, using the Thiessen polygons interpolation method.
- The plain, dashed and dotted purple lines are resp. the simulated streamflow using the worst 1-station (station #1), 2-station (stations #1 and #3) and 3-station (stations #1, #3 and #4) raingauge network, using the Thiessen polygons interpolation method.



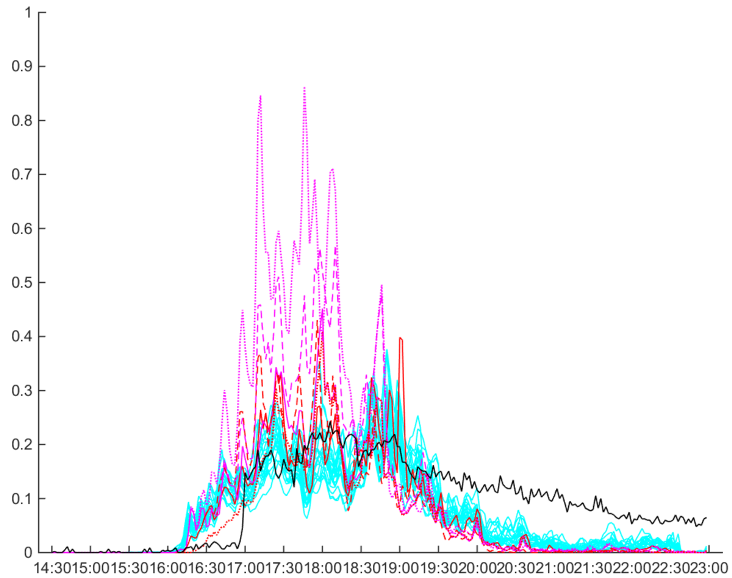
Q event #2 (July 3rd)



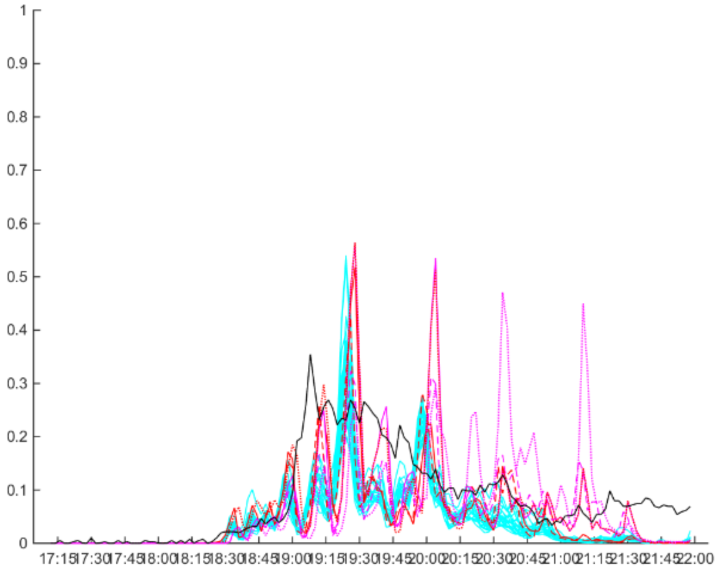
Q event #3 (July 5th)



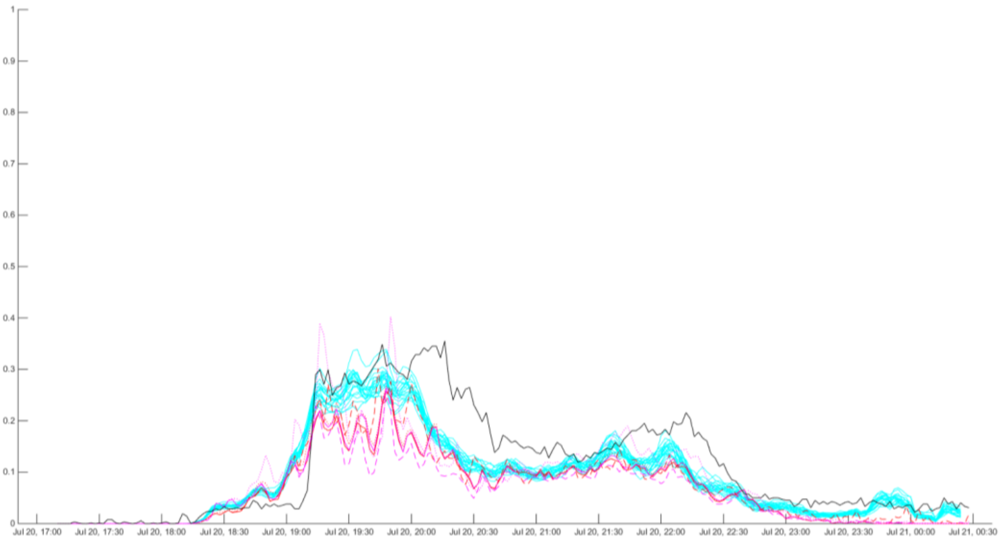
Q event #4 (July 6th)



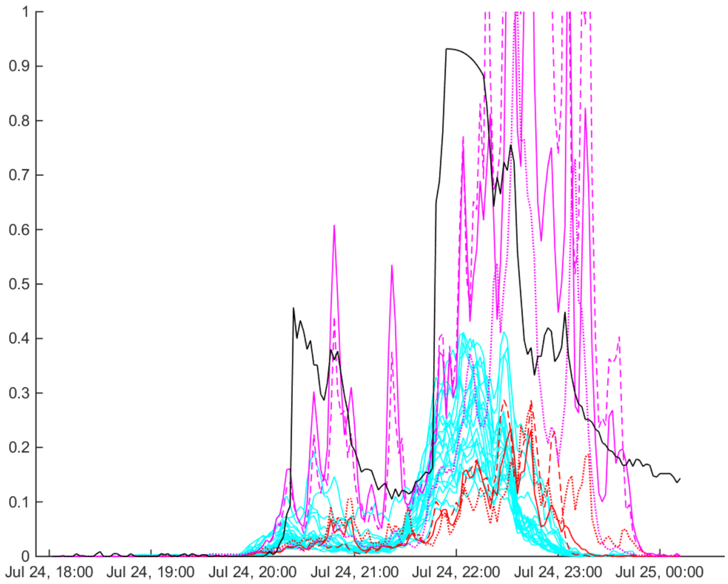
Q event #5 (July 14th)



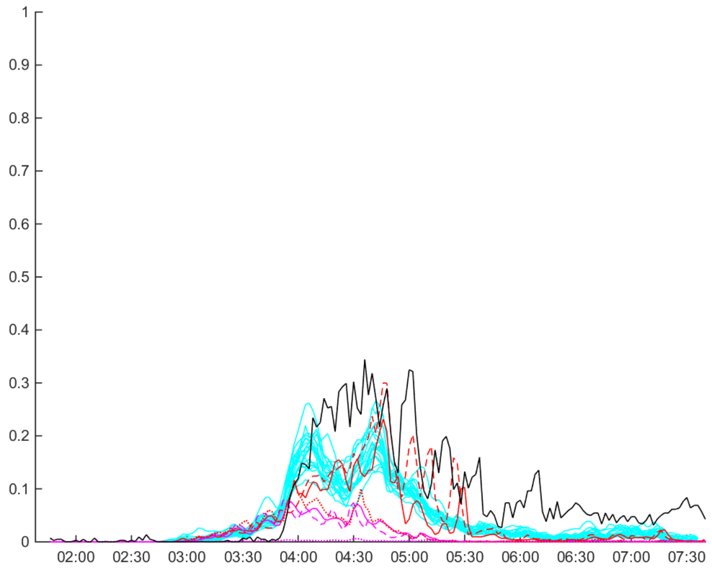
Q event #6 (July 15th)



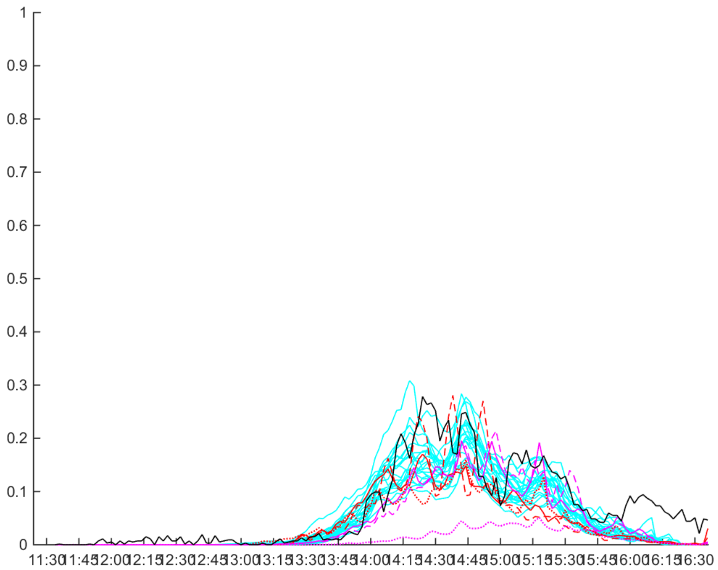
Q event #7 (July 20th)



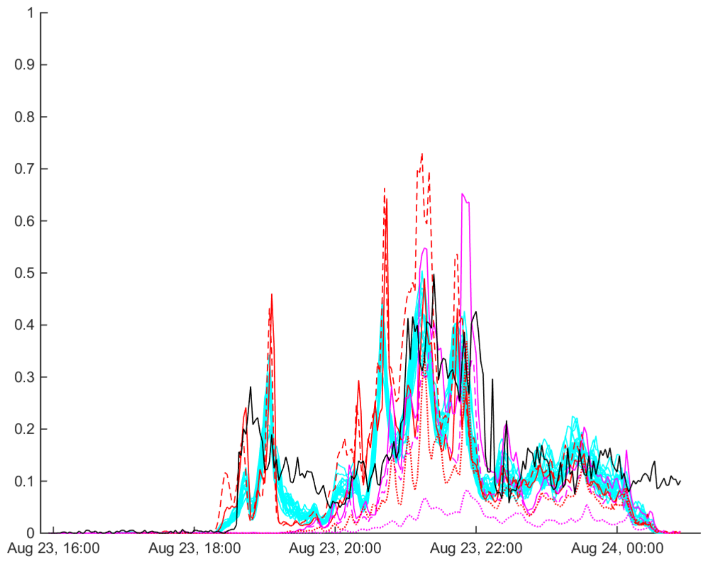
Q event #8 (July 24th)



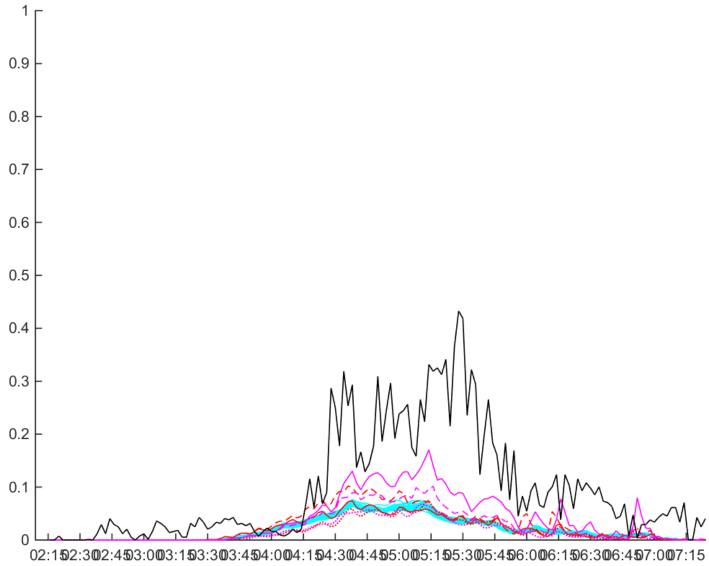
Q event #9 (August 14th)



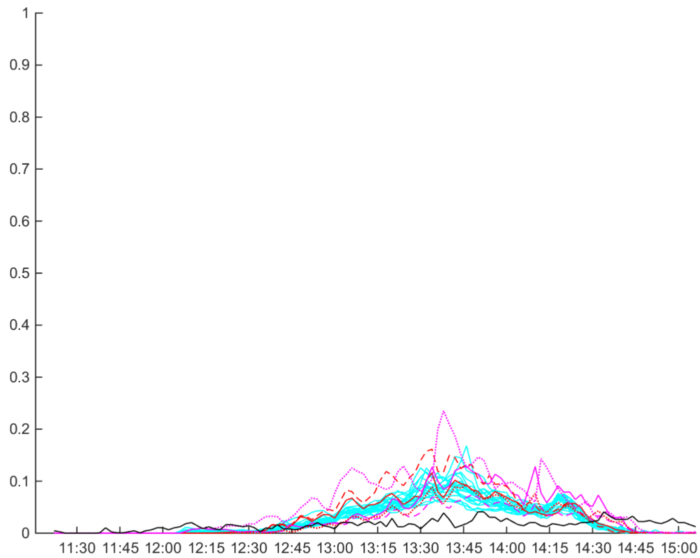
Q event #10 (August 17th)



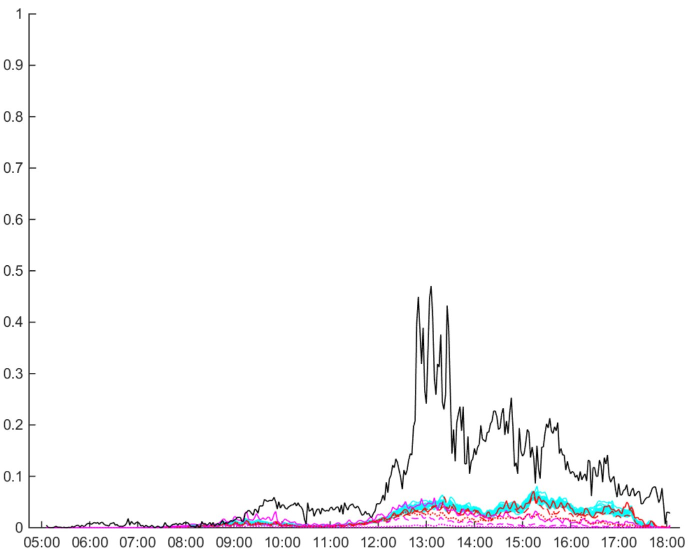
Q event #11 (August 23rd)



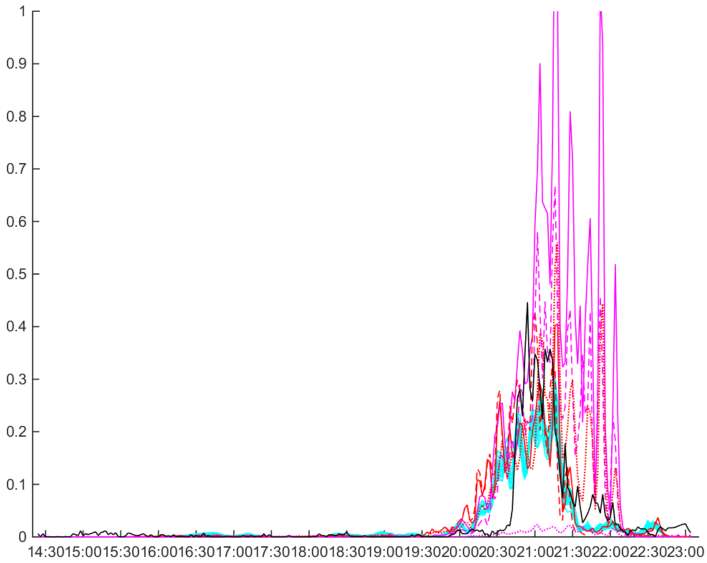
Q event #12 (August 24th)



Q event #13 (August 29th)



Q event #14 (September 1st)



Q event #15 (September 13th)