Authors’ response to Editor decision [hess-2021-628]

We thank the editor for the opportunity to revise the manuscript and addressed the reviewers’ additional comments as indicated below.

In this version we have:

- *reformulated all conclusions about an actual increase in flash floods due to data scarcity*
- *clarified figures and legends (Fig.2, 3 & 4)*
- *made some textual changes based on comments.*

Authors’ response to Reviewer 1 [hess-2021-628-RC1]

We thank the reviewer Ruben Imhoff for his evaluation of our manuscript and his many helpful comments (hess-2021-628). Below we address the reviewer’s comments (full text) indented by arrows and coloured in blue. We appreciate the efforts by the reviewer to take another look at our manuscript and give ideas to further improve it.

General comments

Flash flood database (lines 171 – 179)

In the response to my earlier comment, the authors wrote: *“Before the choice of using a database that was collected through various sources, we analysed discharge data in the region (entire Moselle catchment). Therefore, we collected data for time series as long as possible. Long times series are, however, mainly available for large rivers, such as the Moselle or other bigger stream gauges, but not for catchments, in which flash floods occur. Moreover, data is often only available on a daily resolution. We have conducted several analyses of specific discharge using 79 stations within the region with catchments < 300 km² and found it hard to extract flash floods or high floods from these data. High flows in the past (1980s) were often caused by zonal precipitation in the Vosges mountains. Some regional flash floods that were of major importance and that we know well (i.e. Ernz Blanche 2016 & 2018), were to some extent detected by discharge data, but the overall time series are too short for any long-term analysis. Other events were so small and even outside streams, that they were not even captured by any stream gauge. We concluded that the inconsistencies in this type of streamflow-based dataset would be even bigger than the one presented in the manuscript. Apart from actual flash floods we have also made analyses about the number of scientific reports on the topic, which also started to increase around that time period (beginning 2000), when the topic received more attention. While a better database would be desirable, flash floods rely on site inspections.”*
I think this information is actually very relevant for the reader. Can I ask the authors to put parts of their answer above in the text (either here or in the discussion section)?

→ *We will elaborate on the reasoning for the choice of the database also in the method section.*

Specific comments

Lines 146 – 147 "Unfortunately, the south-western part of the study area is not covered by the RADOLAN data": Perhaps add a reference to Figure 1b here.

→ *Good suggestion, we will add the reference.*

Lines 294 – 296 "Often, soil moisture within the upper and lower soil layer (Swvl10-7 cm, Swvl37-100 cm) is higher during flash flood events compared to general extreme P events (Figure 295 5k, m). The mid-level soil layer (Swvl27-28 cm) shows lower soil moisture before flash flood events (Figure 5l).": These lines still lack some interpretation in my opinion, i.e. do you expect to see these differences between upper/lower and mid layers?

→ *We can extend the speculation in the discussion section a bit (lines 420-427) that the soil is often dry during the summer, but that preceding rainfall events wettened the top layer that eventually hints towards quicker runoff generation in terms of infiltration excess overland flow. We had expected to see wetter top layers and didn't have expectations about lower layers.*

→ *Overall, we however prefer to keep the focus on the atmospheric conditions and not invest too much into soil moisture.*

Lines 301 – 303 "Moreover, sufficient CAPE, high q and weak WS10m-500hPa were identified as the most clearly distinguishing parameters per category to characterize extreme precipitation events, including 75% of all extreme precipitation events and excluding around 75% of all generally occurring parameters values": What about the K-index? No need to change the top three parameters in my opinion, but good to mention the strong signal in this parameter (as the authors already do in their conclusion).

→ *Yes, true, it seems to have gotten lost from your last comments. We will add a small interpretation of the K-Index to chapter 3.4, as this comment especially refers to the strong trends of the K-Index. We did not do that so far, as it is an index and not a more or less independent parameter. Yet, CAPE is also calculated, so we can add the interpretation of the K-Index accordingly.*

Lines 416 – 417 "In recent years they have been increasingly observed, especially in summer (Detring et al., 2021; Lupo, 2020)": You could also add a reference to the July 2021 floods here, for instance Kreienkamp et al. (2021).

→ *While the July 2021 flood mechanism differs a bit from the other, more or less isolated flash floods that we are considering, we agree, that they also occurred during an atmospheric blocking situation and are subject to events within recent years. We will add a corresponding reference here.*

Figure 5 and lines 291 – 293: I think I haven't mentioned this in my previous review, but I can imagine that the difference between the P and FF classes and non-extreme rainfall events might even be larger than the current comparison with "all" classes, as this also included many no-rain time steps (which may also have relatively low wind speeds and shear levels). It is just an idea, but perhaps worth the try if it makes the conclusions stronger.

- → *This is a good idea to split the data into more subsets. However, the rain radar data only starts in 2002, while all ERA5 data is available from 1982. The original idea of including the "all" values, was to give a general idea of what values are observed in the study area throughout the entire time period.*
- → *Identifying light rainfall events is moreover tricky. To identify them correctly, we would have to look for events and merge cells, build averages etc., as we did for the heavy rainfall events. It would moreover be difficult to exclude the 'borders' with lighter rain of the heavy rainfall events.*

Technical corrections

→ *Thank you for your suggestions and pointing out these errors. We will adjust the manuscript accordingly.*

Authors' response to Reviewer 2

[hess-2021-628-RC2]

We thank the reviewer for his evaluation of our manuscript and his many helpful comments (hess-2021-628). Below we address the reviewer's comments (full text) indented by arrows and coloured in blue. We appreciate the efforts by the reviewer, which will help to improve our manuscript.

General comments

Given the temporal inconsistencies in the flash flood reports and lack of a long enough record, I do not believe the linear trend analysis of flash flood occurrence (Fig. 3a) should be included or discussed in depth. The authors even state themselves that "the dataset do not allow drawing conclusions on any robust trends", so why include this figure given its potential to mislead readers? I think it is enough to just state that the linear trend analysis is inconclusive due to the data issues.

- → *We will remove the light-grey dotted trend line and the trend values in Figure 3 (a) and an according sentence within the results description. We will also go through the text again and look for necessary adjustments.*

Additional clarity is needed to distinguish the precipitation events versus the subset that are associated with flash floods. I recommend something like a table to show the number of total precipitation events and the number of precipitation events that are associated with flash floods. That could help make the results more generalizable.

- → *Thank you for your suggestion. This would give a tiny table, that could look like this:*

	P events	*P events associated with FF*

No. of events	*3835*	*37*

→ *We will add the values to the manuscript.*

Specific comments

Lines 51-53: In the U.S., there are nice definitions of flash floods used by the National Weather Service- perhaps you can utilize that or something similar that exists in Europe? In the U.S., the NWS defines a flash flood as "a rapid and extreme flow of high water into a normally dry area, or rapid rise in a stream or creek above a predetermined flood level, beginning within six hours of the causative event" (NWS 2021.)

→ *This is an interesting and wide enough definition we had not yet come across. While we consider much smaller dimensions of floods in central Western Europe compared to the US and the Mediterranean, the character is the same as described in the definition of the NWS glossary https://forecast.weather.gov/glossary.php.*

Line 59-60: I do not believe this is entirely correct description of storm training- please revise this sentence to reflect that "echo training" is when convective cells move in the line-parallel direction leading to repeated cell motion over an area (Peters and Schumacher 2015): Peters, J.M. and R.S. Schumacher, 2015: "Mechanisms for organization and echo training in a flash flood-producing mesoscale convective system". Mon. Wea. Rev., 143, 1058-1085. Doi: https://doi.org/10.1175/MWR-D-14-00070.1.

→ *Thank you for your comment. We will revise this sentence in the manuscript.*

Lines 62-64: Likewise, this sentence needs revising, as forward movement is not halted. Rather the direction of the cell motion and propagation vector cancel out leading to new cells being continuously generated over the same area (Doswell et al. 1996).

→ *Thank you for your comment. We will revise this sentence in the manuscript.*

Line 103: I recommend revising this sentence because the more intense thunderstorms are actually triggered because of high CAPE **and** high CIN.

→ *Thank you. High CIN is often connected to isolated thunderstorm cells rather than large scale thunderstorm conditions, that would potentially lead to the rainfall amounts required for flooding. High CIN mainly indicates that strong lifting mechanisms are needed. High CAPE, however, always has the potential for intense thunderstorms due to strong possible updrafts of a cell.*

→ *We will revise this to "higher CIN levels may lead to higher CAPE values since it prevents premature initiation of convection potentially inhibiting the development of stronger CAPE, and thus possibly increasing the potential of more intense storms."*

Fig. 1d: What is this panel? It is not labeled in the figure caption. Please either omit or add its description to the caption.

→ *Sorry, we forgot to adjust the caption after adding the fourth panel. (d) is what is written for (c). (c) was added new and is an actual digital elevation model of the area at a 1x1 km resolution. We will adjust the caption accordingly.*

Line 150: Do the supplementary rain gauges cover the same time period?

→ *Not all rain gauges cover the same time period as the network became denser over the last years. As we wrote in the lines 155-158, the number of extra rain gauges ranged from 7 to 40 extra rain gauge stations in Luxembourg and 10 to 50 extra rain gauge stations in Germany. Of course, there is an inconsistency but that is always the case when using rain (gauge) data. The physical rain gauges and the environment (wind field, vegetation) around the stations change for example. Nevertheless, the rain gauges are just one part in the processing chain of data adjustment. The main data source is the radar data. Although the radar hardware also changes, as well as the basic radar (internal) quality control and correction algorithms. Thus, a constant time series is an unreachable ideal conception, but we used all sources we could to adjust data as close to reality as we could.*

Line 169-170: I think it would be helpful to at least briefly describe this procedure in one sentence, like you do in the figure caption below.

→ *Ok, we suggest to add: For a small standard P event that lies within one ERA5 grid cell, atmospheric data was averaged over that particular ERA5 grid cell and the eight surrounding ones. Precipitation events at the boundary of the study area do not include the full buffer zone and larger P events covering multiple grid cells include a buffer zone around the ERA5 grid cells of the actual P event.*

→ *We will also add to line 163: "For every P event, we extracted the maximum hourly precipitation intensity at one location within the P event as well as the maximum 5-minute precipitation intensity at one location within the P event."*

- *5 minute max: within one grid cell at one time. Not averaged in space over the entire size of the P event*
- *1 hour sum: moving window over time within one grid cell. Not averaged in space, only at one location*

Line 175-176: How did you determine the spatial threshold of 30 km? Likewise, how did you determine the temporal threshold of one day? What happens if you have multiple hours of precipitation (which count as separate events according to your definition) and one flood?

→ *The 30 km are to reach the next ERA5 grid cell in case no P event > 40 mm/h was identified in the one where the flood occurred. At the actual flooding location, which is not a point either, the hourly precipitation intensity might have been just below the determined threshold or the flood occurred a bit downstream of the P event. If a flood was triggered by a rainfall event not identified as extreme in the radar data, the flood was not considered.*

→ *There would have to be a major temporal and spatial gap (> ½ h, > 2 grid cells = 2 km) in the intense precipitation to make them count as two events. In this case the first would "only" contribute to the pre-event moisture according to our definition and the one closest to the* *flood event caused the flood. By looking at the few floods individually we did not find discrepancies regarding this.*

Figure 2: This is a very helpful figure, although I am a bit confused about the difference between the dashed and solid lines- is one dashed box an ERA5 grid cell and one solid box the multiple ERA5 grid cells used to take the atmospheric condition? If so, please make that clear in the figure description.

→ *Yes, that is what it is. We will improve the legend and make this clearer in the figure description.*

Line 207: How did you obtain 0.5%? Did you calculate it yourself or did you find it in the literature?

→ *We calculated this value ourselves based on the ERA5 data of the area. We will state this clearly.*

Line 263-264: it is difficult to tell from Figure 4, but to me it looks like the median line for max hourly intensity is actually higher in flash flood events than for all P events- can you please see if this is true and provide numbers for these values?

→ *It is true that the median differs from 46.54 mm/h as the mean of all P events and 49.65 mm/h as the mean of all P events leading to flash floods. We will rephrase the sentence to: "P events that eventually led to flash floods (Figure 4c, e) do not differ in the range of precipitation intensities from P events that did not cause flash floods, but their median."*

Figure 4a: I believe the text that states "P events associated with flash floods" is incorrect here, because the text states that these are all P events in the summer- is that true?

→ *We will adjust the legend to be clearer. The blue crosses are indeed the P events associated with flash floods. These are however not shown in Figure 4 (a), just in the other panels. This line of the legend will be moved outside the plot panel.*

Line 294-296: This is a very interesting result!

→ *As indicated by the other reviewer, we add a few sentences of interpretation to the discussion section. Especially the higher moisture in the top soil layer hints to preceding rainfall events and might help explaining some of the quick runoff formation present during flash floods. We however prefer to keep the focus on atmospheric parameters and not go into too much detail regarding the soil moisture.*

Table 2: Is this table for all extreme P events or just those that are associated with flash floods? It would be interesting to show the values for both events.

→ *Table 2 refers to all extreme P events, independent of the occurrence of floods. We rate the threshold values of P events causing flooding, as less statistically robust, as only 37 events would contribute to their calculation. Moreover, they might confuse the reader as to which* *thresholds were used. We have displayed the thresholds for both in the extended table below but prefer to not add the values to the manuscript.*

→ *We will update the table description to be clearer about the values to: "**Table 2:** Threshold values determined as extreme precipitation and flash flood favouring based on the lower/upper quartile of their range of occurrence during extreme precipitation events, including all P events, whether they are associated with a flood or not.*

	Instability			Moisture			Storm motion & organisation			
	CAPE	CIN	Kx	TCWV	q	RH	$WS_{700\ hPa}$	$WS_{10m\text{-}500hPa}$	LLS	DLS
P	≥ 326.9 J kg^{-1}	≤ 183.5 J kg^{-1}	≥ 27.8 °C	≥ 26.5 kg m^{-2}	≥ 0.004 kg kg^{-1}	≥ 59.4 %	≤ 7.1 m s^{-1}	≤ 6.2 m s^{-1}	≤ 3.8 m s^{-1}	≤ 10.4 m s^{-1}
FF	≥ 355.2 J kg^{-1}	≤ 126.4 J kg^{-1}	≥ 27.6 °C	≥ 26.0 kg m^{-2}	≥ 0.004 kg kg^{-1}	≥ 63.4 %	≤ 7.6 m s^{-1}	≤ 6.5 m s^{-1}	≤ 4.4 m s^{-1}	≤ 12.0 m s^{-1}

Line 389: Results are either significant or not- please pick one.

→ *We will rephrase the sentence to: "Increasing trends in low LLS are significant in the south-eastern part of the study area."*

Line 389-390: I don't believe you can state that the storm organization is unchanged, as you did not explicitly study changes in storm structure.

→ *Thank you for pointing this out. We will rephrase to: "Overall, the proxy parameters used for the assessment of organisation and motion of storm systems stayed largely unchanged with tendencies favouring the occurrence of extreme precipitation."*

Line 391: Future studies actually show a decrease in shear with warming and it would be helpful to cite those studies here (Diffenbaugh et al. 2013, Brooks 2013).

- Brooks, H.E., 2013: Severe thunderstorms and climate change. *Atmospheric Research*, **123**, 129-138. https://doi.org/10.1016/j.atmosres.2012.04.002.
- Diffenbaugh, N.S., M. Schere, and R.J. Trapp, 2013: Robust increases in severe thunderstorm environments in response to greenhouse forcing. *PNAS*, **110**, 16361–16366, https://doi.org/10.1073/pnas.1307758110.

→ *Thank you for these references. We will add them.*