Authors' response to Reviewer 3

[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

A major limitation of this study is the lack of flash flood events, particularly before 2006, and how they are identified. While the authors acknowledge this limitation in lines 332-333, I wonder if this is not an issue of a lack of flash flood events in the past, but a limitation of the observational record they use to define flash flood events. Flash floods are defined by news reports, prior literature (I am guessing case studies?), water agency reports, and reinsurance data, which are all prone to human error, including the need for people to observe the flood and report it as noteworthy. I wonder if they can incorporate any physically based indications of a flood event by including streamflow observations. This could address the dearth of floods prior to 2006 and remove some of the inherent bias in the human-based indications of flooding.

- → *As mentioned in detail in the responses to RC1 and RC2, we had collected a discharge database for the entire Moselle catchment, before starting the collection of data through newspapers, reports, etc.. From this data, it was impossible to clearly isolate flash floods in a statistic way. Moreover, the data is not consistent in time either, as long time series are available mainly for large rivers, but not the ones, in which flash floods could occur. Many of the now counted events also occurred in areas without measurements. Therefore, as cleanest option, we offer to take back the sub hypothesis 1 about the trend in flash floods and focus on the atmospheric background during the identified events.*

The connection between the flash flood events and extreme rainfall is unclear in the present form of the manuscript. The only description of how these events is linked are in lines 172–173 where the authors state that they include all floods that directly follow extreme precipitation. What is the temporal scale used to determine what "directly follows" means? Is this an hour, several hours, or a day? Please be explicit in stating this. Also, what is the spatial requirement for a flood event being connected to an extreme precipitation event? Please describe this in more detail. Finally, the independence of flood and precipitation events must be discussed. For example, if a flood event occurs on two consecutive days, is that considered the same event? Again, please discuss this in more detail.

- → *While it is mostly impossible to quantify the exact lag time for most events, as we don't know the time of the flood peak, "within a couple of hours" should cover all events. We will be more precise with that.*
- → *Regarding the spatial requirement: In order to connect a precipitation event and a flash flood event, they have to occur both within the same RADOLAN grid cell (1km*1km) for now. We will however adjust this, as about half of the flash floods failed to be connected to a precipitation event. We will state the new method of matching events precisely in the method section. We will check if a e.g. 5 km radius can already be sufficient to identify related rainfall events.*
- → *For now, precipitation events are grids that cross the 40 mm/h threshold. These are accumulated to the size of ERA 5 grids. If it is crossed during 3h in a row it is for now counted as 3 rainfall events. This will be adjusted for the revised version. We will calculate events more precisely in their temporal and spatial resolution to make sure that one event matches a real*

event. Regarding the further approach of identifying the atmospheric conditions during P events, we will stick to the approach of extracting atmospheric parameters of every ERA5 grid in which at least one RADOLAN grid cell has exceeded the threshold.

→ *As we didn't use discharge data for floods, but reports only, we didn't count an overnight event twice. In one situation in Luxemburg, two floods occurred within two weeks. These were however two independent events (except for the elevated soil moisture and meso-scale atmospheric condition). If two floods occur in neighbouring catchments, they were also counted as two events*

It seems odd to me that despite an increase in more favorable rainfall environments over time, little trend is observed in changes to extreme rainfall. While the authors discuss possible reasons why this is in the discussion section, I wonder if it would be helpful to include other sources of rainfall data (like rain gauges or even ERA5, despite their limitations), to see if the same lack of a trend is reproduced.

→ *As also mentioned more in detail in the reply to RC2, we have analysed daily precipitation station data within the Moselle catchment in a previous approach (Meyer et al., 2020). As we could not identify trends in the stations, where long term data (since 1954) was available, we assumed, that the resolution was too coarse and we had missed too many thunderstorm events in the coarse network. Therefore, we choose to analyse the radar data. We do believe that the limitations of ERA5 data are indeed too high to be a valid additional source. The resolution is very coarse, such that thunderstorms and localized precipitation maxima are not properly simulated and thus significantly underestimated.*

→ *Meyer, J., Douinot, A., Zehe, E., Tamez-Meléndez, C., Francis, O., and Pfister, L.: Impact of Atmospheric Circulation on Flooding Occurrence and Type in Luxembourg (Central Western Europe), EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-13953, https://doi.org/10.5194/egusphere-egu2020-13953, 2020*

Specific comments

Lines 59–62: There are some contradictions in these lines as to how you refer to precipitation events that trigger floods. In line 59, you state that they are characterized by high rainfall amounts over a short period of time, while in the next few lines, you say that the rainfall also lasts over longer periods of time. What do you mean here? Please be clear if these are short or long duration rainfall events. Perhaps providing typical durations could be helpful here.

→ *This sounds indeed contradicting and we will reformulate this to be clearer. We are referring to time scales from 30 minutes to a few hours. The first "short" could be replaced by "sufficient" and the second, we can try to quantify to the above-mentioned time.*

Lines 67–70: I am not sure I understand what you are saying here–which processes are you referring to? Do you mean the upscale growth of convective cells into organized convection, like a mesoscale convection system? Please be more specific.

→ *Yes, upscale growth means that individual cells merge to form a mesoscale convective system (multicell storm). This can happen in an organised way along a well-defined boundary or in a more chaotic way (random cell clustering). We will revise the text to be clearer. Thank you for pointing this out.*

Line 80–81: I would also cite Schroeder et al. (2016) regarding a larger warm cloud depth leading to higher precipitation efficiency.

Schroeder, A., et al., 2016: Insights into atmospheric contributors to urban flash flooding across the United States using an analysis of rawinsonde data and associated calculated parameters. J. Appl. Met. and Clim., 55. Doi: https://doi.org/10.1175/JAMC-D-14-0232.1

→ *Thank you! We will add this source here.*

Lines 83–84: Large rainfall systems can also result in long duration storms (Doswell et al. 1996).

→ *We will rephrase more precisely and comprehensively.*

Line 88: I would start a new paragraph here at "Proxy parameters.." since this paragraph is already quite long.

→ *Thank you, we will consider restructuring this paragraph.*

Lines 91–93: The results of this study seem to contradict your previous sentence stating that bulk wind shear can be used to estimate precipitation efficiency if heavy precipitation occurs over a variety of DSL values. How do you reconcile this conflict?

→ *Weak DLS is present during weakly organized and slow-moving storms. In stronger DLS, cells are organized better and training/back-building multicell storms are possible. In our study area, the first case seems to be the dominant one. Nonetheless, this is a good point, and we will point out the precipitation efficiency and shear conflict a bit more as also already shown by Fankhauser (1988)*

→ *Fankhauser, J. C. (1988). Estimates of Thunderstorm Precipitation Efficiency from Field Measurements in CCOPE, Monthly Weather Review, 116(3), 663-684.*

Line 114: Please also cite Rasmussen et al. (2017), as they were among the first to discover the increasing CAPE/decreasing CIN paradigm:

Rasmussen, K. L., A. F. Prein, R. M. Rasmussen, K. Ikeda, and C. Liu, 2017: Changes in the convective population and thermodynamic environments in convection-permitting regional climate simulations over the United States. Climate Dyn., 55, 383–408, https://doi.org/10.1007/S00382-017-4000-7.

→ *Thank you for the reference, we will add it.*

Line 175–176: Are this 8 neighboring grid cells centered around the precipitation event grid cell? What if the precipitation event takes up multiple grid cells?

→ *This sentence as well as the approach are indeed confusing and will be revised. What we did, was picking the ERA5 grid cell at the location of the flash floods at the time the precipitation threshold of 40 mm/h was exceeded for the first time. As many flash floods were not connected to a P event with this method, we will improve our approach by looking at the neighbouring ERA5 grid cells as it is described in the manuscript. Generally, we had counted one precipitation event accumulating all 1*1km RADOLAN grid cells exceeding 40 mm/h to the margins of an ERA5 grid cell. For the sub-hypothesis 2, we will accumulate precipitation cells* *exceeding thresholds better, independent of ERA5 cells, but dependant on their actual spatial and temporal resolution within the 1*1km grid width. This should however not impact the choice of ERA5 data, that is described here and will refer to the closest ERA5 grid cell in which the precipitation threshold is exceeded. From this grid cell, we picked the atmospheric data. If an event took up multiple grid cells, only the closest one was considered.*

Lines 205–215: Did you perform a sensitivity test to see if taking the RH or winds at different pressure levels aside from 700 hPa affected your results?

→ *This is a good point, we have added a supplement with analyses for the pressure levels of 500 hPa and 850 hPa. Looking at differing pressure levels gave a good overview over tendencies in the atmosphere. The positive trends in specific humidity (q) remain the same, are however less strong and less significant at 850 hPa compared to 700 hPa. At 500 hPa the original threshold of 0.004 kg kg^{-1} is not crossed. We did not identify new thresholds depending on pressure levels for this supplement. While trends do differ in a few grid cells for RH, the overall conclusions drawn from the proxy level of 700 hPa remain valid, as results are not significant. Only for the absolute values of relative humidity (RH) above the threshold of 50% we do see a significant decrease at the lower atmospheric level of 850 hPa that is stronger than at higher levels of the atmosphere. For wind speed, we see trends, that are stronger negative at 500 hPa and therefore also for the mean between the surface level of 10 m and 500 hPa. These trends are however insignificant as well. At 850 hPa, decreases in wind speed are a bit more significant, but at a small range. As we concluded that there are no significant trends in the parameters selected regarding system motion and organisation, the conclusion can be kept also under consideration of the other values. The same stays true for the moisture parameters.*

Line 240: how would your results look if you omitted the year 2016? Would it still be an increasing trend?

→ *As graphically shown in the response to the second reviewer, a positive trend in the flash floods would remain when eliminating both, 2016 and 2018. Therefore, we assume, that it would also be the case when only omitting 2016.*

Line 258: what does "all hourly values" refer to? Is it the time of the precipitation event and the flood event combined or something else? This needs to be described in more detail. Also, how long, on average, are the precipitation events and the flood events?

→ *"All hourly values" are all values as available in the ERA5 dataset independent of P and FF events, so including P and FF events but also all other times without any event. We will rephrase this to be more precise.*

→ *The precipitation events are 1 hour long. If the threshold was crossed in 2 hours following each other, it is counted as two events. While this is not a natural event definition, we will reshape this for the second sub-hypothesis and also consider this for the ERA5 data. The flash flood events do not have a defined duration.*

Lines 261–262: Half of the distribution is above 100 J Kg^{-1} of CIN, which is moderate CIN, so I do not believe saying high values of CAPE are often accompanied by low values of CIN is entirely accurate.

→ *We agree and will rephrase this sentence to "Sufficient values of CAPE are often accompanied by moderate values of CIN."*

Lines 273–274: Given the low wind speed and weak DSL, this likely indicates that these storms are slow-moving single-cell thunderstorms. This is interesting, because many flood-producing storms tend to be larger and more organized mesoscale convective systems (Ashley and Ashley 2008; Schumacher and Johnson 2006):

Ashley, S. T., and W. S. Ashley, 2008a: The storm morphology of deadly flooding events in the United States. Int. J. Climatol., 28, 493–503, https://doi.org/10.1002/joc.1554.

Schumacher, R. S., and R. H. Johnson, 2006: Characteristics of U.S. extreme rain events during 1999–2003. Wea. Forecasting, 21, 69–85, https://doi.org/10.1175/WAF900.1.

→ *Indeed, as also mentioned in one of the previous comments above. Thank you for pointing out this difference as well as referring us to the related references. We will add a paragraph about this in the discussion.*

Line 282: CAPE at or exceeding 100 J kg-1 is not high, but rather weak. I recommend you edit language throughout the paper to reflect this.

→ *This is true that CAPE of 100 J kg^{-1} is not extremely high, but it is in the upper quartile of most values. We will consider revising this to "sufficient CAPE".*

Technical corrections:

→ *Thank you for your comments. We will change the text accordingly.*