Anonymous Referee #3

General comments:

The paper summarizes a comprehensive compilation of heavy precipitation events (HPEs) in the Eastern Mediterranean (EM) based on high-resolution radar data and WRF simulations. This set of events can be representative of the climatology in this area, and is used to quantify the spatio-temporal characteristics of HPE, and the ability to numerically predict the patterns of HPEs. A collection of four diagnostics are used to typify and contrast the radar-based and WRF spatio-temporal precipitation patterns. The events are further classified according to the synoptic situation responsible for the HPEs: namely, Mediterranean cyclone (MC) and active Red Sea trough (ARST). This topic is important as it serves as a benchmark for using numerical weather prediction for flood forecasts, as well as for downscaling of future climate projections. Overall, the presentation is very good, although some excessive text can still be made easier to read, as I suggest in the following. Two major weaknesses of the results and their organization should be fully addressed before the paper can be considered for publication, as detailed in the first two Specific Comments below. Other specific comments should also be clarified.

We thank the reviewer for the comments provided, and for the time and efforts spent in reviewing the manuscript. We have carefully addressed all of the comments and we believe the revised manuscript will benefit from them.

Specific comments:

1. An important distinction is made between HPE under ARST and MC. However, the classification is not maintained throughout the results, having in mind the double goal of the paper: (i) characterize HPE patterns and (ii) evaluate WRF performance. In its current form, the classification is merely mentioned, while referring to previous works on the different spatio-temporal patterns under MC/ARST, but this is not directly shown here, with an exception of Fig. 11a-f. As mentioned in the text, HPE related to ARST is harder for prediction because of the local characteristic convection which dominates the patterns. On the other hand, HPE under MC is characterized by a cold front structure. To enhance the presentation of the results in light of the MC/ARST classification, and to make the discussion and conclusions robust, I suggest to (i) show the spatio-temporal patterns separately for each group (ii) compare the radar/WRF bias between HPE-ARST and HPE-MC. The two aspects can be achieved by reorganization of the presentation of the results, and showing figures such as Fig. 2, 3, 6, 7, 11g and 12 in light of the classification. By doing this, it will be interesting to see if there are consistent differences in the model performance, and substantiate the discussion in Lines 449-458.

We thank the reviewer for raising this point. We agree the distinction between ARST and MC is important both for HPE pattern characterisation and for the ability to forecast the events with a NWP model. Accordingly, we will modify some of the figures, as detailed below, to present this distinction, and further discuss it in the discussion section.

Fig. 2 & 3: HPE identification is based on specific rainfall thresholds that do not take the classification into account but rather the local quantiles (Sect 3.3). We do not think these thresholds should be defined with classification, since it will reduce their robustness. In addition, the distinction between regions that are better observed by the radar (Fig. 3) would

not benefit from synoptic classification. Therefore, synoptic distinction is not relevant for Figs. 2 and 3.

Fig. 6: Although, in principle, the FSS median and range shown in Fig. 6 for all HPEs can be computed for each synoptic type separately, it should be noted that we deal with only 6 ARST-type HPEs, out of them, two are not well simulated. Since we cannot provide a reliable statistic for ARST type we would not include in Fig. 6 the distinction between the two types. We do plan however to refer in the text to some general differences, qualitatively identified, from FSS analysis of the individual HPEs for each type.

Fig 7: We will add the synoptic distinction to the SAL analysis presented in Fig. 7 and in the text discussing these results.

Fig. 10: We will add the synoptic distinction to the figure.

Fig. 11: The DAD analysis is already classified into the two types of synoptic circulation patterns, however, to make a better distinction between rain-fields based on their duration of accumulation and their source (radar-QPE / WRF), we extracted the median curves from each one of the sub figures (a-f) and enlarged them in panel g. We feel that adding a synoptic distinction to this panel, may attract the attention of the reader from the distinction between durations and the source of the rainfall, which was the purpose of panel g.

Fig. 12: We will add the synoptic distinction to the figure and discuss its results.

2. Two individual HPE events are shown in more detail. They are important to get a better grasp of the patterns and the model/radar biases and the diagnostics used. It is, however, remaining unclear if the reader should take these results as representative, and if so, of what. It is mentioned that HPE #1 is of MC type, while HPE #5 is ARSTtype. Are they representative of the two types? Since both cases perform badly in terms of the SAL diagnostic, why do you focus on them? As the message of the work is to demonstrate the overall good performance of WRF, I find this confusing, and suggest to also illustrate the point with a case where WRF performs representatively well. I suggest to clarify this issue by explaining the rationale behind choosing to focus on these events. Further, it will make an easier reading to mark the chosen events onto Figs. 6,7,10,11.

The two events shown in Figs. 4, 5 are meant to represent one well-simulated event (event #1, shown in Fig. 4) and one poorly-simulated event (#5, Fig. 5). It seems there was a confusion with the two poorly simulated events (#5 and #20, Table S1) discussed later on, but this was not the intension; we will modify the text to better clarify this point (see below). The two events shown in Figs. 4 and 5 are given as an example to show what the model is able (Fig. 4) or unable (Fig. 5) to simulate, and they also exemplify a typical MC and a typical ARST cases. It turns out that it is harder for the model to represent the localised rainfall that often happens during ARSTs. The different performance for the two cases is very clear from the SAL analysis (see figure below).

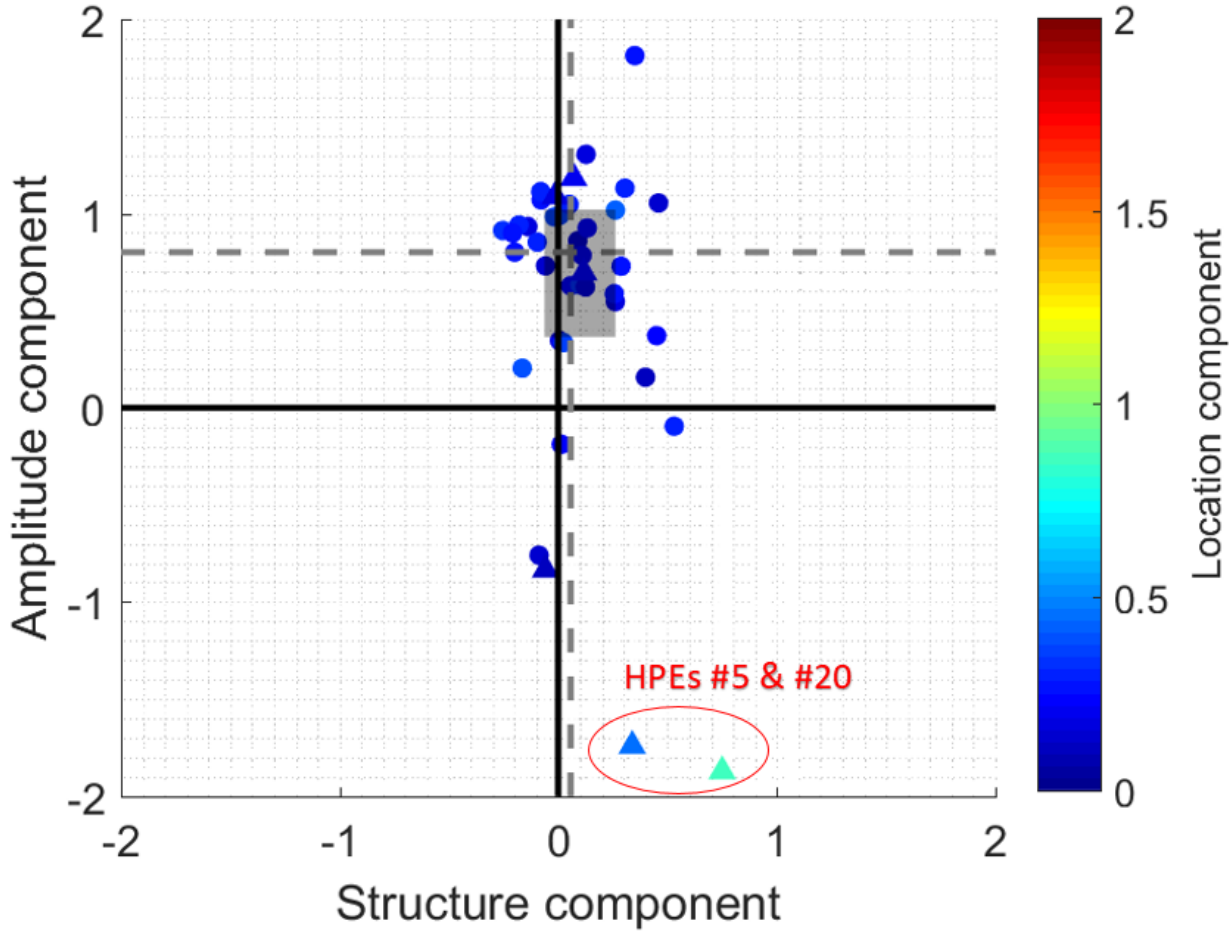

Figure: SAL analysis of the 41 HPEs. MC-type of HPEs denoted with circles and ARST-type with triangles.

In light of your comment, we will better clarify in the revised manuscript the purpose of the closer look at these two events (Section 4.12): "…localised rainfall. Fig. 4a-c presents as an example a well-simulated HPE case (event #1, Table S1). In addition, the distributions of rainfall among pixels is generally well represented (Fig. 4d). At the same time, pixel-based comparisons are deemed inappropriate for such an analysis, as shown in the scatter plot (Fig 4e). Most of the examined HPEs led to similar observations with the exception of two HPEs in which the WRF model clearly failed in representing the rainfall patterns. An example for such a poor simulation is given on Fig. 5 (event #5, Table S1)".

3. Table S1 and Fig. 8: How is HPE duration calculated, and what does it mean if an HPE has a 48-h duration but no shorter durations (e.g., HPE #6)? This is confusing and should be clarified. Consequently, the results in Fig. 8 are confusing, and it is not very clear to me what we can learn from this figure.

The term "event duration" in Fig. 8, Table S1 and possibly in other sections of the paper, does not refer to the total duration of the event but rather to the duration according to which it was selected as HPE. We defined HPEs, in Sect 3.3 by "the exceedance of local, quantile-based thresholds over a sufficiently large area… For a set of durations between 1 and 72 hours we defined the threshold as the 99.5th quantile of the non-zero (i.e. >0.1 mm) hourly amounts observed in each radar pixel… we classify as HPEs all time intervals during which at least 1000 pixels (i.e., 1000 km2) exceeded their local threshold". This is to say that, if enough pixels in the radar archive exceeded their own threshold, for a given examined duration, we defined this event as an HPE for this duration. Obviously, a given event can be selected for several examined durations. Fig. 8 shows that it is hard to separate events according to their duration, i.e. the duration for which rain intensity was exceeded the threshold, because of the above overlapping. However, we see that it is not clear enough. We will clarify it in the text, and possibly move Fig 8. earlier, to relate also to the details in Sect. 3.3.

4. Section 3.5.4 is difficult to understand, and the description of the 2D autocorrelation field, its ellipticity and orientation in Lines 379-392 is not also not clear when not referring to a figure. Please enhance or clarify these parts, possibly with an illustrative figure, such that the analysis can be standalone without referring to the references.

We agree and make sure the description is clear enough by its own in the revised manuscript. However, we do not want to add a lot of text to describe what was already published (e.g., in Marra and Morin, 2018). Therefore, we would try to describe better than we did, but still briefly.

5. I suggest to move the spatio-temporal characteristics in Fig. 9 and 10 to earlier on in the text, even to when presenting the list of events in Sec. 3.4. This seems more natural to understand the events characteristics before assessing the model performance.

We agree that it seems more natural to talk about the characteristics of events before presenting the model performance. However, our goal is to characterise HPEs and to evaluate model capability in reproducing those characteristics. Therefore, we chose to present the examined characteristics from both radar and model, and thus these analyses come after model performance results. However, accounting for suggestions from all reviewers, we would like the re-order the results section as follows:

a) General properties of HPEs (presently shown e.g., in Fig. 8 and 9)
b) Model skill (Fig 3-7)
c) Comparison of characteristics between the radar-QPE and the model (Fig. 10-12)

Technical corrections:

1. Line 10: add ‘spatio-temporal’ before ‘patterns’, and elaborate on what you mean by ‘effects’.

Accepted. We will change this sentence to “Spatiotemporal rainfall patterns govern the hydrological, geomorphological and societal effects of HPEs”.

2. Line 78: replace ‘getting a’ by ‘receiving’.

Accepted.

3. Line 101: Add the coordinates of Ben-Gurion airport.

Agreed. We will add “31.998N, 34.908E”.

4. Line 130: replace ‘Other’ by ‘Additional’.

OK.

5. Section 3.1: add more details about the radar such as: wavelength (The authors mentioned about the C band), radar parameters (reflectivity, doppler, etc). What is the maximum range of the radar observations? We see it very clearly in Fig. 1b, but number will further clarify.

We will add the radar wavelength (5.35 cm), its range (185 km), the fact that it is a non-doppler radar, and that raw radar reflectivity data were translated to QPE using first a fixed Z-R relationship, $Z = 316 \cdot R^{1.5}$, and then into QPE by applying physically based corrections and gauge-based adjustment procedures (see details in Marra and Morin(2015)).

6. Fig. 2: Are the white areas on the eastern side of the circle domain masked out according to the black line in Fig. 3c? If so, this should be mentioned.

Yes. OK, we would add this to the figure caption.

7. Fig. 3: There is no legend of (a), (b), (c) and (d) as mentioned in the caption and text.

Thanks for noting this. We will add the legend to the figure.

8. Figs. 3,4,5: a normalized difference (e.g., (WRF-radar)/radar) would make more sense than a ratio WRF/radar, such that the red areas will not distract the attention from more important biases.

Accepted. We will change our definition for the bias to be a normalized difference (i.e. ((WRF-radar)/radar)).

9. Line 210: Section 3.5.3 please add a sentence to motivate the use of the DAD curve.

We agree that a motivation is needed. Therefore, section 3.5.3 starts with a motivation sentence (“Areal rainfall amounts are crucial drivers of the hydrologic response and are important for understanding rainfall structure and triggering mechanisms”). We think that this text is enough for the aims of this manuscript.

10. Fig. 4e: replace the scatter plot by a density plot, to see the details inside the black area.

Accepted. We will replace it.

11. Fig. 5: add the equivalent Fig. 4d-f to this case.

Panels d-f were added to Fig. 4 to show that even in a well-simulated event, there is a large disagreement on a pixel scale, while the general characteristics (considering all pixels, as in the histogram, or a large neighbourhood, as in the FSS analysis) could still be well simulated. Fig. 5, however, presents a poorly-simulated event and it is not much successful no matter in what perspective we examine it. This is why we did not present further analyses of rainfall patterns. We also think that due to the clarification made in response to the reviewer's 2nd comment this point is now clearer.

12. Line 165: the synoptic classification is based on semi-objective classification by Alpert (2004). This classification is based on parameters such as T, P, U and V at 1000 hPa once per day based on NCEP-NCAR reanalysis with coarse resolution ãA˘U2.5ã ˝ A˘Uˆ ˚ ◦ . The model (WRF) was analyzed with six hourly ERA-Interim reanalysis with 80km horizontal resolution. It is worth mentioning this.

Agreed. We will mention it.

13. Line 266: greater than 99% of pixels': do you mean to write 'corresponding to less than 1% of the pixels in this HPE'?

Yes. This was our intention, and we will edit the text accordingly.

14. Line 278: which bias to you refer to in the square brackets?

It is the bias of the median (inter-event) amplitude score (from the SAL analysis). However, we will explain this better in the revised text.

15. Line 336: missing 'a' after 'are'.

Correct. Thank you for noticing.

16. Line 437: remove 'a' before 'catchments'.

Accepted.

17. Line 490: replace 'weather' by 'numerical weather prediction'.

Accepted.

18. Fig. 11a-f: make the green and blue colors more distinguishable.

OK.

19. Fig. 12: would ‘temporal lag’ be more suitable than ‘temporal distance’?

It is suitable, however we tried to follow the term used in Marra and Morin (2018) that refer to time-distance.

References:

Marra, F. and Morin, E.: Use of radar QPE for the derivation of Intensity–Duration–Frequency curves in a range of climatic regimes, J. Hydrol., 531, 427–440, doi:10.1016/j.jhydrol.2015.08.064, 2015.

Marra, F. and Morin, E.: Autocorrelation structure of convective rainfall in semiarid-arid climate derived from high-resolution X-Band radar estimates, Atmos. Res., 200(September 2017), 126–138, doi:10.1016/j.atmosres.2017.09.020, 2018.