Editor decision

Journal: HESS
Title: Radar-based characterisation of heavy precipitation in the eastern Mediterranean and its representation in a convection-permitting model

Dear authors,

Thanks for submitting the revised version of your manuscript. It addresses most reviewer comments adequately and the quality of English writing and grammar has much improved.
At this point, I have a couple of minor comments that I would like you to address, mainly for clarification of your statements:

Terminology:
- Duration: The way you use the term duration creates confusion (as also mentioned by a reviewer) and even with the added explanation I'm not entirely sure whether you duration refers to "duration above threshold added up over an event" or "aggregation level", i.e. observation averaged over time windows of 1 to 72 hours.
  Please provide a definition to avoid any confusion.
  The term is also used in the abstract, so please make sure its meaning can be unequivocally understood here, too.
- Characterisation: Two of the reviewers mentioned they see the paper as a model evaluation study, with accompanying climatology characterization.
  I agree with the reviewers, in that the paper does not "systematically characterise high-resolution rainfall patterns" as suggested in the Introduction. Rather, it provides a general characterization based on DAD curves and autocorrelation structure, summarized across all events.
  Please clarify this in the Introduction and Abstract.

Results section:
- In 4.3 you discuss FSS scores (p11, l 315-319): from figure 7f it's clear that the scores for 100mm and 125mm are very unstable due to a limited number of observations in these classes. No conclusions can be drawn from such unstable scores – please rephrase.
- In 4.3. p11, l 326 you mention estimation of "minimum scales for skillful forecasts for various cumulative rainfall depths". FSS is a binary score, that evaluates rainfall detection, not rainfall depths.
  Please be more precise in your phrasing and make sure to use terms like "detection" or "occurrence" instead of "rainfall depth" when referring to FSS scores (throughout section 4.3!).
- The discussion of SAL scores is quite limited, it seems there is much more room for discussion, especially of the Structure and Location scores.
  On p 11, L334: "The structure component was well modelled in most cases, showing the ability of the WRF to accurately generate precipitation objects". Can you elaborate a bit more for the reader: were objects well represented in terms of position and/or intensity, can you say anything about what explains the performance range (certain types of events that typically perform better/worse)?
  For Location (p11, L338-344): it is stated that locations of precipitation objects are not well represented, while the structure score seemed to suggest they were. Again, a bit more discussion is needed here.
- The ellipticity of the 2d autocorrelation fields are discussed as a mean across all events. While in reality ellipticity would be expected to vary a lot between events (as indicated by

the wide range of ellipticity values). Please justify why you think comparing means across events is useful?

- Discussion of performance differences: throughout section 4 (Results) you tend to attribute difference in performance between radar and WRF always to radar. You main observations at larger distance are subject to range degradation (L278, 296, 380), those at shorter distance are attributed to clutter (L372-373). In L397-398 you mention that "because radar QPE suffers from temporal inconsistencies".
This way of discussing performance differences (basically attributing all deviations to radar) is unbalanced and I'm not convinced it's correct. Unless your radar product is particularly poor, but if that's the case, then what's left of the value of this "unique dataset"?
Please check your performance evaluation in section 4 and make sure to have a more balanced discussion between the radar observations and WRF model results.

Discussion section:
- The first paragraph of the Discussion section has a lot of repetition of what's already reported in section 4. If it's meant to be a summary, it better fits at the end of section 4.
- L445: "hither and tither" should be hither and thither. Still, it's an unusual term in professional manuscripts, so I suggest to rephrase it
- Section 5.2: a discussion on the usefulness of data in relation to flash flood comes out of the blue here and has no connection with anything in earlier parts of the manuscript. It's almost a lost piece of literature review. Consider to either move it up to Section 1 where literature is discussed or remove it entirely.
- Section 5.3: please clarify whether 2nd paragraph (starting L484) still refers to ARST (as in first paragraph) or whether this is about all HPEs?

Conclusions:
- You present a nice conclusion about the minimum scale that could be used for forecasting rainfall depths in relation to threshold intensities. It would be valuable nice to directly add some numbers here that came out for your analyses, to make the conclusion more quantitative.

Figures:
- In Figure 7f a vertical scale seems to be missing, for the FSS scores?
- In Figure 11 vertical axis title seems incorrect : should be max accumulation instead of Rain threshold?