Anonymous Referee #1

General comments and manuscript summary: In the submitted manuscript, the authors use 24 years of historical radar data to identify historical heavy precipitation events (HPEs) in Israel, based on various threshold criteria. These 41 HPEs are then re-simulated using the WRF model at convection-permitting resolution (1 km grid spacing). Following this, the manuscript is primarily focused on evaluating how realistically the WRF model simulates the precipitation of the 41 HPEs, compared with what the radar shows. In addition to that, the radar data are used to identify common characteristics of HPEs in the study region.

The manuscript is primarily a model evaluation study of high-resolution WRF for eastern Mediterranean HPEs, with some accompanying radar-based climatological analysis. From the scientific/technical perspective, everything seems OK. My comments which follow in the next sections are thus of a technical and minor nature, and the main question I need to answer here as a reviewer is if the paper presents sufficiently "novel concepts, ideas, tools, or data" to justify publication in HESS?

We thank reviewer #1 for acknowledging our scientific and technical work. We hope that our answers and revisions, in part proposed by reviewer #1, result in an improved contribution that justify publication in HESS. The reviewer is highly appreciated for the time and efforts dedicated for improving our manuscript. The additional references suggested by the reviewer will (a) complete the literature review, and (b) further emphasise the advances we made relative to the existing literature. In the revised manuscript we will address the issues raised by the reviewer as detailed below.

The comments made by reviewer #1 helped us understand that we did not emphasise enough the uniqueness of the high-resolution characterisation itself, and we therefore intend to explain it better in the revised manuscript. Long, high-resolution rainfall data records (24 yr) are truly scarce, and we therefore think that this characterisation is interesting even on its own. Currently, the characterisation is detailed in section 4.2. To validate the model, each one of the pattern-related parameters we have characterised was also checked using model simulations of the same events.

This manuscript is certainly not the first to evaluate if "the model description of rainfall during HPEs" in a convection-permitting model (CPM) is "credible", despite the claims of the authors (L62). There is even a study investigating just that with WRF in the eastern Mediterranean (Zittis et al., 2017), which surprisingly wasn't cited. For other studies asking similar questions in other regions see, for example, Berthou et al. (2018), Brisson et al. (2018), Chan et al. (2014), Chen et al. (2001), Hally et al. (2014), Kendon et al. (2012), Lean et al. (2008); many more CPM evaluation studies can be found – both event-based and climatological. This manuscript represents another contribution to this important topic. I think the publication of the manuscript can be justified on the following grounds: (1) the authors' event-based approach incorporates an unusually high number of events, which is different to the most common approaches of either continuous multi-year simulations (e.g. Ban et al., 2014) or just a handful of events (e.g. Coppola et al., 2018); (2) the authors incorporate a nice range of temporal and spatial diagnostics which are (to my knowledge) not prevalent in the extant CPM-evaluation literature, presumably because of the rarity of such long radar archives (24 years) with high spatiotemporal resolution as used by the authors; (3) CPM evaluation studies for this region of the world are not well represented in the literature.

We much appreciate the reviewer's view about our contribution. It is true that we are not the first to answer this question ("Is the model description of rainfall during HPEs credible?" [Line 62]), and we have referred in the text to many of the previous studies of the topic, but, to the best of our knowledge, we are the first to systematically do it for all the available HPEs over a 24-year period. Furthermore, there is still much to contribute in this area (in our case, we address specifically **rainfall space-time patterns** during all the available **HPEs** during a period of 24 yr). We do understand that the wording we chose may be misleading, and we would change it in the revised manuscript, so it will not be read as if we are claiming to be the first to answer this question. We actually did not know the paper you have mentioned (Zittis et al., 2017), and we are glad that you have referred us to it, since it presents a much needed conclusion both about the WRF performance during extreme rainfall events in the eastern Mediterranean and about the need for good precipitation data, even if based on a more limited number (5) of HPEs. Thus, we will refer to this paper in the revised manuscript.

Specific comments:

1. Structure of results. I wonder would the authors consider that it might make more sense to present some of the results from the characterization of rainfall patterns section (S4.2) at the start of the results section, i.e. before model biases are presented? For example, Section 4.2.1 is based on observations rather than model evaluation. It would seem more logical to me to first present the characteristics of the observed HPEs to readers and then examine if these characteristics are reproduced by the model. Indeed, in your abstract (L13-15) you present the manuscript contents in this order. However, this is for the authors to decide!

We understand the reviewer's point and we thought quite a lot on the best order of steps – first HPEs characteristics from radar and then model skill to reproduce those characteristics (as suggested by the reviewer) or first model skill and then HPEs characteristics as manifested in observations (radar) and model. Our tendency towards the latter approach is due to our understanding that radar observations are not perfect and have their own limitations. Therefore, we prefer to present HPE characteristics from the two sources and to emphasise both agreements and disagreements between them. This comparison follows model skill assessment. We do however agree that some of the HPEs characterisation can be moved to the first part of the results section, specifically those that are not relying on pattern analysis, i.e., seasonality and relation between HPEs at different durations (presently shown in Fig. 8 and 9). Therefore, we will make some changes in the structure of the results section: starting with general properties of HPEs, then model skill, following by space-time HPEs characteristics detected from observations and model. Accordingly, we will make small modifications in the abstract and the introduction sections.

E.g., in the introduction (line 71) we will replace the word “Then” by “Although our observations are based on radar data, these are certainly not perfect. Therefore, we quantified several…”

2. Title. It is not really apparent from the title of the manuscript that this is primarily a model evaluation study. I expect your results will be of most interest to readers concerned with the quality of CPM simulations, however I fear that due to the title the manuscript might be overlooked by readers searching for such information and not reach the full audience it deserves. If it was my manuscript, I’d go for a title along the lines of “Heavy precipitation in the eastern Mediterranean and its representation in a convection-permitting model”. This is, of course, for the authors to decide!

We agree, but we do want to keep the “characterisation” part, from the reasons stated above. We suggest to change the title to:
“Radar-based characterisation of heavy precipitation in the eastern Mediterranean and its representation in a convection-permitting model”.

3. Poorly simulated events. Of the 41 HPEs, you identify two which are simulated particularly poorly and observe that these were characterised by short storm durations (L256-257) and were highly localized (L500-501). You also suggest that the poor simulation may be due to a poorly represented moisture field in the ERA-Interim lateral boundary conditions (L466-467). Have you checked this (if possible)? It would be interesting to know if there was any trace of these precipitation events in (i) the ERA-Interim precipitation fields, or (ii) the coarser resolution WRF domains. If the boundary and initial conditions are inadequate, then there is of course no chance for WRF to well reproduce the event. But this doesn’t mean that WRF itself is deficient or is incapable of simulating such events! Maybe WRF could simulate the event using data assimilation techniques beyond the scope of this experiment, or with better boundary conditions.

We agree with the suggestion. We will show (in the supplementary materials) the results of the coarsest WRF domain. This could possibly give an idea of both the WRF simulated rain fields and of the ERA-Interim input. To have an impression of it, we attach below a preliminary analysis of the rainfall for the first of these two events (event #5; 31/3/93-2/4/93), based on the WRF coarsest domain, to be compared with Fig 5. In contrast to most of the simulated HPEs, in which rainfall was simulated quite well in the innermost WRF domain, this event had almost no rainfall simulated in the inner domain. As the figure below shows, rainfall was not produced by the WRF coarsest domain over the area where it was observed (Fig 5), but rather a few hundred km from there – suggesting that the initial conditions were insufficient to produce rainfall in vicinity of the observed one, regardless to the spatial error of the small-scale (innermost) domain. As the reviewer states, it might have been better simulated using data assimilation, or any other better boundary conditions. However, both are beyond the scope of our manuscript.

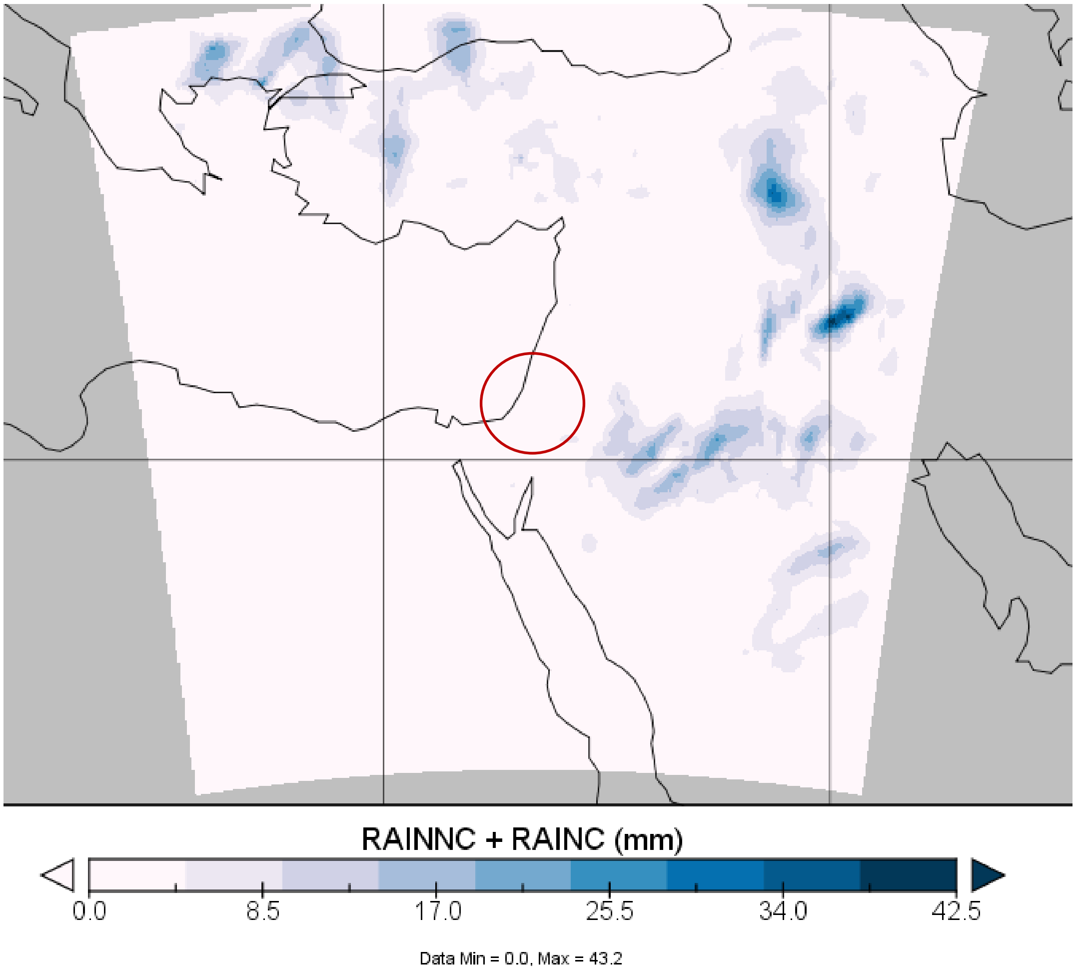

Figure: Rainfall in the coarsest WRF domain during HPE #5 (Table S1) and the approximate range of the Shacham radar (Figure 1).

4. Expectations of CPMs.

My final substantive point is about what we should expect from convection-permitting models, i.e. should we expect them to match radar on a pixel-by-pixel basis? And if they can't do this, does it represent a poor simulation? This is discussed in the introduction of Roberts (2008), where it is argued that the main added value of higher-resolution precipitation forecasts should be seen in area averages – e.g. over a catchment – rather than at specific point locations. I think it's also important to remember that the observed event is also just one possible realisation of the event and WRF will never have perfect initial conditions. You correctly (L469-473) advocate the utility of ensemble simulations for HPEs in the discussion, i.e. as a means of characterizing uncertainty. Similar information to the aforementioned could potentially additionally be presented in the introduction or during the results, as the authors see fit.

The point raised by the reviewer is a crucial one that we want to stress out in the manuscript, and it is actually one of the main points we examine in this manuscript. This is the reason we utilise neighbourhood-based rainfall pattern measures (SAL, FSS), rather than pixel-based indices of success (Fig. 4d, 4f, 6, 7 versus Fig 4e). Moreover, when we compare rainfall patterns, we consider the centre-of-mass of precipitation, Depth-Area-Duration (DAD) curves, and spatial and temporal autocorrelation curves, all of which are not based on point observations. We will better stress this aspect in the revised manuscript. Specifically, we plan to add to the discussion (line 465) the following: "The main added value of convection-permitting models is seen in area averages, rather than over small scale regions (Roberts, 2008). Therefore, over large catchments (e.g., larger than a few hundreds of square kilometres, as suggested by the minimal scale presented in Fig. 6) their forecasts should be relatively useful and accurate. Nonetheless, the use of such a deterministic model is still unsatisfactory in…".

5. Data availability. I think that Section 8 about data availability is inadequate. If someone wants to reproduce your results, a bit more than the two non-specific web domains (L517-518) is needed. Is there a specific web page or ftp server where the radar and rain gauge data can be downloaded? If so, please provide the links. If not, then provide more information about how the data can be found. Additionally, what about the WRF model simulations? Will (have) you upload(ed) them to an openaccess server? If so, provide the download link. Or are they available by contacting the corresponding author? Finally, I suggest uploading the WRF namelist.input as an asset when you are resubmitting the manuscript.

We agree with this comment, however not all of the data are owned by us or can be publicly accessed. We suggest to add to the revised version of the manuscript the specific domain from which one can download the rain gauge data (https://ims.data.gov.il/). These data are not ours to give, however it is available through this data archive (unfortunately, only in Hebrew). The radar data are also not ours to give. It was provided to us by "EMS-Mekorot projects". However, if needed, corrected and gauge-adjusted data (previously published in (Marra and Morin, 2015)) could be given, in the form of images, through a personal communication with the head of the Hydrometeorology lab in the Hebrew University of Jerusalem, Prof. Efrat Morin (efrat.morin@mail.huji.ac.il).

The size of the simulation results is really big (~4.6 TB), so we prefer not to upload those results to the web. We accept your suggestion, and we will add the WRF namelist.input file to the supplementary materials. Using the namelist and the ERA-Interim input files, one will be able to fully reproduce our results.

6. Proof reading. There are a large number of minor grammar errors throughout the text, which are too numerous to list. I therefore suggest a thorough proof reading prior to resubmission.

Accepted. We will proof read the manuscript thoroughly, once we revise it.

Minor and technical comments:

- Section 3.2. Could you please also state (i) the number of vertical levels and height of the model top, (ii) if shallow convection is parametrized in the inner nest, (iii) the interpolation method used, i.e. bilinear, nearest-neighbour, conservative, etc. (i) and (ii) could also be added to table 1, if appropriate.

(i) The number of vertical levels is 68, as stated in Table 1 and the top of the model is at 25 hPa., and we will state this information in the revised version of the manuscript (ii) We use the WRF Tiedtke scheme in the two outer domains (as stated in Table 1) that has a shallow cumulus component, as detailed in (Tiedtke, 1989; Zhang et al., 2011). We will detail this part in the text, as it seems not to be clear from Table 1 only. (iii) The interpolation method used is simply nearest-neighbour, and we will write it in the revised version of the manuscript. Moreover, as suggested, we intend to add the WRF namelist files, so all of the details of our simulations will be clearer.

- Figure 1. It looks like the domain boundaries have been drawn by simply finding the domain corners and drawing straight lines between them. The lower/upper boundaries of Lambert conformal domains shouldn't have constant latitudes. I think you need to extract the outermost rows/columns from WRF's XLONG and XLAT arrays and use these to plot your domain boundaries.

That's correct. The domains are not plotted with their exact extent. We will correct this in the revised version of the manuscript.

- Figure 2. I wonder would it make more sense to compute the %-bias? i.e. instead of bias = WRF/Radar, use bias = 100.*(WRF – Radar)/Radar. With the current formulation the dry biases are lower bounded whereas the wet biases are not upper bounded. With %-bias this would not be the case. I suppose it's not really that big of a deal. The authors can decide for themselves.

This was also mentioned by the other reviewers. We will change the bias definition into normalised difference (i.e. (WRF-radar)/radar).

- Figure 2. Please add "a, b, c, d" labels to the panel plots, to match the text.

We accept this correction, and we will apply it in the revised version of the manuscript.

- L123: Note that it should be possible in WRF to just save precipitation at 10-minute intervals and other variables at a lower frequency, to reduce storage space.

That's correct. Still, after doing this (we actually saved few 2D variables, and not only the precipitation field, however we did not save 3D fields every 10-min), because of the high resolution, the results weigh on average ~112 GB per event.

- L128: I think the reference to "Sect 3.2" is wrong.

Right. This should be corrected to "Sect 3.3" and we intend to add a reference to Table S1 as well.

- L170: The abbreviation "TP" isn't defined anywhere

Correct. We will change this abbreviation to the full synoptic class name (i.e. "Tropical Plume").

- L396-398: It may prove difficult to identify which days to downscale from the GCMs, especially for convective events. There are some papers recently suggesting methods for identifying the best days to downscale (Chan et al., 2018; Meredith et al., 2018; Gómez-Navarro et al., 2019).

This is true, and we can refer to these papers if requested. However, recent studies also tried running long-term downscaled simulations (e.g., Kendon et al., 2014; Liu et al., 2017).

References from reviewer #1:

Ban, N., Schmidli, J., Schär, C. (2014). Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. Journal of Geophysical Research: Atmospheres, 119, 7889–7907.

Berthou, S., Kendon, E. J., Chan, S. C., Ban, N., Leutwyler, D., Schär, C., Fosser, G. (2018). Pan-european climate at convection-permitting scale: A model intercomparison study. Climate Dynamics, 1–25.

Brisson, E., Brendel, C., Herzog, S., Ahrens, B. (2018). Lagrangian evaluation of convective shower characteristics in a convection-permitting model. Meteorologische Zeitschrift, 59-66.

Chan, S., Kendon, E., Fowler, H., Blenkinsop, S, Roberts, N. (2014). Projected increases in summer and winter UK sub-daily precipitation extremes from high-resolution regional climate models. Environmental Research Letters, 9(8), 084019.

Chan, S. C., Kendon, E. J., Roberts, N., Blenkinsop, S., Fowler, H. J. (2018). Largescale predictors for extreme hourly precipitation events in convection-permitting climate simulations. Journal of Climate, 31(6), 2115-2131.

Chen, F., Warner, T. T., Manning, K. (2001). Sensitivity of orographic moist convection to landscape variability: a study of the Buffalo Creek, Colorado, flash flood case of 1996. Journal of the Atmospheric Sciences, 58(21), 3204-3223.

Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., ... Caldas-Alvarez, A. (2018). A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean. Climate Dynamics, 1-32.

Gómez-Navarro, J. J., Raible, C. C., García-Valero, J. A., Messmer, M., Montávez, J. P., Martius, O. (2019). Event selection for dynamical downscaling: a neural network approach for physically-constrained precipitation events. Climate Dynamics, 1-17.

Meredith, E. P., Rust, H. W., Ulbrich, U. (2018). A classification algorithm for selective dynamical downscaling of precipitation extremes. Hydrology and Earth System Sciences, 22(8), 4183-4200.

Kendon, E. J., Roberts, N. M., Senior, C. A., Roberts, M. J. (2012). Realism of rainfall in a very high-resolution regional climate model. Journal of Climate, 25(17), 5791– 5806.

Lean, H. W., Clark, P. A., Dixon, M., Roberts, N. M., Fitch, A., Forbes, R., Halliwell, C. (2008). Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. Monthly Weather Review, 136(9), 3408-3424.

Roberts, N. (2008). Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 15(1), 163-169.

Zittis, G., Bruggeman, A., Camera, C., Hadjinicolaou, P., Lelieveld, J. (2017). The added value of convection permitting simulations of extreme precipitation events over the eastern Mediterranean. Atmospheric research, 191, 20-33.

References:

Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C. and Senior, C. A.: Heavier summer downpours with climate change revealed by weather forecast resolution model, Nat. Clim. Chang., 4(7), 570–576, doi:10.1038/nclimate2258, 2014.

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., Chen, F., Chen, L., Clark, M., Dai, A., Dudhia, J., Eidhammer, T., Gochis, D., Gutmann, E., Kurkute, S., Li, Y., Thompson, G. and Yates, D.: Continental-scale convection-permitting modeling of the current and future climate of North America, Clim. Dyn., 49(1–2), 71–95, doi:10.1007/s00382-016-3327-9, 2017.

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.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Mon. Weather Rev., 117(8), 1779–1800, doi:10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Zhang, C., Wang, Y. and Hamilton, K.: Improved Representation of Boundary Layer Clouds over the Southeast Pacific in ARW-WRF Using a Modified Tiedtke Cumulus Parameterization Scheme, Mon. Weather Rev., 139(11), 3489–3513, doi:10.1175/MWR-D-10-05091.1, 2011.

Zittis, G., Bruggeman, A., Camera, C., Hadjinicolaou, P. and Lelieveld, J.: The added value of convection permitting simulations of extreme precipitation events over the eastern Mediterranean, Atmos. Res., 191, 20–33, doi:10.1016/j.atmosres.2017.03.002, 2017.