Synopsis:
The study by Lin et al. investigates present and future heat wave magnitudes over Europe with RCMs of the Euro-Cordex ensemble. It is found that for present climate conditions, the RCMs are able to capture most of the observed spatial and temporal features of heat wave magnitudes. A central finding in my view is that the uncertainty in a simulated future change of heat wave magnitudes can be attributed almost equally to the difference in model physics and to the driving data. In general I am positive about the work. Still, I raise some issues below which concern the processes behind the observed signals and the choice of evaluation metrics. Further, there are numerous typos in the manuscript which I do not list here. These should be corrected prior to resubmission.

Major:
1) Processes behind climate change signals: I very much appreciate that the authors investigate the link of changes in HWMId to processes of the climate system. However, an important component is missing in my view which is the representation of blocking. It is well known that blocking is an important driver of heat waves in Europe and that the representation of blocking over Europe represents an important challenge to climate models (e.g., Masato et al. 2013). To complete the analysis, I therefore suggest either quantifying blocking biases or at least putting the results of this work in the context of studies that have already investigated blocking in CMIP5 models. The availability of open source code should simplify such an analysis (e.g. https://github.com/jlpscampos/Blocking_Index2d).

2) I understand that the models are evaluated in terms of the MBE, RMSE and Pearson correlation coefficient. However, are the HWMId values normally distributed so that the central limit theorem is valid? Accordingly, I suggest to also evaluate the statistical distributions of the models.
l. 7: The conclusion that "RCMs can reveal spatial features of HWMId associated with small-scale processes" is place quite prominently in the abstract. However, after reading the manuscript it is neither entirely clear to which spatial features the authors are referring to, nor have the small-scale processes been named explicitly. Also what is meant with small-scale process? This is rather colloquial terminology and it would be good to specify explicitly. Are you referring to processes on the meso-scale or processes that are parametrized. Please clarify.

l. 10: The term "expontential-like" is used frequently? However, I am not sure what "expotential-like" means. Either a rise is exponential or it is not. Therefore, I suggest to rethink the terminology used here.

l. 24: I guess you mean "marine heat waves" instead of "maritime heat waves".

l. 26: Is it on purpose to using the terminology "warm spell" here?

l. 48: Perhaps write explicitly "differently then GCMs".

l. 52: "have" instead of "hace".

l. 80: delete "is calculated" at the end of sentence.

l. 84: What is the reasoning for only using a subset? Please explain.

l. 88: Please give reasons why you selected RCP8.5.

l. 98: Do you mean "to ERA-Interim" instead of "the ERA-Interim"?

l. 109: Please clarify this sentence/make it more concise. Also, please specify the processes you are thinking of. To the reader it may not be immediately clear which processes you are referring to.

l. 114: You may want to delete "the following questions" since i) and ii) are not actually formulated as questions.
I. 122: The way the results are presented it is difficult to directly see the differences "of some local features". Therefore, I suggest to revise the figures by showing differences between the RCM results and E-Obs (or ERA-Interim). This would help the reader to spot differences directly. Also, it would be good to name the "local features" explicitly. Are these local features related to land-sea contrasts, the direct effect of topography on temperature, etc?

I. 124: Delete "only" before "focusing".

I. 128: What is the area of the spatial average? Please provide this information somewhere in the text.

I. 161: As for I. 122: In my view, it would easier to follow the discussion if differences between RCM/GCM results and E-Obs were shown. Also, please mention explicitly where the GCMs "miss out both in detailed structure and amplitude". In its current form it is left to the reader to spot these deficiencies.

I. 182: "of" instead of "in".

I. 187: Please split this sentence in two sentences as it is hard to follow the explanation. Also, there is a word missing between "but" and "exists".

I. 200: Please provide a "not shown" as the reader may otherwise try to find the spread information in a Figure.

I. 279: What exactly is meant with the quite general term "orographic effects"? How are they differently represented in the RCMs? Some explanation would be helpful here.

I. 295: This sentence is overly complicated in my view (2x "is expected"). Please revise this sentence.

I. 335: It would be good to mention explicitly the "room for improvement". The interpretation should not be left to the reader.
I. 344: Do you have any insights on what is leading to the weaker drying trend in the RCMs?

I. 349: "Plays" instead of "play".

I. 354: Also here, what is meant with the term "orographic effects"? Are you thinking of the treatment of sub-grid orography? Please specify?

References: