Synopsis:

Comment 1: 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.

Reply: Thanks for the positive evaluation and also the issues raised.

Major:

Comment 2: 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).

Reply: We agree that blocking is an important driver of heat waves. We also appreciate very much the information about the related paper and source code along with the suggestions. Indeed, we are planning a future work going into details about the climate change signals of blocking in GCM-RCM chains and their link with heat waves. An issue that we currently have not a mature solution is that blocking index can only, as latitudinal bands are commonly adopted for existing 2D blocking indicators, be derived within a part of the European domain defined for EURO-CORDEX simulations. Within this subsection of Discussion, we focused on understanding the factors that contribute to the spatial pattern (and the difference between models) of heat wave magnitudes. As such, we have accepted the suggestion that "putting the results of this work in the context of studies that have already investigated blocking in CMIP5 models" and have the following text in the revised version:

> *According to Masato et al. (2013), the three CMIP5 GCMs assessed show a decrease in summertime North Atlantic blockings and an increase in blockings over eastern Europe or Russia indicating an eastward shift of the blocking activity. This implies that the underlying processes of ΔHWMId is possibly beyond the atmospheric dynamics. It is of interest whether, and if so how, the RCMs modify the climate change signals of atmospheric blocking from their driving GCMs, and whether this modification is related to the differences in ΔHWMId patterns between GCMs and RCMs as presented here, which is worth further study, though representing atmospheric blocking is considered a challenge (e.g., Masato et al., 2013; Davini and D'Andrea, 2016; Jury et al., 2019).*

Comment 3: 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.

Reply: This is an interesting comment, which was also mentioned by Reviewer#1. Here, we would like to raise the following two points:

a) Are yearly HWMId values normally distributed? The answer is no. Actually, the distribution of the HWMId values is already reflected in some figures in the manuscript (Fig. 1b and Fig. 9 for spatial; Fig. 2b and Fig. 10 for temporal), which is somehow skewed.

b) Is the central limit theorem valid for the evaluation in terms of the MBE, RMSE and Pearson correlation coefficient? The answer is yes. In probability theory, the central limit theorem (CLT) establishes that, in many situations, when independent random variables are summed up, their properly normalized sum tends toward a normal distribution **even if the original variables themselves are not normally distributed**.

Comment 4: 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 explictly. Are you referring to processes on the meso-scale or processes that are parametrized. Please clarify.

Reply: We accept that the terminology is colloquial. We initially intended to address the local details revealed by RCMs due to higher resolution than GCMs, which we thought to be associated with some small-scale processes. However, we did not (and actually, by looking at the spatial patterns, we doubt we are able to) have an idea about what exactly the underlying processes are. Our special observation is that some local spatial patterns are corresponding to mountainous areas, according to which we consider orographic effects (neither of meso-scale nor of parametrized, but more likely connected to more accurate representation of elevations and/or associated mountainous circulations) to be a plausible underlying process. Thus, in the revised version we have added “(e.g., orographic effects)”.

Comment 5: 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.

Reply: We have used “approximately exponential” instead.

Comment 6: l. 24: I guess you mean “marine heat waves” instead of “maritime heat waves”.

Reply: Corrected

Comment 7: l. 26: Is it on purpose to using the terminology “warm spell” here?

Reply: Changed to “heat wave”.

Comment 8: l. 48: Perhaps write explicitly "differently then GCMs".

Reply: “differently from GCMs” has been added.

Comment 9: l. 52: "have" instead of "hace".

Reply: Corrected.

Comment 10: l. 80: delete "is calculated" at the end of sentence.

Reply: Removed.

Comment 11: l. 84: What is the reasoning for only using a subset? Please explain.

Reply: The corresponding text has been rephrased as (L90–93):

> *The large number of GCM-RCM combinations available from EURO-CORDEX allows us to examine the behaviors (e.g., signal modification and uncertainty transformation) within the downscaling process for simulating heat wave magnitudes. For that we used only a subset of the available ensemble to gain a full GCM-RCM matrix without gaps, to ensure a fair comparison after aggregating along either the GCM dimension or the RCM dimension.*

Comment 12: l. 88: Please give reasons why you selected RCP8.5.

Reply: The corresponding text has been revised as (L96–97):

> *Here, we only focus on RCP8.5 because there are fewer EURO-CORDEX simulations forced by other RCPs, hardly available for formulating such a full simulation matrix.*

Comment 13: l. 98: Do you mean "to ERA-Interim" instead of "the ERA-Interim"?

Reply: We used “to ERA-Interim” instead.

Comment 14: 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.

Reply: The corresponding paragraph has been rephrased as (L119–122):

> *In order to better understand the underlying processes of the simulated climate change signals in HWMId, a preliminary effort was made by exploring the simulated climate change signals in dry conditions and their connection with HWMId. We investigated atmospheric and soil dry conditions, as represented by dry days (with precipitation < 1 mm) and precipitation $-$ evaporation ($P - E$), respectively. As a background of warming, the simulated climate change signals in annual mean T_{max} were also examined.*

Comment 15: l. 114: You may want to delete "the following questions" since i) and ii) are not actually formulated as questions.

Reply: The corresponding text has been rephrased as (L125–127):

> *Among various aspects of heat waves, we want to answer the following questions: i) Does the spatial pattern of the climatological mean reveal observed local information? and ii) Does the regional mean show the same signal/response of large-scale climate variations as observations?*

Comment 16: l. 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?

Reply: Here, we failed to state clearly that the differences referred to are between the RCMs rather than between the RCM results and E-OBS (i.e., bias). We intended to show the spatial pattern of original HWMId values rather than their biases, because we thought bias alone may not sufficiently reveal the performance of a model (and therefore we used also spatial RMSE and r hereafter), although it helps the reader to spot differences directly. Moreover, we also intended to show the spatial pattern on original model grids in order to avoid any loss of information (including so-called “local features”) due to interpolation. The term “local features” used here simply means opposite to those general (large-scale) patterns such as the west-east gradient, and could be related to any known or unknown processes which we were however not going to discuss at this point. To help the reader understand what we are talking about “local features”, we added examples. The corresponding text has been rephrased as (L132–134):

> *Though all RCM simulations agree on the general spatial pattern, they do differ from each other in representing details of some local features. For example, among the four RCMs, RCA4 has particularly high values over central Europe, and REMO2015 is the only RCM capture the observed level of HWMId values (7–9) over eastern Europe.*

Comment 17: l. 124: Delete "only" before "focusing".

Reply: We have followed the suggestion.

Comment 18: l. 128: What is the area of the spatial average? Please provide this information somewhere in the text.

Reply: The following information was added in Section Data and methods (L111):

> *The area of investigation is bounded by 10°W–30°E and 35°N–70°N and is land-only.*

Comment 19: l. 161: As for l. 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.

Reply: Regarding the reason not showing bias w.r.t. E-OBS, please see our response to Comment 16. The following text explains where the GCMs "miss out both in detailed structure and amplitude" (L170–176):

> *The observed west-east gradient is hardly seen in the GCM simulations: EC-EARTH and HadGEM2-ES report no difference between the western and eastern parts (divided by the white line on the map), while NorESM1-M shows a west-east gradient of 1.3, which is even higher than for E-OBS but simulates excessively high HWMId values in the easternmost part of the domain. The downscaling with RCMs improves the representation of the observed spatial pattern by reproducing the west-east gradient, as well as revealing small-scale processes such as orographic effects which are in no way represented by the GCMs due to their coarse resolution.*

Comment 20: l. 182: "of" instead of "in".

Reply: Corrected

Comment 21: l. 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".

Reply: The sentence has been rephrased as (L201–203):

> *Aggregating on the GCM dimension (i.e., calculating mean and spread for each RCM with different GCMs), the ensemble means (first row of Fig. 5) reveal a similar spatial pattern, whereas the spreads (second row of Fig. 5) show considerable differences in the spatial pattern.*

Comment 22: l. 200: Please provide a "not shown" as the reader may otherwise try to find the spread information in a Figure.

Reply: The related text has been revised as (L214–217):

> *Similar to the results of the recent past climate (Sect. 3.2), a large spread of ΔHWMId is seen across the GCMs (5.8–12.8 and 21.6–48.4 for the nearest decades and the end of the century, respectively; Fig. 6). The RCMs decrease the spread (4.9–8.7 and 15.0–31.9 for the nearest decades and the end of the century, respectively; Fig. 6), reflecting less change compared to the driving HadGEM2-ES and NorESM1-M.*

Comment 23: l. 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.

Reply: We used the term "orographic effects" as we observed some orographic traces upon HWMId spatial patterns (especially for mountainous areas) and hence we thought this to be related to orographic-related processes. Considering RCMs have the same spatial grids, we thought the different representation of "orographic effects" comes possibly from the dynamical/thermodynamical interactions with parameterizations. The text referred to has been revised as (L301–306):

> *An interesting finding of this study is that some spatial features of HWMId showing orographic traces, thereby considered to be related to orographic effects, can be seen as one aspect of the added value of RCMs as GCMs cannot represent such features due to too coarse spatial resolution. Nevertheless, the orographic effects are represented differently across the RCMs (i.e., the large ensemble spread along the RCM dimension; Fig. 4), suggesting that the representation of orographic effects, possibly related to dynamical/thermodynamical interaction with parameterizations, is one of the major sources of uncertainty.*

Comment 24: l. 295: This sentence is overly complicated in my view (2x "is expected"). Please revise this sentence.

Reply: The sentence referred to has been revised as (L323–324):

> *As the HWMId is directly calculated from daily T_{max}, the increase in annual T_{max} provides a warming background that the surge in HWMId is associated with.*

Comment 25: l. 335: It would be good to mention explicitly the "room for improvement". The interpretation should not be left to the reader.

Reply: We have removed "but still with room for improvement" because we do not think it is necessary to state this here.

Comment 26: l. 344: Do you have any insights on what is leading to the weaker drying trend in the RCMs?

Reply: No. Here, we would like to leave it as an open question. Note that we have the following paragraph in Discussion (L350–361):

> *As a preliminary effort, concerning only the spatial pattern, the above analysis is however far from building clear causal links. Moreover, we are not yet clear about what is leading to the weaker drying trend in the RCM simulations. Atmospheric blocking, with adiabatic warming of sinking air and anomalous clear-sky radiative forcing, is an important driver of heat waves (e.g., Bieli et al., 2015; Schaller et al., 2018). According to Masato et al. (2013), the three CMIP5 GCMs assessed show a decrease in summertime North Atlantic blockings and an increase in blockings over eastern Europe or Russia indicating an eastward shift of the blocking activity. This implies that the underlying processes of ΔHWMId is possibly beyond the atmospheric dynamics. It is of interest whether, and if so how, the RCMs modify the climate change signals of atmospheric blocking from their driving GCMs, and whether this modification is related to the differences in ΔHWMId patterns between GCMs and RCMs as presented here, which is worth further study, though representing atmospheric blocking is considered a challenge (e.g., Masato et al., 2013; Davini and D'Andrea, 2016; Jury et al., 2019). Regarding the representation of the effect of land-atmosphere interactions, investigation detailed into different sets of parameterizations in GCMs and RCMs as well as some additional sensitivity experiments may be necessary for a better understanding.*

Comment 27: l. 349: "Plays" instead of "play".

Reply: Corrected. Note that the corresponding sentence has been moved to Discussion (L348).

Comment 28: l. 354: Also here, what is meant with the term "orographic effects"? Are you thinking of the treatment of sub-grid orography? Please specify?

Reply: Please see our response to Comment 23.