Response to Reviewers' Comments on egusphere-2022-215

Changgui Lin[1,2], **Erik Kjellström**[1,3], **Renate Anna Irma Wilcke**[1], **and Deliang Chen**[2]

[1]Rossby Centre, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden
[2]Regional Climate Group, Department of Earth Sciences, University of Gothenburg, Gothenburg, Sweden
[3]Department of Meteorology and the Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

Correspondence: Changgui Lin (mapulynn@gmail.com)

We thank the reviewers for their detailed and helpful comments, which we feel will considerably improve the manuscript. Please see our detailed responses below.

1 Reviewer #1

General comments

Comment 1.1: Lin et al. used dynamical downscaling to analyse heatwaves based on simulations carried out with regional climate models (RCMs) from the Euro-CORDEX programme. A particularly relevant topic of this paper is that the authors investigated if there is any added value in the representation of heat waves in the RCMs compared to the driving GCMs. It is an interesting topic definitely worth pursuing.

Reply: Thanks!

Comment 1.2: A general remark is that all researchers discussing evaluation and use of GCM results on regional scales ought to read the paper by Deser et al (2012; DOI:10.1038/nclimate1562), and citing it in a study like this thus should be required. The findings of Deser et al. suggest that the small number of GCMs selected here is insufficient for a proper analysis of future outlooks and model evaluation, due to pronounced chaotic regional variability on decadal scales.

Reply: We agree that a large number of models is required for a proper analysis of future outlooks and model evaluation. It is worth mentioning that we are not trying to make any full-scale assessments of future heat waves in Europe but rather investigate how they differ between GCMs and RCMs by comparing the GCM-RCM chains. Thus, the problem is slightly less problematic. A problem from the downscaling side, though, is the lack of large ensembles of GCM-RCM chains that could be assessed. We conducted the same analyses based on another GCM-RCM combination (with the same size but different GCM/RCM members) and derived similar results, which can alternatively support our conclusions.

In the revised version, we added the following text in an additional subsection of Discussion (L363–368):

> *A full simulation matrix without gaps facilitates a fair comparison after aggregating along either the GCM or the RCM dimension. This requisite limits the size of RCM simulation matrix available for analysis. The limited size of the GCM-RCM simulation matrix could be a shortcoming influencing the robustness of the results in the study, especially for the uncertainty analysis where the uncertainty is described by the spread (maximum – minimum) across three or four ensemble members. In fact, we conducted the same analyses upon another GCM-RCM simulation matrix (with the same size but different GCM/RCM members) and derived similar results (not shown), which can alternatively support the conclusions herein.*

Comment 1.3: The regional climate modelling community also still seems to exhibit a 'silo thinking' behaviour, and in order to try to make som progress in the general thinking about downscaling, I would urge that this paper by Lin et al. also includes work based on empirical-statistical downscaling (ESD). Many papers on RCMs ignore ESD, which becomes invisible and under-appreciated, and this unfortunately seems to create an attitude that RCMs suffice - hence many of the climate services in Europe do not consider ESD. I suspect most people working with RCMs don't read the literature on ESD, but I think there are benefits from consolidating the two approaches - in particular when it comes to the evaluation of RCMs. There are also a few examples of ESD applied to heatwave statistics that merit a mention in the context of this paper (e.g. DOI:10.5194/ascmo-4-37-2018). Nevertheless, ignoring ESD is a weakness, although Lin et al. give a good summary of the limitations of RCMs. RCMs and ESD make use of different sets of assumptions and have different strengths and weaknesses independent of each other, and hence a combination of the two makes the results more robust.

Reply: We appreciate that the reviewer thoughtfully consider the different methods for downscaling, of which the ESD is worth and should be assessed in the context of assessments of climate change. However, this study is not focused on different downscaling approaches but rather the behaviors (e.g., signal modification and uncertainty transformation) within the GCM-RCM chains (see also our reply to Comment 1.2). Thus, we decided not to extend our analysis by including datasets based on ESD. Having said this, we did add the following statement concerning the use of both ESD and dynamical downscaling for deriving regional climate information (L40–44):

> *As a remedy, to improve the quality of the simulated climate and add value compared to GCMs, empirical statistical downscaling employing a wide range of approaches (Benestad et al., 2008, 2018; Hertig et al., 2019;*

Soares et al., 2019) and dynamical downscaling using a regional climate model (RCM) (Torma et al., 2015; Rummukainen, 2016; Strandberg and Lind, 2021) are used, each with its relative strengths. This study focuses on the dynamical downscaling.

Comment 1.4: Often the most severe effects of heatwaves are connected with night-time temperatures not cooling off. It is therefore also of interest to use a heatwave index based on daily minimum temperatures and not the daily maximum. The most pronounced temperature trends also are those of the nights.

Reply: This is an interesting suggestion and is worthwhile pursuing; however, it is out of scope of this paper. To make this point more visible, we added the following text in the revised Discussion section (L369–374):

> *The HWMId can also be applied to other temperature variables, but with different processes/impacts involved. For example, the HWMId applied to daily minimum temperature serve a measurement of heat wave magnitude taking into account also the nighttime cooling effect (Russo et al., 2015). As another example, Apparent Heat Wave Index (AHWI, Russo et al., 2017), the HWMId applied to daily apparent temperature, considers also the impact of air humidity on human beings. Such variants of HWMId are being considered for future studies.*

Comment 1.5: It would be interesting to see the statistical distribution of yearly HWMId values - are they normally distributed? (E.g. is the central limit theorem valid for this statistic aggregated over Europe?) One way to evaluate the models is to compare their statistical distributions (e.g. Kolmogorov-Smirnov Test).

Reply: This is an interesting comment, which was also mentioned by Reviewer#3. We are also grateful for providing the method to test statistical distribution. Below you will find my answers to your questions:

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

b) Is the central limit theorem valid for this statistic aggregated over Europe? 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 1.6: I was a bit surprised by Fig.1 that seems to indicate more heatwave activity in the Nordic countries and less further south on the continent. This also seems to be the case for EOBS and ERAINT - does that mean that perhaps HWMId doesn't represent the typical heatwave reported by the news headlines? It's defined in terms of local variability (IQR) and autocorrelation - and not on any threshold value, as far as I read this paper. At least, this warrants some comments.

Reply: We agree that the HWMId index is somehow complex and assisting materials are needed to fully understand what it represents. The answer to "does HWMId represent the typical heatwave reported by the news headlines?" is yes, for which please refer to Russo et al. (2015, Fig. 2 therein). Moreover, we have added the following text (and associated figures) to the supplementary materials for better understanding the HWMId:

> *The HWMId (Russo et al., 2015) used in the study is a fairly established indicator for classifying heat waves, taking into account both the duration and intensity. The detailed definition of HWMId can be found in Sect. 2.1, as well as Russo et al. (2015). It is worth exploring how it works by presenting examples though. Figures S1 and S2 show the spatial distribution of observed (E-OBS) European HWMId values and the detected heat waves at four selected grid points for Years 2007 and 1989, respectively. Following the definition, heat waves are detected when daily maximum temperature exceeds the daily threshold at least three consecutive days, where the threshold for a given day is 90th percentile of daily maximum temperature within the 31-day window centered at this day and within the reference period (1989–2008 for the examples shown here, same as the results in Sect. 3.1). The HWMId value at each grid point is equal to the maximum red area above the climatological 25th percentile of maximum temperature in Fig. S1b (or Fig. S2b) normalized by the climatological interquartile range (IQR) within the reference period, where the 25th percentile and the IQR are constant at a given grid point. As such, the following points are noted:*
>
> *A) The yearly HWMId values shown on the map may not necessarily represent the spatial distribution of magnitudes of a single heat wave (e.g., occurred in the same period) but the maximum magnitudes of the grid points respectively. For example, Fig. S1b (or Fig. S2b) shows that the heat wave with maximum red areas appeared at different time of the year for the four selected grid points.*

B) The yearly HWMId values are in most cases associated with summertime heat waves due to the higher temperature (Fig. S1b and S2b). However, it does not rule out the case that the heat waves with the maximum magnitudes appear in wintertime; an example is given at grid point "SE" in 1989 (Fig. S2b).

C) The HWMId can certainly identify those outrageous heat waves reported by the news headlines according to Russo et al. (2015, Fig. 2 therein). Here, the heat wave occurred over southern Europe in 2017 summer is clearly visualized in Fig. S1a. However, as an index that is reference period-based, the HWMId might be problematical when quantifying magnitudes of moderate heat waves within the reference period, because a higher threshold can be expected for a location with more heat wave activities. Figure S3 shows that the grid point "SE" compared to points in western Europe has a shorter distance between daily 90th and 95th percentiles (within a 31-day window) of daily maximum temperature within 1989–2008, where the former was used as the threshold for the HWMId within Sect. 3.1, over the long summer season (May–September), indicating very likely less heat wave activities at southeastern Europe. Considering that the yearly HWMId values are in most cases associated with summertime heat waves (i.e., Point B), therefore, the result shown in Fig. S1a explains the spatial pattern of HWMId with values in western coastal areas higher than those in eastern areas (Fig. 1a).

Comment 1.7: Does the result that all RCMs show less agreement with E-OBS in RMSE and r compared to that of ERA-Interim suggest that these RCMs don't add value to that of the global model? Or could it be differences in heat fluxes, cloudiness and topography of the driving and nested models? Perhaps the model domain is so large that the RCMs generate their own dynamics within the interior of their lateral boundaries? Or have they involved spectral nudging to avoid that? See e.g. DOI:10.1007/s00382-022-06219-y (it's also a useful paper to discuss in this context). These questions certainly merit some discussion. The results are nevertheless useful and interesting as they suggest that differences between the RCMs matter.

Reply: The issue of spectral nudging has been discussed vividly in the literature and it is indeed a powerful method for better capturing single events that are governed by the lateral boundary conditions. For the ERA-Interim-driven runs this could have been a meaningful thing to do. A problem in this context is that most RCMs are run without spectral nudging so there is no option of doing such an analysis for practical reasons. Also, as the GCMs do not necessarily represent these events in a realistic way it is interesting to analyse how a free-running RCMs modify the results.

Larger biases w.r.t E-OBS in the RCMs than in ERA-Interim imply that the RCMs do modify the results and that it infers its own biases. At the same time, however, it does not by default mean that it does not add value w.r.t. the GCMs. The better performance shown compared to the GCMs (Table 3 and 4) indicates that they do add value. Related to the size of the domain it is clearly so that the RCMs have some freedom to create their own climate, and that this freedom is generally larger in summer (when we see these heat waves) as the region is less well "ventilated" or "flowed through" by the westerlies that are weaker in these situations (especially, when there are strong high pressure situations - this also speaks against spectral nudging as a solution as it will still be "interior" processes that dominates).

Comment 1.8: I'm not sure that I understand Table 4 and the use of MBE, RMSE and correlation for results derived from GCMs since we don't expect the GCMs to be synonymous with the real world and hence no correlation with observed heatwaves. The only way to evaluate the downscaled results from GCMs is through statistical properties such as statistical distributions and parameters. But perhaps Table 4 shows the correlation in space rather than over time? If so, this ought to be explained more explicitly and clearly. Also if the appearance of the number of heatwaves more or less follows a random process, then we'd expect that it over a given period will follow a Poisson distribution - this can be assumed to be true for both models and the real world. Then the number of observed heatwaves can be compared to a statistical distribution of corresponding number of heatwaves based on the model ensemble by assuming a Poisson distribution (this works if the ensemble is considerably greater than 30 independent runs). Is this possible, or does the HWMId statistic suffice? Also, so-called 'common Empirical Orthogonal Functions' can be used to compare spatial structures and the covariance structures in different data sets - it's an elegant maths-based approach that is surprisingly uncommon. However, this is more general and not specific for a small selection of extreme events. But regarding my comment on Fig 1, I'm a bit unsure what HWMId really represents. Perhaps it also may be of relevance here to mention that one indicator of trends in extremes, including an increasing severity of heatwaves, can also involve an analysis of record-breaking events. There is some literature on this subject connected to climate change.

Reply: Statistics in Table 4 (as well as 2 and 5) are of spatial characteristics, which has been explicitly stated in the revised table caption. Sorry for the opaque manner in the previous version of table caption. We also thanks for the various approaches provided for evaluating the models. Regarding what the HWMId really represents, please see our response to Comment 1.6.

Comment 1.9: The most rapid warming in northern Europe is during winter, but maximum daily temperatures are highest in summer, and it's only summer that defines HWMId? (L348)

Reply: The HWMId analyzed of each year represents the maximum heat wave magnitude throughout the year. Please see our response to Comment 1.6 (especially Point B of supplementary text).

Comment 1.10: The point about 'cascade of uncertainty' is a myth and forgets that each step of analysis also introduces new information (or constraints) in addition to uncertainty. It's only sensible with several model stages as long as we introduce more information than uncertainty for each step (see e.g. DOI:10.1038/NCLIMATE3393). In fact, downscaling can be considered as an act of adding new information to that already provided by GCMs: information about how local geography influences the local climate (as in this case) and information about how local climates depend on the ambient large-scale conditions and teleconnections that the GCMs skillfully reproduce.

Reply: Thanks for providing the insights into RCMs' added value w.r.t. GCMs. Indeed, we have indicated in the manuscript that the concept of "cascade of uncertainty" is questionable by referring to other literature and to results of our study indicating that RCMs add information (and sometimes value). In the revised version, we integrated with more arguments including your insights and have the following text (L283–298):

> *Some studies (e.g., Schiermeier, 2010; Kerr, 2011) show that uncertainty may increase when downscaling GCMs with RCMs as biases from the GCMs are conveyed to the RCMs and RCMs additionally add their own biases, referred to as the 'cascade of uncertainty' (e.g., Wilby and Dessai, 2010). However, many other studies (e.g., Torma et al., 2015; Di Luca et al., 2016; Rummukainen, 2016; Sørland et al., 2018; Strandberg and Lind, 2021) indicate that RCMs can also add value upon the driving GCM simulations. This study demonstrates the added value for heat wave magnitudes that have so far not been studied as extensively as for other aspects of climate and climate change, reflected in adding more detailed geographical patterns and pulling the results closer to the observations (Fig. 3; Table 4 and 5). Such added value confirms the usefulness of RCMs for downscaling coarse-scale GCM simulations, because downscaling can be considered as an act of adding useful information to that already provided by GCMs: information about how local geography influences the local climate (as in this case) and information about how local climates depend on the ambient large-scale conditions and teleconnections that the GCMs skillfully reproduce. Moreover, RCMs may also more realistically represent some atmospheric processes relative to the GCMs. Our analysis of the ensemble spread along the GCM dimension, reflecting uncertainty associated with driving data, reveals that the RCMs alter the spatial HWMId pattern from their driving GCM simulations, and that the alteration is different between the RCMs (Fig. 5 and Table S3). This, on the other hand, suggests that the uncertainties of GCMs in simulating heat wave magnitudes would be transformed by RCMs in a complex manner, rather than simply inherited, due to the nonlinear nature of model dynamics and physics, thus rejecting the concept of 'cascade of uncertainty'.*

Comment 1.11: In summary, the tiny sample of GCMs in this study severely limits the application of these results and there were some points which were unclear and needed elaboration, as pointed out above. One way to improve this is to extend the ensemble of GCMs to the whole of CMIP5 (CMIP6?), and then compare those three selected here in this study with the larger set of GCMs. There are also some issues that merit more discussion, as mentioned above. I also think it's useful to discuss other definitions of heatwaves than HWMId, even if this paper focuses on just this fairly established indicator. Furthermore, it's important to consider ways to connect these results with what can be delivered by ESD (e.g. much larger ensembles than Euro-CORDEX), and in general I suggest that papers on downscaling that ignore one of these strategies do not merit publication.

Reply: In the revised version, we explicitly stated the focus of the study and why we need a full simulation matrix without gaps (L90–93), as follows:

> *The large number of GCM-RCM combinations available from EURO-CORDEX allows us to examine RCMs' 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.*

Regarding the size issue of model members, the suggestion of use of ESD, and other definitions of heatwaves than HWMId, please see our response to Comments 1.2–1.4.

Details:

Comment 1.12: L52 “hace” is misspelt.

Reply: It is actually “have”; corrected.

Comment 1.13: Fig. 2 caption: ‘Scott’s rule’ needs a reference.

Reply: Following the suggestion of Reviewer#2, we have removed the violin-plot and thus there is nothing to do with “Scott’s rule” anymore.

Comment 1.14: L.188: Missing “there” in “shows a similar pattern to the ensemble mean (first row of Fig. 5) but exists considerable differences in the spread (second row Fig. 5) of the RCM ensembles”?

Reply: The sentence has been rephrased as:

> *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.*

2 Reviewer #2

Comment 2.1: This paper analyses the historical occurrence and future projections of heat waves in Europe using simulations from four RCMs driven a) by ERA-Interim and b) by three different GCMs. Heat waves are identified by the Heat Wave Magnitude Index daily (HWMId). The paper finds that RCMs generally reproduce well observed heat wave patterns when driven by ERA-Interim. When using historical simulations driven by GCMs, the (statistical) agreement with observations gets a bit weaker. Future projections reveal increasing heat wave magnitudes throughout Europe but with differing patterns for GCMs and RCMs. This implies that RCM patterns are not only determined by the driving GCM model, but the physical parameterization of RCMs plays an important role as well. Uncertainties of future projections are due to both GCMs and RCMs.

Reply: Thanks for your positive evaluation!

Comment 2.2: The paper is well structured and generally it is easy to follow the argumentation. Some parts would need a more concise language to be better understandable. The paper mostly focuses on analysing the contributions of RCMs and GCMs to the HWMId patterns, but also includes some statements about impacts. For the latter, I would suggest using a different approach (e.g. focusing on warming levels instead of time periods) due to the usage of the high-emissions scenario RCP8.5, as this scenario does not include any climate mitigation efforts and thus likely overestimated impacts at the end of the century.

Reply: We have made efforts during the revision on the improvements of language aspect and hope the readability has been improved.

Regarding the statements about impact, in fact, we did not intend to assess whether impacts are likely to occur or not; rather, we were simply stating the results under RCP8.5 which we analyzed and particularly pointing out our observation about the exponential HWMId increase in both GCM and RCM simulations. We choose to restrict the analysis to RCP8.5 simply because there are more EURO-CORDEX simulations forced by RCP8.5 available for formulating a full "GCM-RCM simulation matrix" with a size as large as possible. We are very grateful for your providing us with the suggestion about another approach, which is constructive and definitely worth further study. However, we tend to keep the current analysis focusing on fixed periods, as assessing the impacts is not the focus of the current study. The focus is rather "RCMs' behaviors (e.g., signal modification and uncertainty transformation) within the downscaling process for simulating heat wave magnitudes" (L90–91).

General remarks

Comment 2.3: Some parts of the result section would benefit from a revision of the language to make the paper easier understandable. Arguments are sometimes rather hard to follow, especially in the parts focusing on the analysis to which extent RCMs and GCMs are responsible for a certain HWMId pattern (e.g. lines 170-194). I would recommend formulating the text in a more concise way and try to keep the same terms throughout the paper to facilitate reading.

Reply: Thanks! We have tried to re-organize structure for some parts of the manuscript and rephrasing some text, and hope the revision make the manuscript more easily understandable.

Comment 2.4: In the results section, the coverage of certain topics is sometimes split (e.g. a table relating to a certain figure is described at a different place than the figure). I would suggest restructuring the results such that each topic is only described/discussed once (i.e. table and figure at the same time). This would make it easier for the reader to follow the argumentation. In the specific remarks I mention some examples for text that could be merged.

Reply: We have followed the suggestion.

Comment 2.5: The paper mentions several times that HWMId rises in an "exponential-like" rate. However, the paper provides no figure or analysis that shows such an exponential increase. I would thus recommend to include a time series plot showing this behaviour. Further, I am not sure if "exponential-like" is the right term: Either the increase is indeed exponentially, or it follows a different functional form.

Reply: In fact, the exponential increase has already, though not directly, been shown in Fig. 9 and 10b, where data were presented on a logarithmic scale. We noticed that approximately equal distances between every neighboring 40-year-periods as mapped to the logarithmic scale, indicating the increase is approximately exponential. Following your suggestion, we have made a time series plot (added as a supplementary figure) also with a logarithmic scale, which clearly shows a "linear" trend indicating the HWMId values varying with time following exponential form.

Comment 2.6: The study currently uses a matrix of 4 RCMs x 3 GCMs. To my knowledge, even a larger matrix of GCMs and RCMs with full RCM/GCM coverage would be available. Is there any reason why only 4x3 models were used? Have the models been specifically selected?

Reply: We did not specifically select the models but based on the full RCM/GCM coverage available at the time we conducted the analysis. It is worth mentioning that, however, EURO-CORDEX simulations forced by CNRM-CM5 are excluded considering a known issue with the CNRM-CM5 boundary conditions provided for CORDEX RCMs (Somot S. 2018, oral presentation at Annual meeting Euro-CORDEX; pdf link: https://nextcloud.meteo.fr/s/jm2p4j95SfgcNM8). Indeed, new EURO-CORDEX simulations become available as time goes forward. In fact, we have formulated another 4x3 simulation matrix with some different members, based on which we conducted the same analysis with similar results derived. We tend not to present these results in this manuscript because of too many figures and tables. Instead, we mentioned this in a new subsection of Discussion (L366–368):

> *In fact, we conducted the same analyses upon another RCM simulation matrix (with the same size but different GCM/RCM members) and derived similar results (not shown), which can alternatively support the conclusions herein.*

Specific remarks:

Comment 2.7: There are some typos and small mistakes in the manuscript that I do not list here. I would recommend to carefully read the manuscript again and correct them.

Reply: We have followed the suggestion.

Comment 2.8: Line 5: I would specify the observations (i.e. E-OBS)

Reply: We have followed the suggestion.

Comment 2.9: Line 6: Higher resolution compared to what?

Reply: We have revised it as "With their higher resolution compared to GCMs".

Comment 2.10: Line 8: What does "satisfactory way" mean? How is this determined?

Reply: We added "(e.g., by reproducing the general west-east gradient)" as an explanation.

Comment 2.11: Abstract: I think it would be nice to finish the abstract with one or two sentences describing the implications of this study.

Reply: The following sentence has been appended to Abstract (L18–19):

> *In summary, our results support the use of dynamical downscaling for deriving regional climate realization regarding heat wave magnitudes.*

Comment 2.12: Line 22: Weren't also other Scandinavian countries in addition to Sweden affected by the 2018 heatwave?

Reply: We have changed "Swedish summer" to "Scandinavian summer".

Comment 2.13: Lines 23-24: Heatwaves can also affect agricultural productions or cause forest dieback (due to lack of water or insects). Might be worth to mention this here as well.

Reply: The related sentence has been revised as:

> *In addition to health problems atmospheric heat waves are often related to water shortages, a decline in agricultural production, and increased risk of forest fires or dieback, all of which can have severe impacts both on natural ecosystems and human society (IPCC, 2014).*

Comment 2.14: Lines 54-59: I think some more argumentation would be good to explain why HWMId is used (instead of other heat wave indicators)?

Reply: The following sentences have been added (L59 and L63):

> *Most existing heat wave-related indicators describe only a single characteristic of heat waves.*
>
> *Being fairly established, therefore, the HWMId is utilized here for representing heat wave magnitudes.*

Comment 2.15: Lines 76-77: From the original publication of HWMId (Russo et al., 2015) I understand that Tmax,ref,25p and Tmax,ref,75p are calculated based on the annual maximum temperature distributions at a certain gridpoint (i.e., 30 values per grid point). It seems that in your study, the total summer distribution was used. This should be checked to make sure the calculation agrees with Russo et al.

Reply: We forgot to mention this small modification. Indeed, we used the whole daily data within the reference period for calculating these two percentiles, differing from Russo et al. (2015) who use annual values. This is because we considered that percentiles from more data will be more stable; if we fully following Russo et al. (2015), it would be 20 values for calculating the two percentiles for the evaluation part of analysis as we use 1989–2008 as the reference period. Those heatwave events reported by Russo et al. (2015) and occurred within 1989–2008 can be all identified from Fig. 2. We have also checked for the spatial pattern of these events, which are almost the same as Russo et al. (2015, Fig. 2a therein). Thus, we confirm the validity of this modification. In the revised version, we have added the following text (L81–83):

> *Here, $T_{\text{max,ref,25p}}$ and $T_{\text{max,ref,75p}}$ (utilized for normalization) differ slightly from Russo et al. (2015) who used annual data. This modification hardly influences the usage of HWMId, and somehow makes calculation more stable.*

Comment 2.16: Line 80: Remove “is calculated” at end of sentence.

Reply: We have followed the suggestion.

Comment 2.17: Line 88: Can you briefly mention why focusing only on RCP8.5?

Reply: We have followed the suggestion by adding the following text (L96–97):

> *because there are fewer EURO-CORDEX simulations forced by other RCPs, hardly available for formulating such a full simulation matrix*

Comment 2.18: Line 101-103: This sentence regarding the different reference periods sounds a bit complicated. As I understand, when RCMs are driven by ERA-Interim the reference period is 1989-2008, and when driven by GCMs it is 1990-2010. Maybe it could be explained like this?

Reply: Thanks for the suggestion. We have added the following text (L109–110):

> *When RCMs are driven by ERA-Interim the reference period is 1989–2008, and when driven by GCMs it is 1981–2010.*

Comment 2.19: Line 106: At which step were the data remapped? Before or after calculating HWMId? And were the data remapped to the ERA-Interim or E-OBS grid? And what about remapping of ERA-Interim or E-OBS?

Reply: Indeed, the original text is not clear. We have added the following information (L110–111):

> *For each dataset, HWMId were calculated upon the original grid points.*

And, we have rephrased this paragraph as:

> *Mean bias error (MBE), root mean square error (RMSE), and Pearson correlation coefficient (r) were adopted as performance indicators for a model in simulating HWMId compared to E-OBS. They quantify the degree of an overall overestimation/underestimation, the degree of closeness in values, and the association in variations, respectively. When these indicators were applied for spatial patterns, they were calculated after HWMId values of all the datasets accounted for other than ERA-Interim were conservatively remapped to the ERA-Interim grids. RMSE and r were also used to determine the similarity in spatial pattern between simulations.*

Comment 2.20: Line 110: I am not sure "effective precipitation" is the right term here, as it would also include runoff. Maybe just use P-E? Alternatively, sometimes "net surface water budget" is used.

Reply: We used $P - E$ instead.

Comment 2.21: Line 132: Which RCM simulations?

Reply: We have removed this sentence, as the previous sentences already state the case of overestimation (L146–147):

> *the RCMs fail in reproducing the ranking of these years by occasionally overestimating or underestimating the HWMId values. For example, RACMO22E overestimates HWMId excessively in 1995 as does RCA4 in 1997.*

Comment 2.22: Lines 139-140: Would be better to include this in the paragraph of lines 117-127.

Reply: We have followed the suggestion.

Comment 2.23: Lines 143-144: I guess this is due to the fact that RCMs are driven by ERA-Interim and not E-OBS.

Reply: Agreed. We have deleted the sentence as we consider it not necessary.

Comment 2.24: Lines 147-151: I would remove this here and only mention it when the respective figures or tables are discussed (see general remark).

Reply: We have followed the suggestion.

Comment 2.25: Lines 155-156: The improvement seems to depend strongly on the model. E.g. RCA4 has a relatively weak pattern correlation in Table 4.

Reply: Thanks and we have added this observation (L164–165).

Comment 2.26: Lines 163-166: Combine with lines 152-156.

Reply: We kept the text referred to where it was. This is because we structured the three paragraphs regarding Fig. 3 and Table 4 and 5. In such a sequence as: the first focusing on the influence of the shift of driving data from ERA-Interim to GCMs, the second focusing on the influence of the shift of time from evaluation-period (1989–2008) to the recent past climate (1981–2020) (along with the reference period for HWMId: 1989–2008 vs. 1981–2010), and the general difference between GCMs and RCMs under the recent past climate; and the third looking into the differences along the GCMs and how the RCMs response to them when driven by these GCMs. We hope our revision have made the mentioned structure clearer.

Comment 2.27: Lines 172: I would not necessarily call this "error". Maybe the term "error" could even be removed here.

Reply: The sentence has been removed in the revised version.

Comment 2.28: Lines 174-175: I do not fully understand this. Could this be rephrased to make it better understandable?

Reply: It has been rephrased as follows (L186–188):

> *Moreover, for each RCM driven by the three GCMs, the downscaling behaves consistently despite the large difference in the spatial pattern of HWMId between the driving GCMs.*

Comment 2.29: Line 193: “are not” instead of “would not be”

Reply: We have followed the suggestion.

Comment 2.30: Line 199: Where does the manuscript contain the information that there is no difference according to the spatial r?

Reply: The information is given in Fig. 6b, for which please refer to the caption of Fig. 6b. We have added “(Fig. 6b)” to the text referred to.

Comment 2.31: Lines 201-202: Something is missing in this sentence after “the driving”.

Reply: The mentioned sentence has been rephrased as (L215–217):

> *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 2.32: Line 215: “Observed” refers to E-OBS, right? If yes, I think it is best to mention it explicitly.

Reply: We have followed the suggestion.

Comment 2.33: Line 222: What is meant by “simulations”?

Reply: We used “RCMs” instead.

Comment 2.34: Line 243: Better performance in terms of what?

Reply: For better understandable, the text referred to has been rephrased as follows (L258–259):

> *To realistically represent HWMId, a model must be able to capture not only the mean magnitude but also the intra-annual temporal evolution of daily $T_{\max}$.*

Comment 2.35: Lines 249-255: In your study, HWMId is already studied in detail, so I am wondering if you could extend the analysis a bit more to the mentioned events. And what would be the benefit of such a detailed analysis (keeping in mind the other CORDEX evaluations that have been carried out already)?

Reply: Thanks for the suggestion. Indeed, an in-depth event-based analysis, as we called for within the text referred to, would be very helpful for understanding the performance of an RCM in representing heat wave magnitudes. In fact, we are considering to conduct some sensitivity modeling experiments, which can be more appropriate for a new study than the discussions in this paper. Following the suggestion, we have added figures about the spatial distribution of HWMId of event-years and an associated table containing MBE, RMSE, and r w.r.t. E-OBS to as supplementary materials, with the following text located in the corresponding place (before stating the need of an in-depth; L265–269):

> *Even larger variability was found in the HWMId spatial distribution of Years 1994, 2003, 2006, and 2007 (Fig. S4–S7) than for the climatological mean, although these events, which were reported by the news headlines (Russo et al., 2015), are generally captured (Fig. 2a) by the RCMs when driven by ERA-interim. Another interesting observation is that RCA4, which performs worse than the other RCMs (Table 2) in representing climatological mean, best reproduces the HWMId spatial distribution of Year 2003 (Fig. S5 and Table S4).*

Comment 2.36: Line 269: How exactly does this study show added value? What does this refer to exactly?

Reply: We have added the following information (L288–289):

> *reflected in adding more detailed geographical patterns and pulling the results closer to the observations (Fig. 3; Table 4 and 5)*

Comment 2.37: Line 273: How does this statement relate to line 263?

Reply: Line 263 corresponding to Fig. S2 (of the initial version) is about the spatial pattern of simulated HWMId values, whereas Line 273 corresponding to Fig. S3 (of the initial version) is about the spatial pattern of ensemble spread along the GCM dimension (i.e., uncertainty associated with driving data).

Comment 2.38: Line 275: Again, what is the added value referred to here?

Reply: The nonlinear manner indicates "rejecting the concept of 'cascade of uncertainty"' (L298).

Comment 2.39: Line 286: RCP8.5 is a high-emission scenario and thus, it is unlikely that the future climate will be as projected by the scenario. Thus, statements about the impacts should be made carefully when using RCP8.5. One option would be to use warming levels instead of time periods, if statements about impacts are made.

Reply: We understand your concern on making statements about impacts under such a high-end scenario. In fact, we did not intend to assess impacts that are likely going to occur; rather, we were simply stating the results under RCP8.5. See also our reply to Comment 2.2. That said, the corresponding text has been revised as follows (L309–312) to avoid that impression:

> *The RCMs, as well as the driving GCMs, project a rise in HWMId values at an exponential rate under RCP8.5 on the European continent. The exponential increase is patent as we see the linear shape of time series when plotting on a logarithmic scale (Fig. S8). As a result, heat waves more severe than the most severe one that has been recorded until now are projected to occur almost every year at the end of the century if we follow the high-end emission pathway (RCP8.5).*

Comment 2.40: Line 288: I would remove "with a strong probability". Also, what is meant by "on the alarm on"?

Reply: We have removed "with a strong probability" and "on the alarm", and have the following two sentences (L312–316):

> *According to the definition of HMWId, the approximately exponential rise can be expected because the projected warming will on one hand increase the daily magnitude (Eq. 2) and on the other hand extend the duration simultaneously. Apart from the agreement on the future severity of heat wave magnitudes under this scenario, the RCMs modify the future climate change signals projected by the driving GCMs, tending to moderate the rise in HWMId values and also deliver some different features in the spatial pattern.*

Comment 2.41: Lines 292-294 & lines 301-302: I think that missing plant-physiological effects in RCMs might also contribute to the difference between RCMs and GCMs (Schwingshackl et al., 2019; https://doi.org/10.1088/1748-9326/ab4949). I am aware that the suggestion to include more papers is always delicate (in particular, if the paper is written by the reviewer). My future review will not depend on the inclusion of this paper.

Reply: Thanks for providing the paper, which greatly helps understand the difference between GCM and RCM simulations. We have the following sentences added to the corresponding places (L318 and L333; respectively):

> *For example, plant physiological CO_2 response may have a positive effect on $T_{\max}$ (Schwingshackl et al., 2019) and thereby on HWMId values.*
>
> *According to Schwingshackl et al. (2019), these results may be linked to the positive effect of plant physiological CO_2 response on $T_{\max}$, since the GCMs except for EC-EARTH consider this response but all the RCMs do not.*

Comment 2.42: Lines 310-328: This paragraph seems rather speculative to me, as it does not include any clear causal links, but remains mostly on comparing patterns. I would suggest to either extend this analysis of potential drivers or to shorten it.

Reply: We agree that clear causal links are missing here. We still think there are some values to compare patterns, like providing a trace of possible processes behind. Thus, the paragraph has been revised as (L341–349):

> *We further examined the correlation relating to spatial characteristics between the change in HWMId and that in each of the three indices accounted for, in order to find any trace whether, and if so how, the drying processes (of either atmospheric or soil) regulate the spatial pattern of ΔHWMId. For the detailed spatial pattern of ΔHWMId, it cannot be explained by the change in the annual mean of daily $T_{\max}$ alone, even though HWMId is calculated based only on daily $T_{\max}$. This is especially true for GCM simulations as they have a poor spatial*

correlation between ΔHWMId and $\Delta T_{\max}$ (Fig. S9). All the GCM and RCM simulations agree on the high ΔHWMId in southern Europe (Fig. 3), which is very likely amplified by the projected drying trend (Fig. S10 and S11), while in northern Europe where rapid warming is projected (Fig. S9), high ΔHWMId is seen in the RCM simulations only (Fig. 3). The r values in Fig. S9–S11 read that the general warming, compared to drying, plays a small role in regulating the spatial pattern of ΔHWMId in GCM simulations, different from the case of the RCMs.

Comment 2.43: Lines 327-329: This sentence seems rather vague. I think it might be better to highlight which open questions arising from your study would be worth to be analysed by future studies.

Reply: Taking into account the comments from another reviewer, the sentence referred to has been re-organized to a new paragraph (L350–361), as follows:

> *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 2.44: Line 335: I am not sure I would expect an exact match with ERA-Interim, given that it only provides the boundary conditions.

Reply: We did not expect exact match with ERA-Interim, as the results were compared to E-OBS. Perhaps, "but still with room for improvement" caused misunderstanding, which has been removed.

Comment 2.45: Line 341: The exponential-like increase is not shown in the paper. I would suggest to include a figure showing it (see general remark).

Reply: Please see our response to Comment 2.5. We have added such a figure into supplementary information (i.e., Fig. S8) and the corresponding text (L310):

> *The exponential increase is patent as we see the linear shape of time series when plotting on a logarithmic scale (Fig. S8).*

Comment 2.46: Line 343: What exactly does "relatively more moderate rising trend" mean?

Reply: Changed to "somewhat more moderate rise" for a clearer expression.

Comment 2.47: Lines 346-347: As mentioned above, I am not really convinced by the analysis regarding the impact of drying trends on HWMId. Thus, this statement currently seems not convincing to me.

Reply: The referred text has been removed from the section of Conclusions.

Comment 2.48: Line 348: Which figure does this refer to?

Reply: The corresponding sentence has been moved to Discussion section and rephrased as (L345–347):

> *All the GCM and RCM simulations agree on the high ΔHWMId in southern Europe (Fig. 3), which may be amplified by the projected drying trend (Fig. S10 and S11), while in northern Europe where rapid warming is projected (Fig. S9), high ΔHWMId is seen in the RCM simulations only (Fig. 3).*

Remarks about figures and tables:

Comment 2.49: Figure 2: I think a time series plot for the different datasets would be easier to understand than the current Figure 2a. Moreover, I am not sure the distributions in 2b are really needed. I personally find them hard to interpret. Another option would be to replace the violin plots by PDFs.

Reply: We have followed the suggestion by removing the violin plots keeping only the box plots.

Comment 2.50: Figure 3: The red rectangle is hard to see. A different colour (e.g. blue) might be better.

Reply: We have followed the suggestion.

Comment 2.51: Table 2: If I understand correctly, this table refers to Figure 1. I would try to highlight this better because it is not instantly clear to me (same for other tables that are connected to certain figures).

Reply: We have followed the suggestion.

Comment 2.52: Figure 4 ff: I find the greyish colormap rather hard to read. Maybe choose a different one? Or adjust the limits?

Reply: We have followed the suggestion by choosing another colormap. Edited figures include Fig. 4, 5, 7, and 8.

Comment 2.53: Figure 10: I think this figure is too busy to understand. And the skewness is hardly visible to me from the violin plots. As above, I would suggest showing PDFs instead of violin plots as they are probably easier to interpret, and potentially also make the skewness visible more clearly.

Reply: We have followed the suggestion by removing the violin plots keeping only the box plots.

3 Reviewer #3

Synopsis:

Comment 3.1: 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 3.2: 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 3.3: 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 3.4: 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 3.5: 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 3.6: l. 24: I guess you mean “marine heat waves” instead of “maritime heat waves”.

Reply: Corrected

Comment 3.7: l. 26: Is it on purpose to using the terminology “warm spell” here?

Reply: Changed to “heat wave”.

Comment 3.8: l. 48: Perhaps write explicitly "differently then GCMs".

Reply: “differently from GCMs” has been added.

Comment 3.9: l. 52: "have" instead of "hace".

Reply: Corrected.

Comment 3.10: l. 80: delete "is calculated" at the end of sentence.

Reply: Removed.

Comment 3.11: 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 3.12: 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 3.13: l. 98: Do you mean "to ERA-Interim" instead of "the ERA-Interim"?

Reply: We used "to ERA-Interim" instead.

Comment 3.14: 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 3.15: 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 3.16: 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 3.17: l. 124: Delete "only" before "focusing".

Reply: We have followed the suggestion.

Comment 3.18: 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 3.19: 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 3.20: l. 182: "of" instead of "in".

Reply: Corrected

Comment 3.21: 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 3.22: 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 3.23: 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 3.24: 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 3.25: 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 3.26: 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 3.27: l. 349: "Plays" instead of "play".

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

Comment 3.28: 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 3.23.