Comment 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: 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 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 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 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 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 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 8: Line 5: I would specify the observations (i.e. E-OBS)

Reply: We have followed the suggestion.

Comment 9: Line 6: Higher resolution compared to what?

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

Comment 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 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 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 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 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 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, $T_{\mathrm{max,ref,25p}}$ and $T_{\mathrm{max,ref,75p}}$ (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 16: Line 80: Remove "is calculated" at end of sentence.

Reply: We have followed the suggestion.

Comment 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 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 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 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 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 22: Lines 139-140: Would be better to include this in the paragraph of lines 117-127.

Reply: We have followed the suggestion.

Comment 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 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 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 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 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 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 29: Line 193: “are not” instead of “would not be”

Reply: We have followed the suggestion.

Comment 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 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 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 33: Line 222: What is meant by “simulations”?

Reply: We used “RCMs” instead.

Comment 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 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 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 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 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 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. 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 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 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 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 ΔT_{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 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 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 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 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 46: Line 343: What exactly does "relatively more moderate rising trend" mean?

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

Comment 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 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 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 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 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 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 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.