Comment on egusphere-2022-215
Anonymous Referee #1


General comments

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.

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.

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 downsampling (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.

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.

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).

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.

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.

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.

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)

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.

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.

Details:

L52 “hace” is misspelt.

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

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”?