Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-253-AC1, 2020




# *Interactive comment on* "The benefits of increasing resolution in global and regional climate simulations for European climate extremes" *by* Carley E. Iles et al.

**Carley E. Iles et al.**

carley.iles@lsce.ipsl.fr

Received and published: 17 March 2020

Anonymous Referee #1 Received and published: 25 November 2019

The paper, in general, is well written and the authors attempt to solve a very critical issue regarding the added value of increasing resolution. However I do have some issues regarding the dataset used and some of the methodologies applied, hence I recommend major revision.

Thank you for your comments and suggestions. Please find our point by point responses below.

Here my major comments:

a) All the datasets used to compare the models' output are labelled as observations when, in reality, they are not, they are reanalyses. So I wonder why the authors chose a reanalysis when there are real observational datasets, like GPCP, CHIRP or TRMM. I do know all those datasets have some caveats, but they are derived from direct observations, (and most of them are of the same resolution like the one that the reanalysis used in the paper). At least for precipitation and temperature, I suggest that the authors re-do the comparisons with some of the observational datasets mentioned. For wind, I understand that there might not be another option that the one that they used.

In fact, for temperature and precipitation we primarily use E-OBS, which is a gridded station based dataset. The MESAN reanalysis for precipitation is not a typical reanalysis, in that it is adjusted using station measurements of precipitation afterwards. We now make the text clearer on this (see below). For wind, yes, it is true that reanalyses were the only readily available option. However, we have now changed the reanalyses used for wind in response to reviewer 2's comments to DYNAD (which is related to MESAN), MESCAN and ERA5. The first two are 5.5km resolution downscalings of the 22km resolution HIRLAM and 11 km UERRA-HARMONIE reanalyses respectively (see section 2.1 for details). ERA5 was not available when we first conducted the analysis, but we now use it to replace the ERA-interim based wind estimates (ECEM and WFDEI). We have now been more careful in the text to specify where we mean observations and where we mean reanalyses.

"results are repeated for precipitation extremes using the 5.5 km resolution MESAN reanalysis (Landelius et al. 2016), which adjusts a downscaled first guess from the 22km resolution HIRLAM reanalysis (Dahlgren et al. 2016) with a network of station-based precipitation observations."

b) I do not understand why the authors chose to use climatological means of Txx and Rx1day. Would not be more useful to have seasonal maps? Mostly for precipitation, as extremes, can occur either in summer or in winter but by very different processes. In this way, we could have seen which type of seasonal extremes are better (or worse) capture and this might also complement the results of the analogues.

Due to space constraints we are not able to show seasons. For wind and temperature this is unlikely to matter (since heatwaves are a summer phenomenon), but for precipitation it might make a difference. We attempted to examine precipitation extremes arising from different processes by examining both Rx1day and Rx5day, the former being more likely to represent convective thunderstorms, and the latter more large scale precipitation (although it is not a perfect division, but seasons would not be either). We did not find large differences between the results for Rx1day and Rx5day, so we only show the former, and mention the latter in some places.

c) On the same idea, why to use a 5-day average (line 568) mainly for precipitation, given the fact that by doing that you are smoothing the extreme event that normally would last one day. For temperature, it might not be a problem, considering that heatwaves per definition last several days(at least 3).

In fact, in the paper we show 1 day precipitation (Rx1day) and 5 day temperature. 5 day precipitation was also examined, but is not shown, although the results are mentioned in a couple of places in the text. We now make this clearer by adding this sentence at line 242 “Rx1day and TXx5day are presented in the figures, whilst the other indices are commented on in the text.” (“other indices” refers to TXx and Rx5day).

General comments:

1. Line 40, please define what do you mean by “small scale”, do you mean synoptic, mesoscale?

We have added :” i.e on the order of a few km to a few hundred km”

2. Line 60, remove “really”

Done

3. Line 62, added to the list of reference: Risanto, C.B.; Castro, C.L.; Moker, J.M., Jr.; Arellano, A.F., Jr.; Adams, D.K.; Fierro, L.M.; Minjarez Sosa, C.M. Evaluating Forecast Skills of Moisture from Convective-Permitting WRF-ARWModel during 2017 North American Monsoon Season. Atmosphere 2019, 10, 694.

We have now added this reference.

4. Line 66-67, I think the justification of centre this paper on Europe, has to other than" "its climate is highly variable and affected by a range of both large and small scale processes which present challenges for adequate simulation", as this is true for several regions in the planet.

The reviewer is correct. This statement was more of an aside point, rather than the main motivation. Europe also has a large number of coordinated RCM simulations at two different resolutions as part of the EUROCORDEX project, which lend themselves to this kind of analysis. We edit the existing sentence as follows:

We will address these questions focusing on Europe, for which a large number of coordinated RCM simulations at two standard resolutions are available as part of the EUROCORDEX project (Jacob et al., 2014), and whose climate is highly variable and affected by a range of both large and small scale processes, which present challenges for adequate simulation.

5. Line 56: " : : :history and trajectory of air masses ARE important : : : (instead of"is")

Done

Anonymous Referee #2 Received and published: 1 December 2019

In the manuscript "The benefits of increasing resolution in global and regional climate simulations for European climate extremes" Iles et al. assess the dependence of simulating extreme event intensity on model resolution and model strategy (regional vs. global) over Europe. They show that higher resolution simulations have generally stronger extremes with higher sensitivities for precipitation and wind extremes than for temperature extremes. Their results generally confirm previous studies and for me, the greatest value of this paper is the combination of a large range of models and investigating three extreme events, which provides a good overview. I also like the upscaling analysis although it is kept fairly brief. However, I have concerns in how the team compared models with different resolutions and would like to see a more careful usage of the word model bias especially in the context of precipitation and wind analysis for which the used observational datasets are of questionable quality. Below is a more detailed description of my major/general and minor comments.

We would like to thank the reviewer for his/her very thorough and constructive review. Please find our point by point responses below.

Major/general comments:

1. I have concerns with using nearest neighbor remapping for extreme value analysis. This remapping method will artificially increase extremes in high-resolution data particularly in situations with strong gradients. For instance, precipitation extremes are typically very localized and have strong spatial gradients. You are remapping the 0.11 EURO-CORDEX simulations to a 0.5deg grid. This means that you pick the closest 0.11deg grid cell to each 0.5deg cell and assume that it rains ass much on the 0.5deg area than on the 0.11deg area. This is certainly not the case and by doing this, you violate mass conservation of precipitation. I strongly recommend using a conservative remapping method that conserves mass and energy while remapping. This will affect many of your results since you generally find that extremes are more intense in the high-resolution simulations. Using conservative remapping will dampen this effect and might change your conclusions.

We agree with the reviewer that using nearest neighbour is not the most appropriate technique for going from high to low resolution, as we only use information from a small subset of grid cells. This can indeed be problematic close to strong gradients. But conservative remapping made the extremes much weaker than any other technique, especially for CORDEX 0.11 (see figure S1) and there seems to be a dampening effect not only from averaging over larger areas (as we expect), but also a further dampening from splitting grid cells that fall on the boundaries of the new grid into two or more- which happens to a large proportion of grid cells. We also see this dampening effect of conservative remapping when going from low to high resolution. We therefore changed the regridding for cordex 0.11 (and similar resolution observational datasets) to bicubic interpolation, since this also replicated the results using the original grid well, whilst using information from all grid cells.

So in summary, we decided to use bicubic interpolation for going from high to low resolution (for CORDEX 0.11, MESAN and MESCAN and ERA5) and nearest neighbour for everything else. Figure S1 shows that this decision (to use nearest neighbor) seems appropriate even for medium resolutions (e.g. CORDEX 0.44 and UPSCALE 25km (N512)).

2. Along coastlines, you have to be careful that you only consider land grid points in your evaluation against land-based observations. As you mention in your wind speed analysis, there is a sharp gradient in wind speed but also for temperature extremes between ocean and land. I am not sure if your interpretation that high wind extremes propagate further inland in coarse-scale models is correct. Could it be that you include ocean grid cells in your analysis, which cause a high bias in low-resolution models?

On closer inspection, we agree that the mask used for wind included too many ocean grid points. This was based on the mask of the WFDEI dataset. Instead we have applied the E-Obs mask (which was also used for temperature and precipitation), which you can see in figure 4 approximates the land much better. This does affect some of the results for wind, and we thank the reviewer for pointing this out. Since model land masks contain information on land area fractions rather than binary land/not land we expect that the coastal grid cells we refer to are a mixture of land and ocean – this still has the effect of making ocean influences appear further inland – just by nature of the grid box overlapping land and ocean areas more- this is what we meant by “propagate further inland". We have made this clearer in the text and added the possibility of there being other differences in the land masks with resolution. We prefer not to select only grid cells that are 100% land from each model (leading to a different land area for each model) since the increasing detail and accuracy of a model's land mask is a fundamental advantage of increased resolution and correcting for it constitutes a sort of bias adjustment. Inaccurate model land masks also have implications for local scale climate impact management decisions.

3. Please use a lapse rate adjustment when comparing temperature extremes between model grid spacings. Your finding that heat extremes are less overestimated in high-resolution simulations in topographic regions is trivial since the coarse resolution models have lower topography and therefore higher temperatures. You could simply use the topography of E-OBS and correct the model temperature with a climatological average lapse rate (e.g., 6.5 deg C per km).

We think that the higher topography associated with higher resolution is a fundamental benefit of increasing resolution, just as the lower topography in low resolution models is a fundamental drawback. Correcting for differences in topography is a kind of bias correction. We prefer to present the models as they are, without such adjustments.

4. Please avoid writing about model biases when you compare extreme precipitation to E-OBS. You mention that E-OBS underestimates precipitation extremes, which means that you would like a model to be wetter than E-OBS. Evan the MESAN precipitation extremes are likely too low. I suggest being very careful how you use the word bias and rather use difference when comparing to precipitation but also wind datasets.

Good point. We have now changed the language in these sections to reflect that these are not necessarily biases, but differences.

5. The separation of CMIP5 and CORDEX results from the UPSCALE results make the paper unnecessary long and harder to read. I recommend combining sections 4.1 and 4.2 since I do not see a good reason to separate the UPSCALE ensemble from the

other datasets. Also, combining figures 1, 3, and 4 with 5, 7, and 8 would be beneficial. You could do this by not showing the results for all and common ensemble members but only one of the two. They are very similar anyway.

We prefer to keep these two sections separate. They are in fact addressing different questions. The CORDEX vs CMIP5 analysis is looking at the benefits of regional dynamical downscaling, whilst the UPSCALE analysis addresses the effects of increasing resolution globally. We feel that combining the sections would make the paper less clear to follow.

6. Also, the summary and discussion section could be combined. After the summary section you again briefly summarize results in the discussion. You would lose very little information by removing the summary section.

We have now combined these two sections.

7. I am concerned with how different the three wind observations are. Are they all equally likely? Looking at the big differences between these observations makes me wonder if you can/should evaluate the models at all. Just looking at your fields in Fig. 4 (c,f,i) makes me wonder if these datasets can even capture topographic effects. The low wind extreme minimum over the Alps looks very unrealistic and event he models seem to do a better job capturing these effects that the observations. Would the use of MESAN winds be a better option?

We agree with the reviewer that the ECEM dataset with the Weibull distribution based bias correction seems unrealistic (this bias correction technique may have worked for mean winds, but not for the extremes). We have opted to replace all three datasets. We now use MESAN winds (actually called DYNAD), as suggested by the reviewer, but also MESCAN (which like MESAN is available at 5.5 km resolution, and is constructed by downscaling of the 11km UERRA-HARMONIE reanalysis with a NWP) and ERA5. The spread amongst the datasets is now much less, although still present. ERA5 seems to have particularly slow winds, including over the Alps, but we show it since it is a new reanalysis that many people are likely to make use of.

8. Please be careful when you use the term model resolution and do not confuse it with model horizontal grid spacing. The model resolution is typically 4 to 8 times the model horizontal grid spacing (Skamarock 2014)

We have now been more careful in the introduction to clarify when we actually mean horizontal grid spacing rather than resolution, and have also added a clarifying sentence at line 72 stating “Throughout the rest of this manuscript we use the term “resolution” to mean model horizontal grid spacing, whilst recognising that a model’s effective resolution, in terms of the scales it can capture, is always less than its grid spacing.”

L37-8: Please explain what a long-standing anticyclone is? Do you mean a stationary anticyclone?

Yes, we have renamed it “stationary anticyclone”

L41: I would say “flash floods” here since river floods involve large-scale processes.

Agreed, we have changed the text accordingly.

L43: I suggest “poorly resolved” since some of these processes can be resolved at fairly large grid spacings.

Agreed, we have changed the sentence to “These are poorly resolved at the resolution of Global Climate Models (GCMs) in CMIP5 (Coupled Model Intercomparison Project Phase 5; Taylor et al., 2012)”

L50: Could you please be more specific than “small-scale processes and features involved”

We have added the following (in bold): “For precipitation and wind extremes, an improvement with resolution could be expected due to the small-scale processes and features involved, including convection and the influence of topography.”

L89: What do you mean with “their” in “to their information”?

“Their” refered to “heatwaves”. We have now changed the text to explicitly say “heatwaves”

L136: E-OBS has quite high station density in some regions (e.g., Slovenia, Germany) but low density in others (e.g., Austria, Spain)

This is a good point. We have updated the sentence to the following: E-OBS has a somewhat non-uniform underlying station density, with relatively high densities in Germany, Sweden and Slovenia, and low densities in other countries (e.g. Spain, France, Austria). It tends to underestimate precipitation extremes relative to higher density regional datasets, especially where it has poor coverage, due to missed extremes which are local in scale (Prein and Gobiet 2017).”

L168L: “us to identify”

Corrected

L173: Are these daily maximum surface wind speeds based on model time step wind speed or hourly maxima of instantaneous wind, or something else?

The variable sfcWindmax in model outputs is based on the model time step wind speed, although figure S7 (which compares return periods of annual maximum wind based on sfcWIndmax with that based on 3 hourly and 6 hourly data for the GCMs that have both) suggests that this is not the case for the IPSL models or CMCC-CM for which annual maximum sfcWindmax has slower speeds than both 3 and 6 hourly estimates. This difference in definition between models is a weakness of the analysis. Also, since we expect the model time step to decrease with increased resolution, we would expect this to result in stronger winds with higher resolution due to the increased sampling frequency. Whilst in some ways this latter point makes the models not strictly comparable, being able to generate stronger winds due to a shorter time step would nevertheless be an inherent feature of higher resolution models.

Comparing all models using 3 hourly (or 6 hourly) data would have been tidier, but this data was simply not available for CORDEX 0.44 and very limited for CORDEX 0.11. These caveats and their implications are made clear in the text.

To the former line 173 we have added "This seems to mostly be based on model timestep wind speed, with a few exceptions (see figure S7). The implications of this are discussed further in the results section." And to the results section we have added some discussion about differences in model time steps.

L174: "Those consist of"

We changed to "These consist of"

L192: What is this alternative dataset for SST?

It was the OSTIA analysis (Operational Sea Surface Temperature and Sea Ice Analysis) (Donlon et al. 2012). We have added this to the text.

L203: Performing your analysis on a 0.5deg grid is fine but you have to say that this will deteriorate the skill of the coarser grid spacing simulations.

We have added further discussion of these points. This paragraph now reads (updates in bold)

"In order to compare models of different resolutions with each other and with observations it was necessary to regrid variables to a common grid. Using a high resolution grid for evaluation would preserve the finer spatial detail and localised extremes for high resolution simulations, but is sometimes considered unfair for coarse resolution models which cannot be expected to simulate the same intensities of extremes even for a perfect simulation due to spatial smoothing effects (Prien et al. 2016). However, the finer spatial detail is an inherent advantage of high resolution and smoothing this out will result in information loss. We use a 0.5° regular longitude-latitude grid since it is in-between the resolution of the CORDEX models and CMIP5, is computationally feasible and E-OBS is also available at this resolution. Some of the benefits of higher resolution may be lost by doing this, putting our results on the conservative side. Nevertheless, sensitivity tests showed that results for MESAN did not change perceptibly by using a 0.5° grid compared to a 0.1° regular grid. We regrid the daily data, before the calculation of annual extreme indices. “

L228: Temperature can have strong gradients along coastlines and in orographically complex regions that you mention quite a lot in your results. Therefore, using bilinear interpolation might also not be the best choice here.

Thank you for raising this point. We agree that using bilinear interpolation will falsely smooth some features. Having now also repeated the sensitivity analysis to regridding that we did for precipitation (fig S1) for temperature, we found a similar (but reduced) sensitivity of temperature extremes to regridding method. Bilinear interpolation reduced the return values seen in the return period plots. We found that the choices we made for precipitation and wind (nearest neighbour for regridding from low to high resolution, or between similar resolutions, and bicubic for high to low resolution) were also the best choices for temperature in terms of replicating the return curves that we get by using the original grid, whilst also preserving the blocky nature of the spatial patterns from the lower resolution models.

L228: I agree that these are rare events but you should mention that you decrease your sample size by looking at rare events, which makes robust statistics more challenging.

We add “One drawback is that this makes robust statistics more challenging.”

L251: I am not familiar with this method of pooling extreme values from an ensemble. You correct the models for a mean bias but you do not correct for the shape and scale of the tail of the distribution. Does this not mean that you base your high return values mainly on the models that have a very long tale? Is there a reference for this method?

We have now replaced the results based on pooling with ensemble means. This was because we realized that pooling interacted with the spatial averaging to change the shape of the distributions (especially for precipitation), making them no longer comparable to the observations. This was also the case for the UPSCALE simulations where there was no issue with different models being combined in the pooling. Instead we use the ensemble mean or median, which retains comparability to the observations.

Regarding the distribution of the models in the pooled distributions, figure S3 (in the original submitted version, now removed) showed that the tails were not dominated by any one model, although models were not spread 100% evenly across the pooled distribution either

L283-7 & 304-8 & 318-9 & 364-8: This information does not have to be duplicated here since it is already mentioned in the figure caption.

This information has now been deleted or shortened in these places.

L332-3: Gauge under catch could not only contribute it definitively does. This can be substantial especially for extreme precipitation that are associated with high wind speed (northern latitudes) and in snow-dominated regimes.

Added: "Gauge undercatch will also contribute to the difference, particularly for precipitation extremes associated with strong winds and in snow dominated regions"

L474: I can also see this in the N96 simulation but the Alps are much wider at this grid spacing.

We change the wording to: "All resolutions have bands of heavy precipitation either side of the Alps, but these move closer together as the Alps become better defined"

L481-3: A model's grid spacing is always higher than its resolution (see e.g. Skamarock 2014).

Thank you. We have now added a sentence on this at line 72 (see response to comment 8 above). We have also updated this sentence to "In addition, it should be noted that models with the same nominal resolution do not necessarily have the same effective resolution, and that the effective resolution is always less than the nominal resolution." The main point we wanted to communicate here was that just because two models claim to be the same or similar resolutions in terms of grid spacing, it doesn't mean they are the same in terms of the actual scales that can be resolved.

L559: Did you detrend your 500 hPa geopotential height before you did the analog analysis. There is a high chance that the 500 hPa geopotential increased during your simulation period, which might affect your analog analysis.

500 hPa was not detrended in the original analysis. We have re-run the analogue analysis and replaced the relevant figures using pattern correlations as the measure of distance between circulation states instead of Euclidean distance. This should get around issues relating to trends in geopotential height. Results were not sensitive to this change in method. We have updated the text accordingly: "Similarity between circulation states is quantified using pattern correlation, which is not affected by trends in geopotential height with global warming"

L542-73: This should go into the methods section.

We think that it improves the readability of the paper to keep the description of the circulation analogues method here. The circulation analogue analysis builds on the analysis presented in the rest of the paper, and we feel that moving the description to the methods section would overload the reader, forcing them to imagine many steps that become more concrete after seeing the figures in the rest of the paper.

L696-7: Getting benefits from upscaling might demand convection-permitting climate simulations (Hart et al. 2018).

Thank you, we have added this citation.

L715: This is not true in Ban et al. (2015). They show very similar increases in extreme precipitation between their 12 km and 2 km model results.

This is true for daily precipitation, but they do see a greater increase for sub-daily summer precipitation in the 2km model. We have modified this sentence to specify only summer sub-daily intensities (rather than both daily and sub-daily), and made the sentence read less definitively.

Literature:

Skamarock, W.C., 2004. Evaluating mesoscale NWP models using kinetic energy spectra. Monthly weather review, 132(12), pp.3019-3032.

Hart, N.C., Washington, R. and Stratton, R.A., 2018. Stronger local overturning in convectiveâĔŸARĔĞ permitting regional climate model improves simulation of the subtropical annual cycle. Geophysical Research Letters, 45(20), pp.11-334.

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-253, 2019.