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

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

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

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 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 convectiveâ˘AR˘ permitting regional climate model improves simulation of the subtropical annual cycle. Geophysical Research Letters, 45(20), pp.11-334.

The benefits of increasing resolution in global and regional climate simulations for European climate extremes

Carley E. Iles[1], Robert Vautard[1], Jane Strachan[2], Sylvie Joussaume[1], Bernd R. Eggen[2] and Chris D. Hewitt[2]

[1] Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91198 Gif-sur-Yvette, France
[2]Hadley Centre, Met Office, Exeter, UK

Correspondence to: Carley E. Iles (carley.iles@lsce.ipsl.fr)

Abstract. Many climate extremes, including heatwaves and heavy precipitation events, are projected to worsen under climate change, with important impacts for society. Future projections, required for adaptation, are often based on climate model simulations. Given finite resources, trade-offs must be made concerning model resolution, ensemble size and level of model complexity. Here we focus on the resolution component. A given resolution can be achieved over a region using either global climate models (GCMs) or at lower cost using regional climate models (RCMs) that dynamically downscale coarser GCMs. Both approaches to increasing resolution may better capture small-scale processes and features (downscaling effect), but increased GCM resolution may also improve the representation of large-scale atmospheric circulation (upscaling effect). The size of this upscaling effect is therefore important for deciding modelling strategies. Here we evaluate the benefits of increased model resolution for both global and regional climate models for simulating temperature, precipitation and wind extremes over Europe at resolutions that could currently be realistically used for coordinated sets of climate projections at the pan-European scale. First we examine the benefits of regional downscaling by comparing EURO-CORDEX simulations at 12.5 and 50 km resolution to their coarser CMIP5 driving simulations. Secondly, we compare global scale HadGEM3-A simulations at three resolutions (130, 60 and 25 km). Finally, we separate out resolution dependent differences for HadGEM3-A into downscaling and upscaling components using a circulation analogue technique. Results suggest limited benefits of increased resolution for heatwaves, except in reducing hot biases over mountainous regions. Precipitation extremes are sensitive to resolution, particularly over complex orography, with larger totals and heavier tails of the distribution at higher resolution, particularly in the CORDEX vs CMIP5 analysis. CMIP5 models underestimate precipitation extremes, whilst CORDEX simulations overestimate compared to E-OBS, particularly at 12.5 km, but results are sensitive to the observational dataset used, with the MESAN reanalysis giving higher totals and heavier tails than E-OBS. Wind extremes are somewhat stronger and heavier tailed at higher resolution, except at coastal regions where large coastal grid boxes spread strong ocean winds further over land. The circulation analogue analysis suggests that differences with resolution for the HadGEM3-A GCM are primarily due to downscaling effects.

1 Introduction

Climate extremes, such as heatwaves and heavy precipitation events are projected to worsen under climate change, with important impacts for society (Seneviratne et al., 2012). Such projections are generally based on numerical climate model simulations. However, given finite computational resources, trade-offs between model resolution, ensemble size and the level of model complexity are necessary. For extreme events driven by large-scale processes such as ~~long-standing~~stationary anticyclones, the proper simulation of the amplitude of extremes is limited by dynamics but also by land-atmosphere feedbacks and the many physical processes involved in the surface energy budget. Such extremes are typically heat waves, droughts and cold spells. Many other types of extreme event are by nature small scale, i.e on the order of a few kilometers to a few hundred kilometers. Such is the case of convective precipitation, flash floods, extratropical wind storms, cyclones and medicanes. These are poorly resolved at the resolution of Global Climate Models (GCMs) in CMIP5 (Coupled Model Intercomparison Project Phase 5; Taylor et al., 2012)~~ does not allow these events to be resolved explicitly~~. Increased resolution in GCMs may improve the representation of small-scale processes and features, including orography and coastlines (downscaling effect), but potentially may also improve the representation of the interaction between small and large scale dynamical processes and ultimately improve the large-scale atmospheric flow (upscaling effect). For instance, a better representation of baroclinic eddies may help to better simulate large Rossby waves such as those inducing long-lived anomalies, due to the inverse energy cascade. This may improve the simulation of the frequency and duration of heat waves and cold spells, and related anomalies such as summer droughts. 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. However, upscaling effects may also have benefits by improving storm-track location, and duration of wet spells. An alternative approach to increasing the resolution of global-scale models is to use regional climate models (RCMs) driven by coarser GCMs to achieve a given high resolution over a limited area at lower cost. However, this technique only captures downscaling effects, since the RCM inherits the large scale circulation from the driving GCM.

Current generation GCMs commonly used for climate projections (e.g. CMIP5 models) have a horizontal grid spacing~~resolution~~ ranging from about 70 to 250 km~~ resolution~~, although 25 km GCMs are starting to be run under projects such as PRIMAVERA and HighResMIP (part of CMIP6; Haarsma et al., 2016). For coordinated RCM experiments, such as CORDEX (Coordinated Regional Downscaling Experiment; Giorgi et al., 2009), grid spacing~~resolutions are~~ is generally between 10 to 50 km (e.g. Jacob et al., 2014). In order to simulate convective precipitation a grid spacing~~resolution~~ of <5 km is~~ really~~ needed, which is very computationally expensive, but such ensembles of convection permitting RCMs are currently in development (e.g. Coppola et al., 2019; Risanto et al. 2019). An important question is the extent to which increased resolution benefits the simulation of extreme events for both global and regional models for the kind of resolutions that can realistically be run for coordinated pan-continental climate projections. Particularly, whether using global high resolution adds further benefits over regional high resolution due to an improved large scale circulation. 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. We focus on extreme precipitation, temperature and wind, to cover a range of events that may be affected by resolution in different ways. 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 (Skamarock 2004; Klavar et al. 2020).

The benefits of increased resolution for European precipitation extremes are well documented, whilst the effects on heatwaves, cold spells and wind extremes are less well known. In GCMs, global precipitation tends to increase with resolution, and for grid point models the fraction of land precipitation and moisture fluxes from land to ocean increases, largely due to better resolved orography (Vannière et al., 2019; Terai et al., 2018; Demory et al., 2014). Precipitation extremes tend to get heavier and agree better with observations (Wehner et al., 2010, O'Brien et al., 2016; Kopparla et al., 2013; Shields et al., 2016; Vannière et al., 2019), unless the parameterisation schemes are not suited to the resolution (e.g. Wehner et al., 2014). In Europe, Schiemann et al. (2018) find that both mean and extreme precipitation are simulated better with increased resolution in HadGEM3A, mostly originating from better resolved orography. In contrast, Van Haren et al. (2015a) find that improvements in Northern and Central European mean and extreme winter precipitation with resolution are mostly associated with improved storm tracks in EC-Earth. For RCMs, extreme precipitation is improved with resolution when compared to high resolution observations, particularly over orography, including frequency-intensity distributions and spatial patterns, (see e.g. Torma et al., 2015 and Prein et al., 2016 for EUROCORDEX at 12.5 km vs 50km and vs the driving GCMs, and Ruti et al., (2016) for Med-CORDEX). However, benefits are smaller for regional and seasonal mean precipitation. Convection permitting models (<4km ~~resolution~~grid spacing) are particularly beneficial in simulating summer extreme and sub-daily precipitation, including the diurnal cycle of convection, but can overdo extreme precipitation (e.g. Prein et al., 2015; Kendon et al., 2012; 2014).

For heatwaves, increasing horizontal resolution does not lead to obvious benefits in RCM simulations (see e.g. Vautard et al., 2013 for EURO-CORDEX), except improved spatial detail (Gutjahr et al., 2016). However, increased resolution may have more impact in global models since the large scale circulation that contributes to heatwave~~their~~ formation may be affected. This remains a largely unstudied question, with the exception of a few studies such as Cattiaux et al. (2013) who find that increasing resolution in the IPSL GCM leads to a reduction in the cold bias of both cold and warm extremes in Europe, along with improved statistics, such as duration and frequencies and improved weather regimes.

For wind extremes, stronger winds and better spatial detail with resolution have been found for regional models (e.g. Pryor et al., 2012; Kunz et al., 2010). Donat et al. (2010) found that observed storm loss estimates for Germany could be reconstructed more accurately through dynamical downscaling compared to using the coarser resolution driving ERA-40 data directly. Ruti et al., (2016) found improvements in Mediterranean cyclogenesis in coupled Med-CORDEX RCMs relative to the ERA-interim driving data, whilst extreme winds over the Mediterranean generally improve (i.e. are stronger) with higher resolution RCMs (e.g. Ruti et al. 2016; Hermann et al. 2011). However, most GCM studies focus on the simulation of extratropical cyclones rather than wind directly. Such studies find an improvement in the representation of various aspects of Northern Hemisphere extratropical cyclones with increased resolution, including frequency, intensity and the position of the storm tracks (Colle et al., 2013; Jung et al., 2006; 2012), even in the higher resolution CMIP5 models (~<130 km; Zappa et al., 2013). Vries et al., (2019) found that the resolution of Atlantic Gulf-Stream SST fronts affects winter extratropical cyclone strength. Whether these improvements translate into an improvement in wind extremes remains to be assessed.

Persistence of weather regimes, such as blocking or the phase of the North Atlantic Oscillation, can be important drivers for extreme events in Europe. Using the ECMWF IFS model, Dawson et al., (2012; 2015) find that such weather regimes cannot be simulated realistically at typical CMIP5 resolution (~125 km grid spacing), but are improved at 40 km, and well-simulated at 16 km. Cattiaux et al., (2013) find improvements at more modest resolutions in the IPSL model. Blocking frequency tends to be underestimated by CMIP5-resolution climate models (Anstey et al., 2013). This tends to be improved with resolution, particularly over the North Atlantic (Jung et al., 2012, Anstey et al., 2013; Matsueda et al., 2009, Berckmans et al., 2013, Davini et al., 2017a; 2017b), although results tend to be somewhat sensitive to season and model considered (Schiemann et al., 2017) and compensating errors may be involved (Davini et al., 2017a for EC-EARTH). O'Reilly et al. (2016) find that having a well-resolved Gulf stream SST front is also important for European winter blocking and associated cold spells. An important question is whether these improvements in the large scale circulation translate into an improvement in the simulation of European climate extremes.

Here we examine the benefits of increased resolution for global models compared to regional models for the simulation of European heatwaves, heavy precipitation events and wind storms. We further break down any resolution related differences for a global model into upscaling and downscaling components. This will shed light on whether potential improvements in the large scale circulation suggested in the literature translate into an improved representation of climate extremes. This is an important consideration in choosing how to distribute finite resources between global and regional models. We focus on the kind of models widely used to provide climate projections at a European scale, applying a consistent approach across model types. Firstly, the benefits of regional dynamical downscaling are explored by comparing EURO-CORDEX simulations at 50 and 12.5 km resolutions to their coarser driving CMIP5 GCMs. Secondly, the benefits of increased resolution for a global model are examined using HadGEM3-A at 130, 60 and 25 km resolution. Finally, the roles of upscaling versus downscaling will be examined using a circulation analogue technique applied to HadGEM3-A.

2 Observational and model data

2.1 Observations

Model simulations are evaluated using observational and reanalysis datasets. For daily precipitation and daily maximum temperature, we use the gridded station based dataset E-OBS on a 0.5° latitude-longitude grid (Haylock et al. 2008). This covers the European domain from 1950 to present. Gridded datasets tend to reduce the magnitude of extremes compared to station data through smoothing effects, but are more comparable to the grid box averages from GCMs (Haylock et al. 2008). 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,~~Nevertheless, E-OBS has a relatively low underlying station density, and tends to underestimate precipitation extremes relative to higher density regional datasets),~~ due to missed extremes which are local in scale (Prein and Gobiet 2017). However, such high resolution datasets are not available at a pan-European scale. As a compromise, results are repeated for precipitation extremes using the 5.5 km resolution MESAN reanalysis (Landelius et al. 2016), which ~~combines~~adjusts a downscaled first guess~~ information~~ from the ~~high~~ 22km resolution HIRLAM reanalysis (Dahlgren et al. 2016) with a network of station-based precipitation observations. For much of Europe these are the same as those used for E-OBS, but with the addition of Swedish Meteorological and Hydrological Institute (SMHI) stations over Sweden, and a high density of Meteo-France stations over France (Landelius et al. 2016). MESAN provides daily precipitation data for the more limited period 1989-2010.~~ We use the version available on a 0.11° rotated grid.~~ Prein and Gobiet (2017) find that it gives heavier extremes than E-OBS in some regions (France, Spain, the Carpathians), but generally not as high as the high resolution regional datasets (except in France). Neither dataset is corrected for gauge undercatch, which tends to be around 3-20% for rain, and up to 40% for snow, or even 80% for non-shielded gauges (Førland and Institutt 1996; Goodison et al. 1997).

Wind extremes tend to happen on sub-daily time scales, necessitating the use of sub-daily data to avoid missing as many events (although events, or their peak magnitude, will still be missed). We use 10 m wind speed from three ~~observational wind~~ reanalysis datasets. These are the ~~–~~EURO4M DYNAD (Landelius et al. 2016), UERRA MESCAN-SURFEX (Bazile et al. 2017) and ERA5 (Hersbach et al. 2019) reanalyses. The former is available at 6 hourly intervals on a 5.5km rotated grid over Europe for the period 1979-2013 and is computed through dynamical adaptation a downscaled version of the 22km resolution HIRLAM reanalysis to 5.5 km resolution orography using DYNAD (a simplified version of HIRLAM). MESCAN is also available at the same spatial and temporal resolution over Europe from 1961 onwards, but is computed through dynamical downscaling of the 11 km UERRA-HARMONIE reanalysis. Both HIRLAM and UERRA-HARMONIE are forced by the ERA interim global reanalysis (ERA40 before 1979 for the latter). Finally, ERA5 is available globally at 0.25° and at hourly resolution from 1979 onwards. We sub-sample ERA5 to 6 hourly data in order to be consistent with the other reanalyses.

~~These are all based on the ERA-Interim reanalysis (Dee et al. 2011), but differ in the way they are processed. The first is the WFDEI dataset (WATCH-Forcing-Data-ERA-Interim; Weedon et al. 2014) and the other two are ECEM~~

~~datasets (European Climate Energy Mixes; Jones et al. 2017). One ECEM dataset is bias corrected using a Weibull distribution based on the HadISD station dataset (Dunn et al. 2012) applied to each grid cell (ECEM-wbc), whereas the other version contains no bias correction (ECEM-noc). WFDEI is available at 3 hourly resolution, whereas ECEM is 6 hourly. Therefore, 6 hourly data is used from both datasets for consistency. All datasets are available on a 0.5° regular latitude longitude grid for the period 1979-2016. Although neither ECEM-noc or WFDEI are bias corrected, they nevertheless give different values, presumably due to differences in interpolation method from the original 0.7° grid of ERA-Interim.~~

2.2 Climate model data

2.2.1 EURO-CORDEX and CMIP5

In order to examine the effect of dynamical downscaling for climate extremes, we make use of the EURO-CORDEX (Jacob et al. 2014) RCM simulations for the historical period over the European domain which are driven by lower resolution global scale coupled CMIP5 GCMs. The GCMs are forced by observed records of anthropogenic and natural forcings, such as greenhouse gases, anthropogenic aerosols, land use changes, solar variability and volcanic aerosols to allow comparability to historical records. For the most part the RCMs inherit the effects of these forcing agents from the GCMs, with the exception of greenhouse gases, which are prescribed. A comparison of the RCM simulations with their driving CMIP5 simulations allows us to identify any value added by regional high resolution. The EURO-CORDEX simulations are available at 0.11° and 0.44° (12.5 km and 50 km respectively), allowing an assessment of the difference that increased regional resolution brings. By examining the subset of GCM-RCM combinations that are common to both CORDEX resolutions along with their driving GCMs we can isolate the effects of changing resolution.

Daily precipitation (pr), daily maximum temperature (tasmax), and daily maximum surface wind speed (sfcWindmax) were taken from both CORDEX and CMIP5. The simulations used are shown in Table S1. These~~is~~ consist~~s~~ of 23 and 19 simulations for precipitation for the 0.44° and 0.11° simulations respectively, with 15 common to both categories with data also available from their driving GCMs (from now on referred to as “common to all” or “common subset”); 22 and 18 respectively for temperature, with 14 common to all, and 15 and 14 for wind with 6 that are common to all. We also extend the analysis to all other historical CMIP5 GCMs with the relevant variables, with 126 simulations from 41 GCMs for precipitation, 115 from 39 models for temperature and, 61 simulations from 28 models for wind. For wind, using 3 or 6 hourly data would have made results more comparable to the ~~observational~~ reanalysis wind datasets and across models (see above). However, such data were not available for the 0.44° CORDEX simulations, and very limited for ~~only three~~ CORDEX ~~simulations at~~ 0.11° ~~resolution which also had data for their driving GCMs, all three of which use the same RCM (RCA)~~. We therefore use the variable sfcWindmax (daily maximum surface wind speed) which was available for many models. This seems to mostly be based on model timestep wind speed, with a few exceptions (see figure S7) ~~The~~The –implications of this are discussed further in the results section.

2.2.2 UPSCALE simulations

In order to examine the benefits or otherwise of differences in resolution for a global model, we make use of simulations undertaken as part of the UPSCALE project (UK on PRACE: weather-resolving Simulations of Climate for globAL Environmental risk; Mizielinski et al. 2014). This consists of the atmosphere only version of the Hadley Centre Global Environment Model 3 (HadGEM3-A) run at three different resolutions: N96 (130 km), N216 (60 km) and N512 (25 km), all with 85 vertical levels for the period 1985-2011, with 5, 3 and 5 ensemble members respectively (or 3, 3 and 5 for wind data). The simulations are forced by observed records of greenhouse gases, aerosols, ozone, solar variability and volcanic forcings following the AMIP-II procedure (Taylor et al. 2000), but using the higher resolution OSTIA analysis for ~~and an alternative dataset for~~ sea surface temperatures (SSTs) and sea ice (Donlon et al. 2012). Very few parameters differ between the resolutions, enhancing the comparability of the three ensembles. We use daily precipitation data, daily maximum temperatures and 3-hourly wind (subsampled to 6-hourly).

2.2.3 Regridding

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 (Prein 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. ~~the resolution of the majority of the observational datasets used (E-OBS, ECEM and WFDEI) and is computationally feasible.~~ Some of the benefits of higher resolution may be lost by doing this, putting our results on the conservative side. Nevertheless~~However~~, sensitivity tests showed that results for MESAN did not change perceptibly by using a 0.5° grid ~~as~~ compared to a 0.1° ~~regular~~ grid. ~~(chosen to be close to the original 0.11° rotated grid).~~ We regrid the daily data, before the calculation of annual extreme indices.

Sensitivity of results to regridding technique was investigated ~~for precipitation and wind~~ for a number of models of different resolutions and compared with results based on using the original grids (Figure S1). For the coarser resolution models (e.g. HadCM3) results for precipitation extremes were ~~very~~ particularly sensitive to regridding technique, with much weaker extremes for some techniques e.g. distance-weighted average remapping and bilinear interpolation, with unrealistic artefacts in the spatial patterns for many methods. For high resolution models, regridding technique did not make much difference to results, although conservative remapping tended to dampen extreme precipitation, particularly for CORDEX 0.11. Overall the nearest neighbour method was chosen for precipitation for everything except CORDEX 0.11 -and MESAN since it gave results very close to using the original grid for all model resolutions, preserving the amplitude of extremes, and also having minimal artefacts when plotting spatial patterns of precipitation extremes. For going from high to lower resolution (e.g. 0.11° to 0.5°) nearest neighbour is less appropriate since information from only a subset of grid cells is incorporated. Therefore, bicubic remapping was used for CORDEX 0.11 and MESAN, which also replicated results using the original grid very well (Figure S1). Wind and temperature results were also somewhat sensitive to regridding technique, particularly for the coarser models. The above choices also seemed appropriate for these variables (nearest neighbour in most cases, but bicubic for CORDEX 0.11, MESCAN, ERA5 and DYNAD), both in terms of replicating return period results using the original grid, and retaining the blocky nature of the low resolution simulations in the spatial patterns.~~Whilst nearest neighbour may not be the best choice in regridding from high resolution to lower (e.g. for MESAN and CORDEX 0.11), since information from only a subset of grid cells is incorporated, results were the same when repeated using bicubic remapping. Results for wind for the coarser models were also sensitive to regridding technique; the nearest neighbour method was again chosen since it also performed well here, both in terms of minimising artefacts and replicating results using the original grid. For temperature, which tends to be more uniform over large areas, bilinear interpolation was used, since the choice of regridding technique is anticipated to be less important.~~

3 Methods

3.1 Extremes Indices

In order to examine extremes, we adopt indices based on the ETCCDI indices (Zhang et al. 2011). For precipitation these are the annual maximum daily precipitation (Rx1day) and the annual maximum consecutive 5-day total (Rx5day). For temperature we use the annual maximum daily maximum temperature (TXx) and the annual maximum consecutive 5-day mean of daily maximum temperature (TXx5day). Rx1day and TXx5day are presented in the figures, whilst the other indices are commented on in the text. For wind we use the annual maximum of daily maximum wind, which we refer to as (WindXx). This is based on sfcWindmax for the CMIP5 and CORDEX models, and on 6-hourly data for the UPSCALE simulations and the ~~observational~~ reanalysis wind datasets. These are therefore much ~~more rare~~rarer extremes than those based e.g. on the 95th or even 99th percentile which would happen on average 1 in 20 days and 1 in 100 days respectively. One drawback is that this makes robust statistics more challenging.

In order to examine how well the climate models simulate extremes and the differences between different resolutions, we first examine the spatial patterns of the climatological mean values of the indices and their biases with respect to observations. We then examine return period plots (see definitions below) for a number of regions for each index, which highlights any differences in the shape of the tails of the distribution of the extremes. The regions used are based on the PRUDENCE regions (Christenson and Christenson 2007) and the IPCC SREX regions (Seneviratne et al. 2012) and are shown in Figure S2 and Table S2. A subset of representative regions are presented here, with some comments about the others.

3.2 Return periods

In order to calculate regional return periods and return values we first sort the data into ascending order for each grid cell.~~, and then calculate the area weighted regional averages.~~ The return periods are calculated as N/k where N is the number of years of data, and k is the rank, with k=1 for the largest value. Return periods are therefore the inverse of the probability of an event exceeding a given value (called the "return value"). The area weighted regional average is made, for given return periods, over the associated return values. To avoid complications from missing data, grid cells in E-OBS with more than 5 days of missing data in any year during the period examined were masked for the whole period. Having one or more years missing would complicate the calculation of regional mean return periods and values. Models and observational datasets are masked to have the same spatial coverage, which is land only. A common time period, across the models being examined and the observations they are being compared to, are chosen to allow comparability. For the CMIP5 and CORDEX analysis 1970-2005 is used for temperature and precipitation and 1979-2005 for wind. For the UPSCALE runs we use 1985-2011 for temperature, and 1989-2010 for precipitation to allow comparisons with MESAN (1986-2011 is used for the analogue analysis, see below) and 1986-2011 for wind.

~~Return values and periods are also calculated for the "pooled ensemble". For each grid cell, all simulations of a certain type are combined into one long time series before being sorted into ascending order, and then regional means are calculated as above. The models are first bias adjusted by subtracting the difference between their climatology of the index in question and the climatology of the observations at a grid cell level. This adjustment avoids, for example, models with particularly hot extremes dominating the ends of the tails of the distributions and allows differences in the shapes of the distribution tails of different models to be compared more easily. Figure S3 shows the resulting spread of models across the distributions.~~

In order to allow comparability of results between the EURO-CORDEX ensembles at both resolutions and their driving CMIP5 GCMs, we picked a subset of models that are consistent across each category; that is the same GCM-RCM combinations are used across both the 0.11 and 0.44° CORDEX categories, and are compared to the CMIP5 model runs that were used to drive them (Table S1). We refer to these simulations as the "common subset" (see section 2.2.1). The only exception is that the EC-EARTH ensemble member "r3" was not available for download from ESGF, so r2 was substituted instead. Since more than one EURO-CORDEX RCM is driven by the same ensemble member of the same GCM, we repeat these GCMs when calculating the CMIP5 ensemble mean ~~and pooled results~~ for the common subset. For the CMIP5 vs CORDEX analysis we first bias adjust models before plotting return period curves in order to allow the shapes of the distributions to be compared more easily. We do this by subtracting the difference between the model climatology of the index in question and the climatology of the observations for each model at a grid cell level. We use E-OBS as the reference for temperature and precipitation, and MESCAN for wind. For the UPSCALE simulations, since the same version of the same model is used across each resolution, results can also be examined without bias adjusting the extremes climatology, and this provides some interesting insights.

Confidence intervals for observations are calculated using a bootstrapping method. If, for example, the analysis period was 1970-2005 (i.e. 36 years), 1000 random samples of 36 years from this period are chosen from the same observations~~/ simulation(s), ~~, allowing the same year to be chosen more than once per iteration. For each random sample, the chosen values are sorted for each grid cell and a regional average is calculated as above, effectively yielding 1000 return period curves per region. The 5th and 95th percentile of these values are then calculated to give the confidence intervals.

Finally, for the HadGEM3-A GCM simulations, a circulation analogue technique is used to split any differences in results according to resolution into upscaling (i.e. improved large scale circulation) and downscaling effects. This is described in section 4.3.

4 Results

4.1 The benefits of regional high resolution: EURO-CORDEX versus CMIP5

4.1.1 Temperature extremes

Figure 1 shows the spatial patterns of the climatological mean of TXx5day for the period 1970-2005 for E-OBS, and the multi-model means (MMM) of CMIP5, and CORDEX at both resolutions, along with their biases with respect to E-OBS. ~~The first two columns are based on a subset of CORDEX simulations that use the same GCM-RCM combinations at both resolutions, whilst the CMIP5 MMM is based only on the CMIP5 simulations that drive these RCMs, with repetition of the GCMs that drive more than one RCM. The last two columns are based on the mean of all available simulations for each category to check how representative the results based on the subset are of the whole ensembles. ~~The same general pattern can be seen in both the observations and the models, with hotter extremes in the south and cooler extremes in the north and over the mountains. At higher resolution the colder warm extremes over the Alps and Carpathians become more distinct. For the “common subset” the pattern of biases relative to E-OBS is similar for each model category with cold biases in the North and West and hot biases in the South-East. However, the hot biases over the mountains reduce with higher resolution since the model topography is higher. The cold bias over Scandinavia is also larger in CORDEX than in CMIP5. Biases using the whole ensemble are very similar as those for the CORDEX subset, although for CMIP5 the hot biases over the south-east, and over mountain ranges are stronger. Findings for TXx are similar, but hotter (not shown).

To give an idea of the level of consistency of results between models, results for individual models are shown in figure S34. Although the CMIP5 models agree on the general spatial pattern of temperature extremes, their absolute magnitudes vary considerably, although all are too hot over the Alps. There are also substantial differences between results from different RCMs, including those driven by the same GCM. Biases of individual RCMs do not appear systematically smaller than that of their driving GCM. Patterns are very similar for the same GCM-RCM chains at the both 12.5 and 50 km resolutions. Results for different ensemble members of the same GCM or GCM-RCM chain are very consistent, suggesting that the differences between models are not due to internal variability.

In order to assess the shape of the statistical distribution of temperature extremes, figure 2 (left column) shows return period against magnitude for TXx5day for CMIP5, CORDEX at both resolutions and E-OBS, for individual models (thin lines) and the pooled ensembles (circles) both for the common subset of models (darker circles) and all models (lighter circles) (see Methods). Results are shown for Northern, Central and Southern Europe, and are representative of the subregions. There is no obvious difference in the shape of the tails between CMIP5 and CORDEX, apart from marginally heavier tails for CORDEX 0.44 in central Europe. Agreement with E-OBS is good for the pooled ensemblemulti model median, although many individual ensemble members lie outside the range of the observational uncertainty, particularly on the heavy tailed side.

In summary, temperature extremes appear to be relatively insensitive to dynamical downscaling based on comparing CMIP5 to CORDEX at 0.11° and 0.44°, except over mountains where hot biases decrease with resolution.

4.1.2 Precipitation extremes

Now we consider precipitation extremes for CMIP5 compared to CORDEX. Figure 3 shows the climatological mean of Rx1day over the period 1970-2005 for E-OBS and the MMMs of CMIP5 and CORDEX at both resolutions, and their differencesbiases with respect to E-OBS. The left two columns show results for the "common subset" of simulations across the model categories, and the right two columns for all simulations. The heaviest annual maximum precipitation totals in E-OBS occur over the Alps and the western side of coastal mountain ranges, including western Norway and north-eastern Spain. A similar spatial pattern of precipitation distribution can be seen in the models, although totals are lower in CMIP5, and higher in CORDEX. CMIP5 exhibits is drier than E-OBSa dry bias over most of Europe, particularly over the areas of maximum observed precipitation in E-OBS (i.e. over or near mountains), whilst CORDEX is generally wetter than observedexhibits a general wet bias, particularly in these same locations, and at higher resolution. Results using the entire ensembles are very similar to using the common subset of simulations. Previous studies suggest that E-OBS underestimates precipitation extremes since it is not corrected for gauge undercatch and has a relatively low underlying station density (e.g. Prein and Gobiet 2017). Therefore, we also repeat results relative to the MESAN reanalysis (Figure S45) for the shorter period 1989-2005. MESAN uses a particularly high density of stations in France (see Data section). The climatology of Rx1day is wetter in MESAN than in E-OBS over most of Europe, most noticeably over the Alps and surrounding areas. This leads to the dry bias in CMIP5 appearing bigger, and the wet bias in CORDEX decreasing, although it is still present in the 0.11° simulations. Using regional-scale very high resolution datasets could improve agreement with the 0.11° simulations, since they tend to give heavier precipitation extremes (Prein and Gobiet 2017). Gauge undercatch willcould also contribute to the difference, particularly for precipitation extremes associated with strong winds and in snow dominated regions.

Figure S5~~6~~ shows results for individual models. Again, whilst models agree on the general pattern of precipitation extremes – i.e. wettest over mountains, there are considerable inter-model differences concerning the magnitude, particularly over complex orography. A number of CMIP5 models have too light extremes everywhere, but all underestimate precipitation extremes over mountainous regions to a greater or lesser extent. RCMs systematically simulate heavier precipitation extremes compared to their driving GCMs, particularly over mountains, and these extremes tend to become heavier when moving from 0.44° to 0.11° in most cases. Many of the RCMs have heavier precipitation extremes than seen in E-Obs~~show a heavy bias~~ over much of Europe at 0.44°, although this difference may disappear if compared to MESAN.~~, and~~ T~~t~~his difference~~bias~~ gets bigger at higher resolution and is largest over mountainous regions. Again results are very consistent between ensemble members of the same models.

Figure 2 (middle column) shows return period curves for Rx1day for Northern, Central and Southern Europe. There is a clear separation in the tails of the distribution according to resolution, with CMIP5 having the lightest tails, CORDEX 0.44 in the middle, and CORDEX 0.11 with the heaviest tails across all regions (including the subregions – not shown). Results using the common subset of models or the full ensembles are similar to each other.~~ In order to compare with observations, E-OBS should be compared to the thin lines for individual models rather than the pooled ensemble results, since pooling seems to affect the shape of the distribution, causing it to lie below that of the single models~~. EOBS tends to lie~~ at the heavy end of the CMIP5 range for southern Europe,~~ between CMIP5 and CORDEX 0.44 for central and southern Europe, and closer to CORDEX 0.44 in northern Europe. Using MESAN gives slightly heavier tails in central Europe (figure S6~~7~~) (particularly in France, where station density is highest –not shown) and more so in southern Europe, causing the best agreement to occur with CORDEX 0.44 everywhere. Results for Rx5day are similar, but with marginally less separation between the resolutions, whilst over Northern and Central Europe the best agreement with E-OBS happens at a slightly higher resolution than for Rx1day – i.e. either with CORDEX 0.44 or the lower end of the range of CORDEX 0.11 (not shown).

In summary, precipitation extremes are wetter and heavier tailed with higher resolution, especially over mountainous regions. CMIP5 has a dry bias, particularly over mountains, whilst CORDEX tends to be too wet relative to E-Obs , particularly at 0.11°, but results are sensitive to observational dataset used, with wet biases for CORDEX reducing when compared to the higher resolution MESAN dataset.

4.1.3 Wind Extremes

Finally, we examine annual maximum wind (WindXx). Figure 4 shows the multi model means of climatological mean annual maximum wind ~~over the period 1979-2005~~ for CMIP5 and CORDEX at 0.44° and 0.11°~~° for the common subset of simulations and for all simulations~~ compared to three ~~observational~~ reanalysis datasets. Note however that the model results are based on the annual maximum of the daily maximum~~ of~~ surface wind (variable “sfcWindmax”), whilst the ~~observations~~ reanalysis estimates are based on the annual maximum of 6-hourly data. As a sensitivity test, ~~F~~for CMIP5 models that had both sfcWindmax and 3-hourly data, we compared results using sfcWindmax, 3-hourly and 6-hourly data (Figure S7~~8~~). 6-hourly data tends to give lower values than using 3-hourly data or sfcWindmax since some events will be missed due to the lower sampling frequency. SfcWindmax appears to be mostly based on the model timestep, and gives higher wind speeds than using 3 or 6 hourly data, with some exceptions, e.g. the IPSL models and CMCC-CM where it gives lower values. This apparent difference in definition between models is a weakness of this analysis. Furthermore, since different models have different time steps, and the time step generally decreases with increased resolution, we might expect stronger winds with increased resolution purely due to the difference in sampling frequency. Whilst it could be argued that this makes the models not strictly comparable, being able to generate stronger winds due to a shorter time step could nevertheless be considered an inherent feature of higher resolution models.~~In addition, sfcWindmax does not always appear calculated the same way across all models- although it often gives higher wind speeds than using 3-hourly or 6-hourly data, the degree to which this is the case seems to depend on the model, and suggests a different sampling frequency, whilst for a few models sfcWindmax gives the lowest wind speeds.~~ It would have been cleaner to~~Ideally an analysis based on a~~ use a metric that is more ~~-~~consistentt~~t metric~~ across models, such as 3 hourly or 6 hourly wind speeds~~, would be performed~~. However, CORDEX at 0.44° does not have this data available, whilst CORDEX at 0.11° only has it for a small number of simulations, all of which are based on RCA, and only 3 of which have data for the driving GCM. Therefore, the reader is invited to interpret results with this caveat in mind. Model sfcWindmax estimates may also differ in terms of the treatment of surface roughness length and the method for calculating wind at 10m from wind at a higher level.

~~All three observational datasets are based on ERA-interim, but WFDEI and ECEM-noc are processed differently, whilst ECEM-wbc is bias corrected using a Weibull distribution based on the station based dataset HadISD (see Data section). It is notable that all three datasets give different results, particularly ECEM-wbc, despite all being based on ERA-Interim. WFDEI and ECEM-noc show the same overall pattern of annual maximum wind, with a belt of stronger winds running zonally across the middle of Europe, including particularly high wind speeds over the British Isles, with lower wind speeds to the north and south of this band. ECEM-noc has slightly faster wind speeds everywhere. ECEM-wbc is very different with very high maximum wind speeds over southern Europe and Scandinavia, which are areas with low wind speeds in the other two datasets. These differences are most apparent for wind extremes- the climatological means of 6-hourly wind speeds are shown in Figure S9. Although patchier in nature, mean wind in ECEM-wbc appears broadly similar to the other datasets. However, whilst the other two datasets have similar spatial patterns between mean and extreme wind, for ECEM-wbc the patterns are very different. One assumption is that the Weibull correction works well for mean wind, but is not suited to extremes, which are more sensitive to the parameters of the Weibull correction. It should also be noted that as a reanalysis, ERA-interim is itself model based, albeit with assimilation of observations.~~

Examining figure 4, the MESCAN and DYNAD reanalyses show strong extreme winds over the UK, the Norwegian mountains and the NW coastline of France through to Denmark. Relatively strong winds are also seen over the Spanish plateau, and a belt of strong winds running zonally across central Europe between slower winds to the North and South. The datasets differ in the magnitude of the winds, with MESAN having more contrast between areas of low and high wind. ERA5 has notably slower winds, particularly over mountaineous regions, but a similar overall zonal tripole pattern can be seen.

The CMIP5 driving model mean shows a similar overall pattern of WindXx as ~~WFDEI and ECEM-noc~~the reanalyses, with a pattern of weaker winds in the north and south, and a belt of stronger winds in the middle, but do not tend to have stronger winds over mountains like in DYNAD and MESCAN ~~, but with a lower magnitude than the observations~~. Using the whole CMIP5 ensemble gives slightly stronger extreme winds. Absolute magnitudes are not directly comparable to the ~~observational~~ reanalysis estimates, which would be expected to have slightly slower winds due to differences in sampling frequency.~~, but they are nevertheless broadly similar to WFDEI and ECEM-noc, but with too light winds in the central zonal belt.~~ The CORDEX multi model means show generally higher wind speeds than CMIP5, and capture~~a different spatial pattern, with~~ the high~~est~~ wind speeds along western coastlines and over mountainous terrain. Differences between the 0.11° and 0.44° runs appear small. Results for the common subset of simulations are very similar to those obtained from the complete CORDEX ensembles. Biases are not shown due to the difference in temporal resolution with respect to the ~~observations~~reanalyses.

Figure S8~~10~~ shows that there is a large variety between different models, particularly for CMIP5, but also according to RCM. CanESM2 and IPSL-CM5A-LR are notable outliers, and this may be related to the timestep of the wind data used to calculate sfcWindmax in these models. The zonal tripole pattern can be seen in a number of GCMs, as can stronger winds along western and Mediterranean coastlines, and lower wind speeds over the Alps. Spatial patterns for the RCMs are very RCM specific and relatively insensitive to driving GCM. All RCMs agree on higher winds over the British Isles and weaker winds over northern Europe, but notably the mountainous regions have either low or high wind speeds depending on the model, which must relate to how wind speed is calculated there - it can be imagined that the wind speed in a valley would be somewhat different to that at the top of a mountain. In terms of differences between the two resolutions of CORDEX, some RCMs show increased wind speeds with higher resolution e.g. RACMO, HIRHAM5, and others less so. Again, ensemble members of the same model give similar results.

Figure 2 (right column) shows the return period plots for WindXx for CMIP5 and both resolutions of CORDEX. All models are shifted to have the same climatology of annual maximum wind for each grid cell, which goes some way to adjusting for differences in sampling frequency, although there is evidence that the shape of the tails is also affected for some models (Figure S7~~8~~). The results for the common subset of CORDEX runs should at least be more directly comparable to each other, although the sampling frequency should still increase at higher resolution. The British Isles are shown instead of Northern Europe, since they are particularly affected by wind extremes, and for comparison with the results for the UPSCALE simulations, where this region shows distinctive results. The distribution of annual maximum sfcWindmax has heavier tails in CORDEX 0.11 compared to 0.44 which is in turn heavier than CMIP5, regardless of the subset of models used in calculating the multi-model median ~~creating the pooled ensemble~~ in almost all regions examined. ~~The CMIP5 r~~Exact values~~esult~~s are somewhat sensitive to the models included for some sub-regions (not shown). Results based on DYNAD and MESCAN tend to lie in between the two CORDEX resolutions, whilst CMIP5 is closest to ERA5. ~~The model results appear relatively consistent with~~

~~the WFDEI and ECEM-noc observations (note the different sampling frequency). ECEM-wbc is much heavier tailed in southern and Central Europe.~~

In summary, winds tend to be stronger, with heavier tails at higher resolution, with a large spread between models. ~~Observational~~ Reanalysis datasets give ~~very~~ fairly diverse results.

4.2 Global high resolution: UPSCALE

We now examine the benefits or otherwise of global high vs. standard resolution simulations for simulating climate extremes. Global high resolution may allow an improved representation of the large scale circulation that cannot be captured by regional models, which may in turn affect the representation of climate extremes. For this we examine the UPSCALE simulations (Mizielinski et al. 2014), which consist of a small ensemble of HadGEM3-A simulations at three different resolutions: 130km (N96), 60km (N216), and 25km (N512) (see Data section).

4.2.1 Temperature extremes

Figure 5 shows the ensemble mean climatological mean of TXx5day for the UPSCALE simulations over the period 1985-2011 at all three resolutions, and their biases relative to E-OBS. The same general pattern of hotter extremes in the south and colder in the north and over mountainous regions can be seen at all three resolutions, but temperature extremes are hotter at higher resolution in the south and east, and colder over mountains. The same pattern of biases is seen as for CORDEX and CMIP5 with cold biases in the north and hot in the south-east and over mountains. The mountain biases reduce with higher resolution, as the orography becomes better defined, whilst the hot bias in the SE and SW increases and the northern cold bias improves slightly. A coastal cold bias at low resolution disappears at higher resolution as the model land mask becomes more detailed~~, presumably because the ocean influence is carried further over land at low resolution in the large grid boxes~~. Note that the SSTs are prescribed and are the same for all simulations. Results for TXx are similar but hotter (not shown).

Figure 6 (left column) shows regional return period plots for TXx5day for the UPSCALE simulations. Results are a little less consistent across regions for UPSCALE compared to the CMIP5 vs CORDEX analysis, so we split Northern Europe into the British Isles and Scandinavia, and add the Alps, to better capture regional variations. ~~Again, the thin lines are individual simulations, and the circles are for results pooled across the ensemble members for each resolution separately~~. Since the ~~pooled~~ ensemble means~~s~~ are only based on one model, results are presented without adjusting according to the climatology of TXx5day, although bias adjusted results can be seen in Figure S9~~11~~ and allow differences in the shapes of the tails to be seen more clearly. TXx5day seems to be somewhat hotter with higher resolution over many~~ost~~ regions, although this is not always clear cut.~~,~~ The Alps are a~~with the~~ notable exception~~ of the Alps~~, where the higher elevations with higher resolution give rise to colder temperature extremes. There are notable biases relative to the observations, with the models being too cold in the north, especially at low resolution, whilst in the south the colder subset of models (N96, the lowest UPSCALE resolution) agree best with the observations. Over the Alps, again the low resolution simulations agree best with observations, with the warmest temperatures, but this will depend on the height of the meteorological stations. This apparent contradiction to the reduced orographic hot bias with resolution in figure 5 comes from the stronger cold bias of the surrounding areas at low resolution. Figure S9~~11~~ shows that differences between the shape of the tails with resolution are not systematic across regions and are mostly small~~in general there is not much difference between the shape of the tails with resolution, with only slightly heavier tails with increased resolution over the British Isles and Southern Europe~~. Agreement with E-OBS is~~ very~~ good everywhere. Results for TXx are similar.

In summary, hot biases of temperature extremes over mountains reduce with increased resolution for HadGEM3-A. Elsewhere extremes tend towards getting~~get~~ hotter with resolution, whilst the shapes of the statistical distributions are insensitive.

4.2.2 Precipitation extremes

For precipitation, Figure 7 shows the ensemble mean climatological mean of Rx1day for the period 1989-2010 for the three UPSCALE ensembles and their differences~~biases~~ relative to E-OBS and MESAN. The overall pattern of Rx1day is similar to that in E-OBS, with heaver precipitation extremes and finer spatial detail with increasing resolution over complex orography. All resolutions have bands of heavy precipitation either side of the Alps, but these move closer together as the Alps become better defined~~The N96 runs have an area of heavy precipitation stretching from France into Germany, whilst the N216 and N512 simulations show instead a pattern of heavy precipitation either side of the Alps, with a drier area in-between~~. All simulations are generally wetter than E-OBS across most of Europe, ~~A general wet bias can be seen at all resolutions over Europe,~~ whilst the dry bias over orography in the Alps, Southern Norway and Scottish Highlands reduces with resolution and a wet bias on the southern edge of the Alps and the coastal side of the Dinarie Alps in the Balkans appears as resolution increases. Comparing to MESAN instead of E-OBS, the general wet bias disappears, and the dry mountain bias over orography at low resolution increases. The differences between resolutions appear smaller than for the CMIP5 versus CORDEX analysis: all the UPSCALE simulations look most similar to CORDEX at 0.44°. However, UPSCALE does not reach as fine a resolution as CORDEX at 0.11° (25 km vs 12.5 km), and CMIP5 is on average slightly coarser than the N96 simulations. 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 (Skamarock 2004; Klavar et al. 2020).~~ a model's nominal resolution does not always accurately reflect the spatial scales that it can represent~~. Results are similar for Rx5day (not shown).

Figure 6 (middle column) shows the return period plots for Rx1day for the three resolutions of UPSCALE ensembles. Slightly heavier precipitation extremes are found at higher resolution in all the regions shown (exceptions are France and Mid Europe- not shown), although differences are small, they are more obvious in southern Europe and especially in the Alps. Figure S9~~11~~ shows that there is not much difference in the shape of the tails for most regions, although there are very slightly heavier tails at higher resolution for southern Europe (more so in the

Mediterranean sub region- not shown) and more obvious differences over the Alps in the same direction, both of which are regions where convective precipitation is important. E-OBS tends to lie just below the model simulations for most regions (Figure 6~~ – compare with the thin coloured lines~~), although it agrees with the models for the British Isles, and is between the low and medium resolution simulations over the Alps. MESAN gives higher values for observed Rx1day which improves agreement in regions where E-OBS lay below the models, and causes a higher resolution subset to agree better in the other regions (Figure 6). For the bias adjusted ~~versions~~ curves E-OBS tends to lie just on the lower end of the ensemble for most regions, whilst MESAN gives slightly heavier tails and tends to improve agreement with models (Figure S9~~11~~). Results for Rx5day are broadly similar (except that both sets of observations lie above all the models for the British Isles).

In summary, precipitation extremes are somewhat wetter and heavier tailed with increasing resolution mostly in southern Europe and the Alps for HadGEM3-A. Dry orographic biases decrease with resolution but wet biases appear in the south next to mountain ranges instead.

4.2.3 Wind extremes

For wind extremes, Figure 8 shows the spatial patterns of climatological mean annual maximum wind based on 6-hourly data for UPSCALE and the same for three ~~observational~~ reanalyses~~datasets.~~. In this case the models and ~~observations~~ reanalyses are directly comparable since they share the same temporal resolution. The spatial patterns are similar for the three different model resolutions, with the highest winds over the British Isles and coastal regions, lower wind speeds over the Alps, and the zonal tripole pattern described above~~, although this does not extend as far east as in the observations (i.e. ECEM noc and WFDEI).~~ The main differences are that the lower resolution model (N96) has stronger winds around the British Isles and western coastlines. This is likely because the larger coastal grid boxes overlap more with the ocean, which tends to have higher wind speeds, or due to differences in the model land mask itself with resolution~~, presumably because the larger grid boxes overlap more with the ocean, which tends to have higher wind speeds~~. The wind speeds at higher resolution are a little stronger overall, most obviously in the central European zonal belt, and over the Alps and Norwegian mountains. All resolutions show stronger winds than ERA5 over most of Europe. Compared to MESCAN winds are too weak in the northern and southern Europe, particularly over mountainous regions, and a little too strong inbetween. Relative to DYNAD the pattern of differences is similar as for MESCAN, but with stronger negative differences over the Norwegian mountains and positive differences in other parts of Northern Europe. There are positive coastal biases relative to all reanalyses that reduce with increased resolution.

~~As noted before, the observational estimates vary significantly and therefore the biases depend on the observational dataset used, i.e. extreme winds are slightly weak compared to ECEM noc over much of Europe; compared to WFDEI, winds are too strong in the north and south, and too weak in the east; and compared to ECEM wbc, winds are far too weak in the north and south and too strong in-between.~~

Figure 6 (right column) shows the return period plots for some example regions for annual maximum wind for the UPSCALE simulations, without shifting the climatology. Over all regions examined (except the Mediterranean- not shown), the N512 simulations have stronger winds than the N216 simulations. The position of the curve for N96 is strongly related to how much coastline there is relative to land area per region, e.g. with faster winds than the other simulations over the British Isles and southern Europe, but relatively slower winds over central Europe, and particularly over the Alps (not shown) ~~most regions the strongest extreme winds are found at the highest resolution, with the exception of the British Isles (and the Iberian Peninsula- not shown) where the low resolution models have the strongest winds. This is likely related to the large coastal grid boxes overlapping windy ocean areas as discussed above~~. ~~As noted above, ~~T~~t~~here are fairly large differences between ~~observational ~~reanalysis estimates, with ERA5 always having the slowest winds, and the model simulations tending to lie between ERA5 and the other two reanalyses for most regions. ~~with ECEM-wbc having considerably higher values and heavier tails than the other two datasets and models over most regions, except the British Isles. ECEM-noc tends to agree best with the model simulations, whilst WFDEI tends to lie at the lower end of the model range or underneath. ~~For the bias adjusted versions of the return period plots (Figure S9~~11~~), differences in the shapes of the tails with resolution are generally small, although with marginally heavier tails with increasing resolution over a number of regions (not all are shown). The shape of the tails is generally close to the reanalysis estimates. ~~Agreement of the shape of the tails with ECEM-noc and WFDEI is good.~~

In summary winds are slightly stronger and heavier tailed at higher resolution in HadGEM3-A, except over coastal areas where large coastal grid boxes at low resolution bring strong ocean winds further over land.

4.3 Circulation Analogues

For the global model results, any differences in the representation of extremes according to resolution could come from either upscaling or downscaling effects. Upscaling effects could include a better representation of the large scale circulation, whilst downscaling allows a better representation of small scale processes, such as convection, and an improved representation of orography and coastlines. In order to investigate which of these effects leads to the differences between the low (N96) and high resolution (N512) HadGEM3-A simulations, we employ a circulation analogue technique (e.g. Vautard et al., 2016), which is frequently used in attribution studies (see e.g. Stott et al., 2016; Cattiaux et al., 2010). The idea is to determine whether the simulation of climate extremes changes between the two resolutions if both were to have the same large scale circulation –i.e. isolating the downscaling effect, or conversely whether circulation differences explain any differences in extreme events whilst circulation-variable (e.g. precipitation) relationships stay the same –i.e. the upscaling effect.

For each day in the lower resolution simulations we pick the nearest circulation analogue from anywhere in the higher resolution simulations, providing it happens at the right time of year (i.e. within a 30-day window centred on the day of the year in question). We then record the associated temperature, precipitation and wind values from the higher resolution simulations to make a “*u*-chronic” dataset (e.g. Jézéquel, et al. 2018) that contains data from the high resolution simulations but follows the daily sequence of circulation patterns from the low resolution models. We then repeat the analysis of return periods and value as above. We also do the reverse (find analogues for the N512 circulation in the N96 ensemble and record the N96 temperature). Since results using analogues are not directly comparable to the original results, due to lack of exact analogue match, we also perform "self-analogues" -i.e. finding circulation analogues for the N96 simulations within the N96 ensemble, (excluding the same year from the same ensemble member) and creating a u-chronic time series, and the same for the N512 ensemble). Comparing the resulting return period curves tells us about the contribution of large-scale circulation and downscaling to differences in extremes between the two resolutions. For example, comparing the N96 self-analogue return curve to the version based on N512 circulation but with N96 precipitation shows us the contribution of any differences in the large scale circulation between the resolutions i.e. the upscaling effect. Comparing the N96 self-analogue to the version based on N96 circulation with N512 precipitation shows us the downscaling effect – i.e. any difference between the relationship between the large scale circulation and precipitation.

Analogues are defined using geopotential height at 500 hpa, since this avoids complications relating to surface heat lows associated with heat waves in anticyclonic conditions that occur in summer, whilst also avoiding incomplete data due to mountain ranges. Geopotential height is regridded to a 2° grid using bilinear interpolation. This choice ensures that we are comparing analogues with the same resolution and do not penalise small-scale differences. Similarity between circulation states is quantified using pattern correlation, which is not affected by trends in geopotential height with global warming ~~is calculated using the Euclidean distance~~. For precipitation and wind the European domain used is -16 to 44° E and 34 to 72° N (roughly the same as the domain plotted in the map-based figures). For temperature, a larger domain is used, since the history and trajectory of air masses are~~is~~ important for temperature extremes. This domain is loosely based on the domain used by Cattiaux et al. (2010) and extends over the N. Atlantic as well as Europe, (-62 to 44°E and 24 to 80° N). However, results are very similar if the smaller domain is used (not shown). For the 5-day variables (Rx5day and TXx5day); daily geopotential height, precipitation and temperature datasets were smoothed using a 5-day running mean first, and then analogues were calculated, and the u-chronic datasets constructed. We also tried doing the 5-day means last rather than first, i.e. calculating analogues using daily data and smoothing the u-chronic dataset. The relationship between the different curves was largely consistent between the two techniques, but absolute values differed and the shape of the distributions changed a little. Results presented here are based on the first technique since it replicates better the autocorrelation structure of the original analysis.

Figure 9 shows the results of the analogue analysis. The blue curves show the results for the N512 self-analogues, grey represents the N96 self-analogues, red represents results using the circulation patterns from the N96 runs but with the N512 circulation-variable relationships, and green indicates N512 circulation with N96 circulation-variable relationships. The difference between the blue and red curves (or the grey and green curves) shows the contribution from differences in the large scale circulation with resolution, whilst the difference between the blue and green curves (or the red and grey curves) indicates the downscaling effect.

For TXx5day downscaling effects are dominant over regions that have a clear difference between resolutions, although circulation differences also have a small effect in some regions such as the British Isles (Figure 9). For Rx1day the different curves are very close together for most regions, making it difficult to discern the relative contributions from upscaling and downscaling. However, it generally seems to be downscaling effects that are the most important, and this can be seen more clearly for the Alps and Southern Europe where there are larger differences with resolution. Interestingly, these are regions where convective precipitation is particularly important for precipitation extremes. For wind extremes downscaling effects also dominate~~, but the large scale circulation also plays a role in Scandinavia~~. Results for TXx and Rx5day are very similar to those for TXx5day and Rx1day respectively (not shown).

Also shown, using ~~thick solid~~ dashed lines, are the original ~~pooled~~ ensemble mean results without using analogues. By comparing these with the self-analogue results ~~(i.e. compare the blue line with the blue circles for N512, and the grey line with the grey circles for N96),~~ we can see how successful the analogue technique is in recreating the original distributions. The self-analogue results tend to be close to~~, but slightly below~~ the original results for wind and Rx1day, but above them for Tx5day.~~, with a slight difference in the shape of the tail at the far right for Rx1day. This can be explained by the fact that the analogues are not perfect, and since the circulation patterns associated with climate extremes are rare, the nearest analogues are likely to represent slightly less severe events. The original results are beneath the analogue results and a different shape for TXx5day,~~ This effect seems to be enhanced by the 5 day averaging, but is still present ~~which seems to be associated with the 5-day averaging, and is much less marked~~ for TXx (not shown). Undertaking the 5-day averaging last rather than first (see Methods) shifts analogue results downwards, underneath the original curves, but otherwise gives the same results (not shown). ~~The same~~ A similar phenomenon is seen for Rx5day (not shown).

In summary, for all three types of extreme events, downscaling effects appear to dominate the differences seen between the 130km and 25km HadGEM3-A simulations. This suggests that at least for this model, any large scale circulation differences obtained with global high resolution do not affect the statistics of these extreme events much.

~~5 Summary and Discussion~~

~~5.1 Summary~~

Comment [c1]: combined with below

~~We evaluated climate model simulations of temperature, precipitation and wind extremes over Europe, addressing three questions: 1) The benefits of dynamical downscaling using regional climate models by comparing EURO-CORDEX simulations at two resolutions (12.5 and 50 km) to their driving coarser resolution CMIP5 models; 2) The benefits of increased resolution for global models by comparing HadGEM3-A simulations at three resolutions (130, 60 and 25 km); and 3) whether any differences according to resolution in the global model comes from differences in the large scale circulation (upscaling) or the representation of small scale processes, and features (downscaling) using a circulation analogue method.~~

For temperature extremes, increased resolution did not make much difference to results for the CORDEX vs CMIP5 analysis, both in terms of the shapes of the distributions, which all agreed well with observations, or in terms of biases, apart from reducing hot biases over mountains. This reduction in orographic bias with increased resolution was also seen in the HadGEM3-A GCM simulations, along with a general increase in magnitude of hot extremes elsewhere, which reduces biases in the north, and increases them in the south. Overall the benefits of increasing resolution were limited, or region dependent.

Precipitation extremes were more sensitive to resolution, particularly in the CMIP5 vs CORDEX analysis, with heavier tails at higher resolution across all regions. Spatially, CMIP5 shows a general dry bias compared to E-OBS, particularly over mountainous regions, whilst CORDEX shows the opposite, with increasing wet biases at 0.11° compared to 0.44°, which appears to be systematic across models. The higher resolution MESAN reanalysis gave wetter extremes and heavier tails than E-OBS, agreeing best with the 0.44° resolution CORDEX simulations, highlighting the importance of the choice of observational dataset. Differences according to resolution were smaller for the global scale HadGEM3-A simulations, although these span a smaller range of resolutions. Differences were most obvious in southern regions and the Alps, with heavier tails and wetter extremes at higher resolution. Dry biases over orography decreased with increasing resolution; however, wet biases next to some mountain ranges in the south emerge. Return period curves for HadGEM3-A tended to agree well with MESAN, but were too wet compared to E-OBS.

For wind extremes, higher resolution gave stronger winds and heavier tails for most regions for both the CORDEX vs CMIP5 analysis and to a lesser extent for HadGEM3-A. The largest differences were between CMIP5 and CORDEX at 0.44°, with less difference between the two resolutions of CORDEX. Differences between observational estimates made model evaluation difficult, whilst inconsistencies in the way daily maximum wind is calculated in different models were also an issue.

The circulation analogue analysis suggested that for the global scale HadGEM3-A simulations, differences according to resolution for all three phenomena were dominated by downscaling effects, with only small contributions from differences in the large-scale circulation.

5.2 Discussion and Conclusions

We evaluated climate model simulations of temperature, precipitation and wind extremes over Europe, addressing three questions: 1) The benefits of dynamical downscaling using regional climate models by comparing EURO-CORDEX simulations at two resolutions (12.5 and 50 km) to their driving coarser resolution CMIP5 models; 2) The benefits of increased resolution for global models by comparing HadGEM3-A simulations -at three resolutions (130, 60 and 25 km; referred to as the "UPSCALE" simulations); and 3) whether any differences according to resolution in the global model comes from differences in the large scale circulation (upscaling) or the representation of small scale processes, and features (downscaling) using a circulation analogue method.

~~For temperature extremes, our results imply that increased resolution in both regional and global models is of limited benefit at the resolution range considered here, except in reducing the hot bias over mountainous areas. In particular, for resolutions used in the UPSCALE experiments, we do not find strong upscaling nor downscaling effects.~~ For temperature extremes, increased resolution did not make much difference to results for the CORDEX vs CMIP5 analysis, both in terms of the shapes of the distributions, which all agreed well with observations, or in terms of biases, apart from reducing hot biases over mountains. These findings agree with Vautard et al. (2013)~~ for regional models~~, who find limited benefits in simulating various aspects of heatwaves between the 0.44° and 0.11° versions of the EURO-CORDEX models. This reduction in orographic bias with increased resolution was also seen in the HadGEM3-A GCM simulations, along with a general tendency towards hotter extremes elsewhere, which reduces biases in the north, and increases them in the south. Overall the benefits of increasing resolution were limited, or region dependent. However, our results for the global model analysis are based on only one model and the new model simulations and analyses being generated as part of the PRIMAVERA and HighResMIP projects (https://www.primavera-h2020.eu/; Roberts et al. 2018; Haarsma et al. 2016) will be very useful for determining how representative our results for HadGEM3-A are of other GCMs. For instance, improvements in the simulation of summer blocking, which can be involved in heatwave generation is very model dependent (Scheimann et al. 2014). Furthermore, Cattiaux et al. (2013) find that the frequency, intensity and duration of summer heatwaves improve in the IPSL model with resolution, associated with a better representation of the large scale circulation. In addition, here we examine only one aspect of heat waves (intensity), and it could be that results are different for others aspects, such as frequency, duration and timing.

Precipitation extremes were more sensitive to resolution, particularly in the CMIP5 vs CORDEX analysis, with heavier tails at higher resolution across all regions. Spatially, CMIP5 shows a general dry bias compared to E-OBS, particularly over mountainous regions, whilst CORDEX shows the opposite, with increasing wet differences at 0.11° compared to 0.44°, which appears to be systematic across models. This is consistent with results for mean precipitation in EURO-CORDEX in Kotlarski et al. (2014). The higher resolution MESAN reanalysis gave wetter extremes and heavier tails than E-OBS, agreeing best with the 0.44° resolution CORDEX simulations~~For precipitation extremes, we found that the CMI~~. ~~P5 models were too dry whilst CORDEX was a little too wet at 0.44° and more so at 0.11° when compared to E-OBS. This was particularly the case over complex orography. This is consistent with results for mean precipitation in EURO-CORDEX in Kotlarski et al. (2014). However, our results depend on the observational dataset compared against, with MESAN giving heavier extremes than E-OBS and agreeing reasonably well with the 0.44° simulations.~~ Other studies suggest that country-scale higher resolution precipitation datasets give heavier precipitation extremes still, which may agree best with the 0.11° simulations. Similarly, for mean precipitation, Prein and Gobeit (2017) find that RCM biases are a similar size to the differences between different observational estimates. For extreme precipitation, Prein et al (2016) and Torma et al (2015) find that various aspects (biases, frequency-intensity distributions, spatial patterns) of mean and extreme precipitation improve in EURO-CORDEX at 0.11° compared to 0.44° when compared to such datasets for Europe and the Alps respectively. Prein et al (2016) ascribe this mostly to the better representation of orography at higher resolution, but also the ability to capture the larger scales of convection. However, aside from improved spatial patterns Casanueva et al (2016) found only limited evidence for improvements in precipitation intensity, frequency and derived indicators over the Alps and Spain with resolution in EURO-CORDEX. ~~However, s~~Some of the differences with resolution in our results may also be explained by parameterisation schemes that tend to be tuned to one resolution and can behave sub-optimally at others.

For the UPSCALE global simulations, there was less difference with resolution, with the biggest differences in southern regions or over or near mountains, with heavier tails and wetter extremes at higher resolution. This reduced dry biases over orography, but wet biases next to some mountain ranges in the south emerged instead~~.~~ However, these simulations span a narrower range of resolutions, i.e. not reaching the same high resolutions as CORDEX 0.11°, but also not as coarse as some CMIP5 models. Other global model studies also tend to find an increase in precipitation extremes with increased resolution for Europe, which is continent-wide in summer, and concentrated in mountainous regions in winter (Volosciuk et al. 2015; Wehner et al. 2014). This sometimes improves agreement with observations (e.g. Kopparla et al. 2013; Wehner et al. 2014 for winter), but can overestimates summer extreme precipitation if parameterisation schemes are not retuned (Wehner et al. 2014).

For wind extremes, higher resolution gave stronger winds and heavier tails for most regions for both the CORDEX vs CMIP5 analysis and to a lesser extent for HadGEM3-A, except for regions dominated by coasts for the latter, where large coastal grid boxes at lower resolution brought strong ocean winds further over land. Stronger winds with higher resolution is also found in~~For wind extremes, our findings of stronger winds and heavier tails with increased resolution are consistent with~~ previous studies (e.g. Pryor et al. 2012;~~ Champion et al. 2011;~~ Kunz et al. 2010). The largest differences we found were between CMIP5 and CORDEX at 0.44°, with less difference between the two resolutions of CORDEX. Differences between reanalysis based estimates made model evaluation difficult, whilst inconsistencies in the way daily maximum wind is calculated in different models were also an issue.
~~However, observational issues made model evaluation difficult.~~

The results of the circulation analogue analysis on the HadGEM3-A GCM simulations suggested that downscaling effects were the dominant cause of differences with resolution for all three phenomena, with limited effects of any differences in the representation of the large scale circulation. If this result also applied to other GCMs, it would suggest that dynamical downscaling with more economical limited area models would be a better strategy for simulating European extreme events, whilst GCM efforts could focus on other aspects such as multiple members or multi-physics ensembles. However, we cannot reach this conclusion based solely on this analysis, since we examine only a single model, which may not be representative of other models, and because the range of resolutions considered may be too narrow. Furthermore, a number of studies do find improvements in the large-scale circulation with resolution, including for extra-tropical cyclones and storm tracks (Colle et al. 2013; Jung et al 2006; 2012, Zappa et al. 2013), Euro-Atlantic weather regimes (Dawson et al. 2012; 2015; Cattiaux et al. 2013) and blocking

(Jung et al. 2012, Anstey et al. 2013; Matsueda et al. 2009, Berckmans et al 2013; Scheimann et al. 2014; Davini et al 2017a; 2017b; see also Introduction). Interestingly, Scheimann et al. (2017) find improvements in Euro-Atlantic blocking with resolution in all seasons in the same HadGEM3-A simulations as we analyse here. However, the net effects on extremes, given all uncertainties, was not explicitly investigated. Our study does not seem to be able to discern such effects. Other studies suggest that benefits from upscaling may require convective permitting simulations (Hart et al. 2018).

Overall our results suggest that whether or not increased resolution is beneficial for the simulation of extreme events over Europe depends on the event being considered. Benefits appear limited for heatwaves, whereas wind extremes and particularly precipitation extremes are more sensitive. We do not find any particular advantage in using a global high resolution model compared to regional dynamical downscaling, with the caveats that this investigation needs to be extended to other GCMs, and a wider range of resolutions should be investigated.

In order to fully address the question of the benefits of increased resolution for European climate extremes, a number of aspects remain to be investigated. Firstly, the analysis could be widened to other types of extremes, for example, sea level rise and storm surge, or other aspects of extremes could be considered e.g. timing, frequency and duration of events. The global simulations we investigated were atmosphere-only, and the role of increased ocean resolution and also vertical resolution and model top height should be considered. Finally, we assume that better historical performance translates into more accurate future projections. Lhotka et al. (2018) find low sensitivity of heatwave projections to resolution in EURO-CORDEX RCMs. However, Van Haren et al. (2015b) find stronger future summer drying and heating in central Europe with increased resolution in the EC-Earth GCM due to differences in atmospheric circulation. Concerning precipitation, future projections for large scale and seasonal mean precipitation are consistent between large scale regional and convective permitting models, whereaswhilst there is evidence that summer daily and sub-daily intensities increase more in the future in convection permitting models (Kendon et al. 2014; 2017; Ban et al. 2015; Kendon et al. 2014). For wind, Willison et al. (2015) find a larger response of the North Atlantic storm track to global warming with higher resolution in the regional WRF model. Furthermore, Baker et al. (2019) find that in winter the polar jet, storm tracks and associated precipitation shift further North over the Euro-Atlantic region the future with increased resolution in the same HadGEM3-A set up as used here. The sensitivity of projections to resolution nevertheless remains an area that needs further research.

Data and code availability

The CMIP5 and CORDEX data used for this analysis are available from the Earth System Grid Federation portals, and are detailed in Table S1. The HadGEM3-A UPSCALE simulations are available from the CEDA-JASMIN platform. E-OBS can be downloaded here https://www.ecad.eu/download/ensembles/download.php, MESAN is available here http://exporter.nsc.liu.se/620eed0cb2c74c859f7d6db81742e114/, ERA5 and MESCAN are available from the Copernicus Climate Data Store https://cds.climate.copernicus.eu, whilst DYNAD winds are available from Tomas Landelius at SMHI.access to WFDEI is detailed here http://www.eu-watch.org/gfx_content/documents/README-WFDEI%20(v2016).pdf and ECEM data are available from the Copernicus Climate Data Store https://cds.climate.copernicus.eu.

Author contributions

CI, RV and SJ conceptualised the study, CI carried out the analysis and wrote the manuscript, JS managed the CRECP project together with CH and BE, and all co-authors were involved in discussions to prepare the study and helped improve the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This work is published in the name of the European Commission, with funding from the European Union through the Copernicus Climate Change Service project C3S_34a Lot 3 (Copernicus Roadmap for European Climate Projections). The Commission is not responsible for any use that many be made of the information contained. We acknowledge the WCRP's Working Group on Regional Climate, and the Working Group on Coupled Modelling - the coordinating body of CORDEX and the panel responsible for CMIP5 respectively. We thank the climate modelling groups for producing and making available the model output listed in Supplementary Table 1, which is available at http://pcmdi9.llnl.gov. For CMIP, the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We thank the modelling team that produced the UPSCALE simulations, and acknowlegdge the JASMIN and IPSL mesocentre computing clusters on which this analysis was performed. We thank Tomas Landelius from SMHI for making the DYNAD wind data available. We also acknowledge helpful input from the CRECP project scientific advisory board and useful discussions with UK Met Office Scientists, in particular Malcolm Roberts and Carol McSweeney.

References

Anstey, J. A., Davini, P., Gray, L. J., Woollings, T. J., Butchart, N., Cagnazzo, C., Christiansen, B., Hardiman, S. C., Osprey, S. M. and Yang, S.: Multi-model analysis of Northern Hemisphere winter blocking: Model biases and the role of resolution, J. Geophys. Res. Atmos., 118, 3956–3971, doi: 10.1002/jgrd.50231, 2013.

Baker, A.J., Schiemann, R., Hodges, K. I., Demory, M., Mizielinski, M. S., Roberts, M. J., Shaffrey, L. C., Strachan, J. and Vidale, P. L.: Enhanced Climate Change Response of Wintertime North Atlantic Circulation, Cyclonic

Activity, and Precipitation in a 25-km-Resolution Global Atmospheric Model. J. Climate, 32, 7763–7781, https://doi-org.ezproxy.is.ed.ac.uk/10.1175/JCLI-D-19-0054.1, 2019

Ban, N., Schmidli, J. and Schär, C.: Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster?, Geophys. Res. Lett., 42, 1165–1172, doi: 10.1002/2014GL062588, 2015.

Bazile, E., Abida, R., Verrelle, A., Le Moigne, P. and Szczypta, C. : Report for the 55years MESCAN-SURFEX re-analysis, deliverable D2.8of the UERRA project, pp. 22, available from: http://www.uerra.eu/publications/deliverable-reports.html, 2017

Berckmans, J., Woollings, T., Demory, M. E., Vidale, P.-L. and Roberts, M.: Atmospheric blocking in a high resolution climate model: influences of mean state, orography and eddy forcing, Atmos. Sci. Lett., 14, 34–40, doi:10.1002/asl2.412, 2013.

Casanueva, A., Kotlarski, S., Herrera, S., Fernández, J., Gutiérrez, J.M., Boberg, F., Colette, A., Christensen, O. B., Goergen, K., Jacob, D., Keuler, K., Nikulin, G., Teichmann C. and Vautard, R.: Daily precipitation statistics in a EURO-CORDEX RCM ensemble: added value of raw and bias-corrected high-resolution simulations, Clim Dynam, **47**,719-737. https://doi.org/10.1007/s00382-015-2865-x, 2016

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., and Codron, F.: Winter 2010 in Europe: A cold extreme in a warming climate, Geophys. Res. Lett, 37, L20704, doi: 10.1029/2010GL044613, 2010.

Cattiaux, J., Quesada, B., Arakélian, A., Codron, F., Vautard, R, Yiou, P.: North-Atlantic dynamics and European temperature extremes in the IPSL model: sensitivity to atmospheric resolution, Clim. Dynam., 40, 2293–2310, doi:10.1007/s00382-012-1529-3, 2013.

~~Champion, A. J., Hodges, K. I., Bengtsson, L. O., Keenlyside, N. S. and Esch, M.: Impact of increasing resolution and a warmer climate on extreme weather from Northern Hemisphere extratropical cyclones, Tellus A, 63, 893-906, doi:10.1111/j.1600-0870.2011.00538.x, 2011.~~

Christensen, J. H. and Christensen, O. B.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic Change, 81, 7–30, doi: 10.1007/s10584-006-9210-7, 2007.

Colle, B. A., Zhang, Z., Lombardo, K., Liu, P., Chang, E. and Zhang, M.: Historical evaluation and future prediction in Eastern North America and western Atlantic extratropical cyclones in the CMIP5 models during the cool season, J. Climate., 26, 882–903, doi: 10.1175/JCLI-D-12-00498.1, 2013.

Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., Ban, N., Bastin, S., Belda, M., Belusic, D., Caldas-Alvarez, A., Margarida Cardos, R., Davolio, S., Dobler, A., Fernandez, J., Fita Borrell, L., Fumiere, Q., Giorgi, F., Goergen, K., Guettler, I., Halenka, T., Heinzeller, D., Hodnebrog, Ø., Jacob, D., Kartsios, S., Katragko, E., Kendon, E., Khodayar, S., Kunstmann, H., Knist, S., Lavín, A., Lind, P., Lorenz, T., Maraun, D., Marelle, L., van Meijgaard, E., Milovac, J., Myhre, G., Panitz, H.-J., Piazza, M., Raffa, M., Raub, T., Rockel, B., Schär, C., Sieck, K., Soares, P. M. M., Somot, S., Srnec, L., Stocchi, P., Tölle, M., Truhetz, H., Vautard, R., de Vries, H. and Warrach-Sagi, K.: A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean, Clim. Dynam., 1-32, https://doi.org/10.1007/s00382-018-4521-8, 2018.

Dahlgren, P., Landelius, T., Kållberg, P. and Gollvik, S., A high-resolution regional reanalysis for Europe. Part 1: Three-dimensional reanalysis with the regional HIgh-Resolution Limited-Area Model (HIRLAM), Q.J.R. Meteorol. Soc., 142, 2119-2131, doi:10.1002/qj.2807, 2016

Davini, P., Corti, S., D'Andrea, F., Rivière, G., and von Hardenberg, J.: Improved Winter European Atmospheric Blocking Frequencies in High-Resolution Global Climate Simulations. J. Adv. Model. Earth Syst., 9, 2615–2634, https://doi.org/10.1002/2017MS001082, 2017a.

Davini, P., von Hardenberg, J., Corti, S., Christensen, H. M., Juricke, S., Subramanian, A., Watson, P. A. G., Weisheimer, A., and Palmer, T. N.: Climate SPHINX: evaluating the impact of resolution and stochastic physics parameterisations in the EC-Earth global climate model, Geosci. Model Dev., 10, 1383-1402, https://doi.org/10.5194/gmd-10-1383-2017, 2017b.

Dawson, A. and Palmer, T. N.: Simulating weather regimes: impact of model resolution and stochastic parameterization. Clim Dynam, 44, 2177-2193, https://doi.org/10.1007/s00382-014-2238-x, 2015.

Dawson, A., Palmer, T. N., and Corti, S.: Simulating regime structures in weather and climate prediction models, Geophys. Res. Lett, 39, L21805, https://doi.org/10.1029/2012GL053284, 2012.

Dee, D. P., Uppala, S. M., Simmons, A. J, Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J., Park, B., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Demory, M. E., Vidale, P. L., Roberts, M. J., Berrisford, P., Strachan, J., Schiemann, R., and Mizielinski, M. S.: The role of horizontal resolution in simulating drivers of the global hydrological cycle, Clim. Dynam., 42, 2201–2225, https://doi.org/10.1007/s00382-013-1924-4, 2014.

Donat M. G., Leckebusch G. C., Wild S., Ulbrich U.: Benefits and limitations of regional multi-model ensembles for storm loss estimations, Clim. Res., 44, 211-225. https://doi.org/10.3354/cr00891, 2010.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E.,and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, Remote Sens. Environ., 116, 140–158, doi:10.1016/j.rse.2010.10.017, 2012.

~~Dunn, R. J. H., Willett, K. M., Thorne, P., Woolley, E. V., Durre, I., Dai, A., Parker, D. E., Vose, R. S.: HadISD: A Quality Controlled global synoptic report database for selected variables at long-term stations from 1973-2011, Clim. Past, 8, 1649-1679, doi: 10.5194/cp-8-1649-2012, 2012.~~

Førland, E. and Institutt, N. M.: Manual for Operational Correction of Nordic Precipitation Data. Norwegian Meteorological Institute., 1996.

Giorgi F., Jones C., Asrar G. R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bull., 58:175–183, 2009.

Goodison, B. E., Louie, P. Y. and Yang, D.: The WMO solid precipitation measurement intercomparison. World Meteorological Organization-Publications-WMO TD, Report No. 67, 65–70, 1997.

Gutjahr, O., Schefczyk, L., Reiter, P. and Heinemann, G.: Impact of the horizontal resolution on the simulation of extremes in COSMO-CLM, Meteorol. Z., 25, 543 – 562, doi: 10.1127/metz/2016/0638, 2016.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fuckar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185-4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Hart, N. C. G., Washington, R., and Stratton, R. A.: Stronger local overturning in convective-permitting regional climate model improves simulation of the subtropical annual cycle. Geophys. Res. Lett., 45, 11334–11342, https://doi.org/10.1029/2018GL079563, 2018.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. J. Geophys. Res. Atmos., 113, D20119. doi:10.1029/2008JD010201, 2008.

Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Simmons, A., Soci, C., Dee, D.: Global reanalysis: goodbye ERA-Interim, hello ERA5. ECMWF, doi:10.21957/vf291hehd7. https://www.ecmwf.int/node/19027, 2019

Herrmann, M., Somot, S., Calmanti, S., Dubois, C., and Sevault, F.: Representation of spatial and temporal variability of daily wind speed and of intense wind events over the Mediterranean Sea using dynamical downscaling: impact of the regional climate model configuration, Nat. Hazards Earth Syst. Sci., 11, 1983-2001, https://doi.org/10.5194/nhess-11-1983-2011, 2011.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Ko-vats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., Meijgaard, E. van, Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, Reg. Environ. Change, 14, 563–578, doi:10.1007/s10113-013-0499-2, 2014.

Jézéquel, A., Yiou, P. and Radanovics, S.: Role of circulation in European heatwaves using flow analogues. Clim. Dynam. 50, 1145-1159, https://doi.org/10.1007/s00382-017-3667-0, 2018.

~~Jones, P. D., Harpham, C., Troccoli, A., Gschwind, B., Ranchin, T., Wald, L., Goodess, C. M., and Dorling, S. Using ERA-Interim reanalysis for creating datasets of energy-relevant climate variables, Earth Syst. Sci. Data, 9, 471-495, https://doi.org/10.5194/essd-9-471-2017, 2017.~~

Jung, T., Gulev, S. K., Rudeva, I. and Soloviov, V.: Sensitivity of extratropical cyclone characteristics to horizontal resolution in the ECMWF model. Q.J.R. Meteorol. Soc., 132, 1839-1857, doi:10.1256/qj.05.212, 2006.

Jung, T., Miller, M. J., Palmer, T. N., Towers, P., Wedi, N., Achuthavarier, D., Adams, J. M., Altshuler, E. L., Cash, B. A., Kinter, J. L., Marx, L., Stan, C., and Hodges, K. I.: High-Resolution Global Climate Simulations with the ECMWF Model in Project Athena: Experimental Design, Model Climate, and Seasonal Forecast Skill, J. Climate., 25, 3155–3172, doi:10.1175/JCLI-D-11-00265.1, 2012.

Kendon, E. J., Roberts, N. M., Senior, C. A., and Roberts, M. J.: Realism of rainfall in a very high-resolution regional climate model, J. Climate., 25, 5791–5806. doi: 10.1175/JCLI-D-11-00562.1, 2012.

Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., and Senior, C. A.: Heavier summer downpours with climate change revealed by weather forecast resolution model, Nat. Clim. Change, 4, 570–576, doi: 10.1038/nclimate2258, 2014.

Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., Evans, J. P., Fosser, G. and Wilkinson, J.M.: Do Convection-Permitting Regional Climate Models Improve Projections of Future Precipitation Change?. B. Am. Meteorol. Soc., 98, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1, 2017.

Klaver, R., Haarsma, R., Vidale, P. L., Hazeleger, W.: Effective resolution in high resolution global atmospheric models for climate studies. Atmos Sci Lett., 1– 8. https://doi.org/10.1002/asl.952, 2020

Kopparla, P., Fischer, E. M., Hannay, C., and Knutti, R.: Improved simulation of extreme precipitation in a high-resolution atmosphere model, Geophys. Res. Lett., 40, 5803-5808, doi: 10.1002/2013GL057866, 2013.

Kunz, M., Mohr, S., Rauthe, M., Lux, R., and Kottmeier, C.: Assessment of extreme wind speeds from Regional Climate Models – Part 1: Estimation of return values and their evaluation, Nat. Hazards Earth Syst. Sci., 10, 907-922, https://doi.org/10.5194/nhess-10-907-2010, 2010.

Landelius, T., Dahlgren, P., Gollvik, S., Jansson, A. and Olsson, E.: A high-resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind, Q. J. R. Meteorol. Soc, 142, 2132–2142, doi:10.1002/qj.2813, 2016.

Lhotka, O., Kyselý, J. and Farda, A.: Climate change scenarios of heat waves in Central Europe and their uncertainties. Theor Appl Climatol., 131, 1043-1054, https://doi.org/10.1007/s00704-016-2031-3, 2018.

Matsueda, M., Mizuta, R. and Kusunoki, S.: Future change in wintertime atmospheric blocking simulated using a 20-km-mesh atmospheric global circulation model, J. Geophys. Res., 114, D12114, doi: 10.1029/2009JD011919, 2009.

Mizielinski, M. S., Roberts, M. J., Vidale, P. L., Schiemann, R., Demory, M.-E., Strachan, J., Edwards, T., Stephens, A., Lawrence, B. N., Pritchard, M., Chiu, P., Iwi, A., Churchill, J., del Cano Novales, C., Kettleborough, J., Roseblade, W., Selwood, P., Foster, M., Glover, M., and Malcolm, A.: High-resolution global climate modelling: the UPSCALE project, a large-simulation campaign, Geosci. Model Dev., 7, 1629–1640, doi:10.5194/gmd-7-1629-2014, 2014.

O'Brien, T. A., Collins, W. D., Kashinath, K., Rübel, O., Byna, S., Gu, J., Krishnan, H. and Ullrich, P. A.: Resolution dependence of precipitation statistical fidelity in hindcast simulations, J. Adv. Model. Earth Syst., 8, 976–990, doi: 10.1002/2016MS000671, 2016.

O'Reilly, C. H., Minobe, S. and Kuwano-Yoshida, A.: The influence of the Gulf Stream on wintertime European blocking, Clim. Dynam., 47, 1545- 1567, https://doi.org/10.1007/s00382-015-2919-0, 2016.

Prein, A. F. and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis, Int. J. Climatol., 37, 305-327, doi:10.1002/joc.4706, 2017.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., Van Lipzig, N. P. M., and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Rev. Geophys., 53, 323–361. doi: 10.1002/2014RG000475, 2015.

Prein, A.F., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., Fox Maule, C., van Meijgaard, E., Déqué, M., Nikulin, G., Vautard, R., Colette, A., Kjellström, E., and Jacob, D.: Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: high resolution, high benefits?, Clim. Dynam.. 46, 383-412, doi: 10.1007/s00382-015-2589-y, 2016.

Pryor, S. C., Nikulin, G., and Jones, C.: Influence of spatial resolution on regional climate model derived wind climates, J. Geophys. Res., 117, D03117, doi:10.1029/2011JD016822, 2012

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-ARW Model during 2017 North American Monsoon Season. Atmosphere, 10, 694, doi:10.3390/atmos10110694, 2019

Roberts, M. J., Vidale, P. L., Senior, C., Hewitt, H. T., Bates, C., Berthou, S., Chang, P., Christensen, H. M., Danilov, S., Demory, M. E., Griffies, S. M., Haarsma, R., Jung, T., Martin, G., Minobe, S., Ringler, T., Satoh, M., Schiemann, R., Scoccimarro, E., Stephens, G. and Wehner. M.F.: The benefits of global high-resolution for climate simulation: process-understanding and the enabling of stakeholder decisions at the regional scale.. B. Am. Meteorol. Soc., 99, 2341–2359 https://doi.org/10.1175/BAMS-D-15-00320.1, 2018.

Ruti, P. M., Somot, S., Giorgi, F., Dubois, C., Flaounas, E., Obermann, A., Dell'Aquila, A., Pisacane, G., Harzallah, A., Lombardi, E., Ahrens, B., Akhtar, N., Alias, A., Arsouze, T., Aznar, R., Bastin, S., Bartholy, J., Béranger, K., Beuvier, J., Bouffies-Cloché, S., Brauch, J., Cabos, W., Calmanti, S., Calvet, J., Carillo, A., Conte, D., Coppola, E., Djurdjevic, V., Drobinski, P., Elizalde-Arellano, A., Gaertner, M., Galàn, P., Gallardo, C., Gualdi, S., Goncalves, M., Jorba, O., Jordà, G., L'Heveder, B., Lebeaupin-Brossier, C., Li, L., Liguori, G., Lionello, P., Maciàs, D., Nabat, P., Önol, B., Raikovic, B., Ramage, K., Sevault, F., Sannino, G., Struglia, M. V., Sanna, A., Torma, C., and Vervatis, V.: Med-CORDEX Initiative for Mediterranean Climate Studies. B. Am. Meteorol. Soc., 97, 1187–1208, https://doi.org/10.1175/BAMS-D-14-00176.1, 2016

Schiemann, R., Demory, M. E., Shaffrey, L. C., Strachan, J., Vidale, P. L., Mizielinski, M. S., Roberts, M. J., Matsueda, M., Wehner, M. F. and Jung, T.: The resolution sensitivity of Northern Hemisphere blocking in four 25-

km atmospheric global circulation models, J. Climate., 30, 337–358, https://doi.org/10.1175/JCLI-D-16-0100.1, 2017.

Schiemann, R., Vidale, P. L., Shaffrey, L. C., Johnson, S. J., Roberts, M. J., Demory, M.-E., Mizielinski, M. S., and Strachan, J.: Mean and extreme precipitation over European river basins better simulated in a 25 km AGCM, Hydrol. Earth Syst. Sci., 22, 3933-3950, https://doi.org/10.5194/hess-22-3933-2018, 2018.

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C. and Zhang, X.: Changes in climate extremes and their impacts on the natural physical environment. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change, edited by: Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K., L. Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M. and Midgley, P. M., Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 109-230, 2012

Shields, C. A., Kiehl, J. T., and Meehl, G. A.: Future changes in regional precipitation simulated by a half-degree coupled climate model: Sensitivity to horizontal resolution, J. Adv. Model. Earth Syst., 8, 863–884, doi: 10.1002/2015MS000584, 2016.

Skamarock, W.C.: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Mon. Wea. Rev*., 132, 3019–3032, https://doi.org/10.1175/MWR2830.1, 2004

Stott, P. A., Christidis, N., Otto, F. E., Sun, Y., Vanderlinden, J., van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P. and Zwiers, F. W.: Attribution of extreme weather and climate-related events. WIREs Clim. Change, 7, 23-41. doi:10.1002/wcc.380, 2016.

Taylor, K., Williamson, D., and Zwiers, F,: The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations, PCMDI Rep. 60, Tech. Rep. 60, PCMDI, 25 pp., available at: http://www-pcmdi.llnl.gov/publications/ab60.html, 2000.

Taylor, K. E., Stouffer, R. J. and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485498, doi: 10.1175/BAMS-D-11-00094.1, 2012.

Terai, C. R., Caldwell, P. M., Klein, S. A. Tang, Q. and Branstetter, M. L.: The atmospheric hydrologic cycle in the ACME v0.3 model. Clim. Dynam., 50, 3251- 3279. https://doi.org/10.1007/s00382-017-3803-x, 2018.

Torma, C., Giorgi, F. and Coppola, E.: Added value of regional climate modeling over areas characterized by complex terrain—Precipitation over the Alps, J. Geophys. Res. Atmos., 120, 3957–3972, doi: 10.1002/2014JD022781, 2015.

Van Haren, R., Haarsma, R. J., Van Oldenborgh, G. J. and Hazeleger, W.: Resolution Dependence of European Precipitation in a State-of-the-Art Atmospheric General Circulation Model, J. Climate., 28, 5134–5149, doi: 10.1175/JCLI-D-14-00279.1, 2015a.

Van Haren, R., Haarsma, R. J., de Vries, H., van Oldenborgh, G. J., and Hazeleger, W.: Resolution dependence of circulation forced future central European summer drying, Environ. Res. Lett., 10, 055002, doi:10.1088/1748-9326/10/5/055002, 2015b.

Vanniere, B., Vidale, P. L., Demory, M.-E., Schiemann, R., Roberts, M. J., Roberts, C. D., Matsueda, M., Terray, L., Koenigk, T., Senan, R.: Multi-model evaluation of the sensitivity of the global energy budget and hydrological cycle to resolution, Clim. Dynam., 52, 6817- 6846, https://doi.org/10.1007/s00382-018-4547-y, 2019

Vautard, R., Gobiet, A., Jacob, D., Belda, M., Colette, A., Déqué, M., Fernández, J., García-Díez, M., Goergen, K., Güttler, I., Halenka, T., Karacostas, T., Katragkou, E., Keuler, K., Kotlarski, S., Mayer, S., van Meijgaard, E., Nikulin, G., Patarcic, M., Scinocca, J., Sobolowski, S., Suklitsch, M., Teichmann, C., Warrach-Sagi, K., Wulfmeyer, V., Yiou, P. : The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project, Clim. Dynam., 41, 2555-2575, doi: 10.1007/s00382-013-1714-z, 2013.

Vautard, R., Yiou, P., Otto, F., Stott, P., Christidis, N., van Oldenborgh, G. J. and Schaller, N.: Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events, Environ. Res. Lett., 11, 114009, https://doi.org/10.1088/1748-9326/11/11/114009, 2016.

Volosciuk, C., Maraun, D., Semenov, V.A. and Park, W.: Extreme Precipitation in an Atmosphere General Circulation Model: Impact of Horizontal and Vertical Model Resolutions, J. Climate., 28, 1184–1205, https://doi.org/10.1175/JCLI-D-14-00337.1, 2015.

Vries, H. de, Scher, S., Haarsma, R., Drijfhout, S., and Delden, A. van.: How Gulf-Stream SST-fronts influence Atlantic winter storms, Clim. Dynam., 52, 5899-5909. https://doi.org/10.1007/s00382-018-4486-7, 2019

~~Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J. and Viterbo, P., The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., 50, 7505–7514, doi:10.1002/2014WR015638, 2014.~~

Wehner, M. F., Smith, R. L., Bala, G. and Duffy, P.: The effect of horizontal resolution on simulation of very extreme US precipitation events in a global atmosphere model, Clim. Dynam., 34, 241-247. https://doi.org/10.1007/s00382-009-0656-y, 2010.

Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Chen, C.-T., Paciorek, C., Gleckler, P. J., Sperber, K. R., Collins, W. D., Gettelman, A., and Jablonowski, C.: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1, J. Adv. Model. Earth Syst., 6, 980–997, doi:10.1002/2013MS000276, 2014.

Willison, J., Robinson, W.A. and Lackmann, G.M.: North Atlantic Storm-Track Sensitivity to Warming Increases with Model Resolution,. J. Climate., 28, 4513–4524, https://doi.org/10.1175/JCLI-D-14-00715.1, 2015.

Zappa, G., Shaffrey, L. C. and Hodges, K. I., The Ability of CMIP5 Models to Simulate North Atlantic Extratropical Cyclones, J. Climate., 26, 5379-5396, doi: 10.1175/JCLI-D-12-00501.1, 2013.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C. Trewin, B. and Zwiers, F. W.: Indices for monitoring changes in extremes based on daily temperature and precipitation data, Wiley Interdiscip. Rev., Clim. Chang., 2, 851–870, doi:10.1002/wcc.147, 2011.

Figures

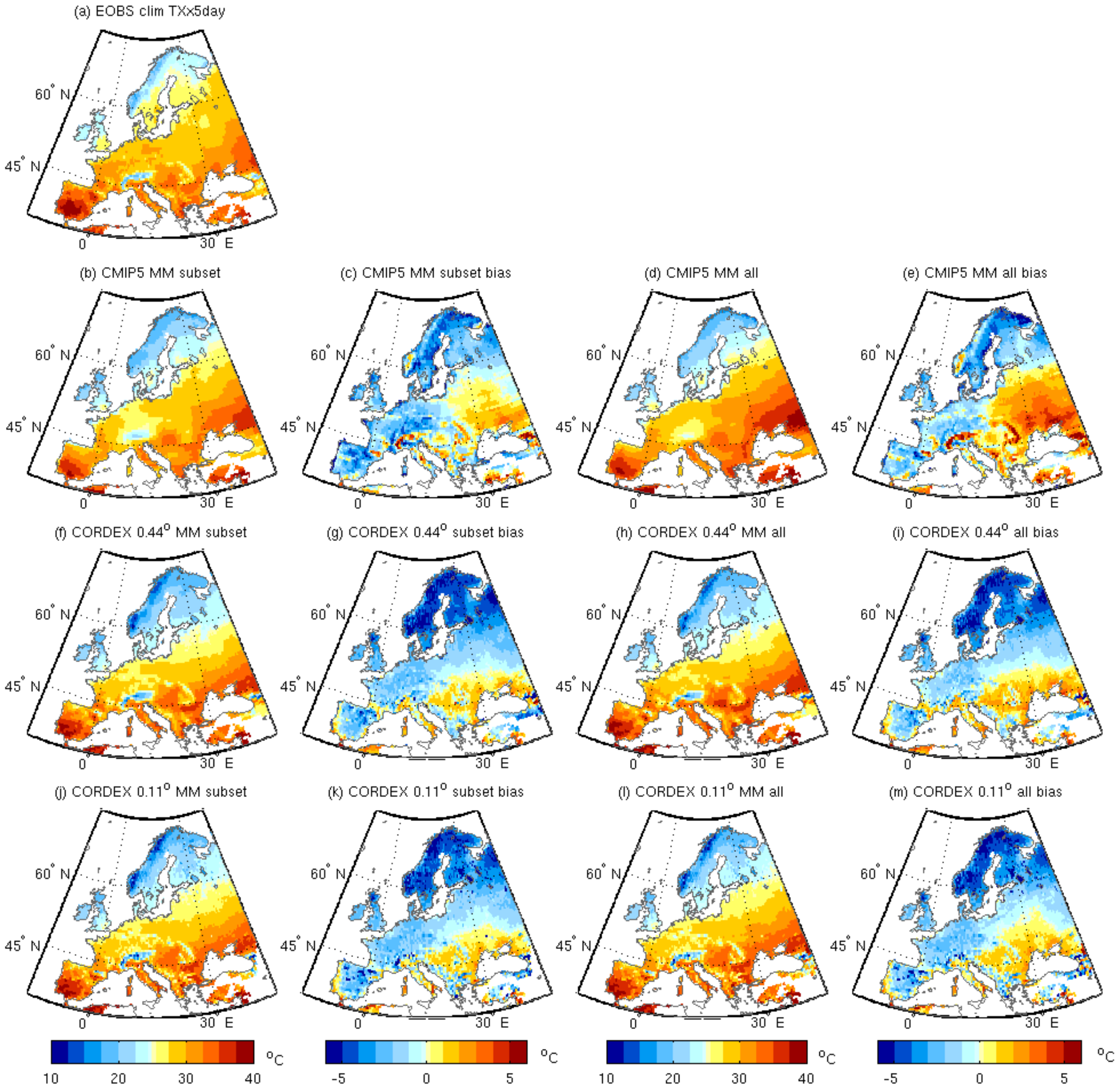

Figure 1: Climatological mean of TXx5day for the period 1970-2005 for (a) EOBS, the multi model mean of the common subset of models (see Methods) for (b) CMIP5, (f) CORDEX 0.44° and (j) CORDEX 0.11°, (c, g, k) their biases with respect to EOBS, and (d,e,h,i,j,k) the same for the full ensembles of CMIP5, and CORDEX. Units °C.

Comment [c2]: Replaced with different regridding

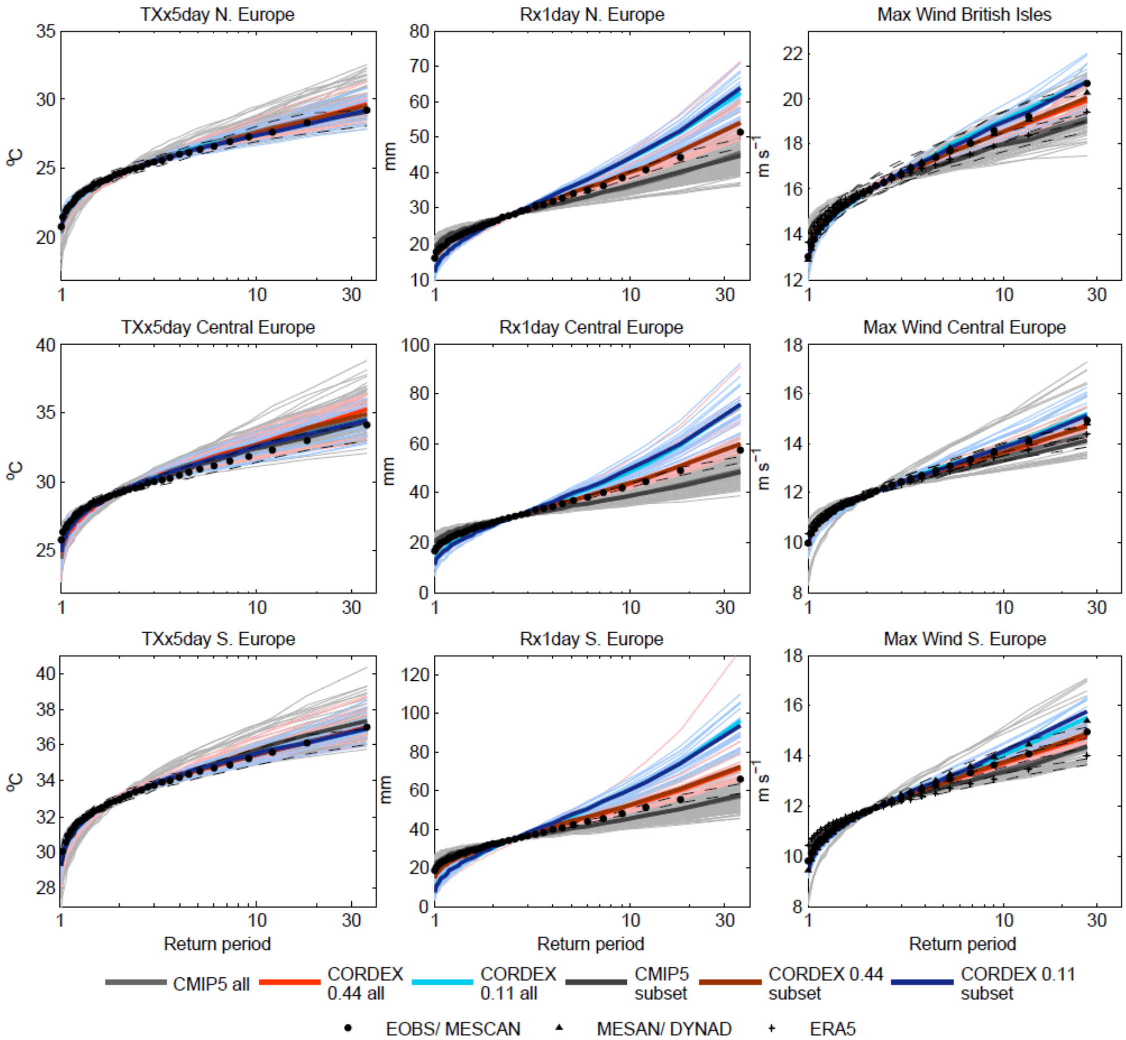

Figure 2: Return period plots for (left) TXx5day, middle column Rx1day and (right) annual maximum wind, for CMIP5 and CORDEX for Northern Europe (top row (except top left = British Isles)), Central Europe (middle row) and Southern Europe (bottom row). CMIP5 is shown in grey, CORDEX 0.44° in red and CORDEX 0.11° in blue. Thin lines are individual ensemble members, thick lines are multi model medians,~~circles represent the pooled ensembles~~, lighter shades for the full ensembles, and darker shades for the subset of models common to CMIP5, and both CORDEX resolutions. Observations are shown in black, circles for E-OBS temperature and precipitation and MESCAN~~WFDEI~~ wind, triangles for MESAN precipitation and DYNAD~~ECEM noc~~ wind and crosses for ~~ECEM wbc~~ERA5 wind. Confidence intervals based on bootstrapping are shown with dashed lines for the observations. The time periods considered are 1970-2005 for TXx5day and Rx1day, and 1979-2005 for wind.

Comment [c3]: Pooling replaced my multi-model medians.
Regridding changed for temperature, and cordex 0.11, MESAN

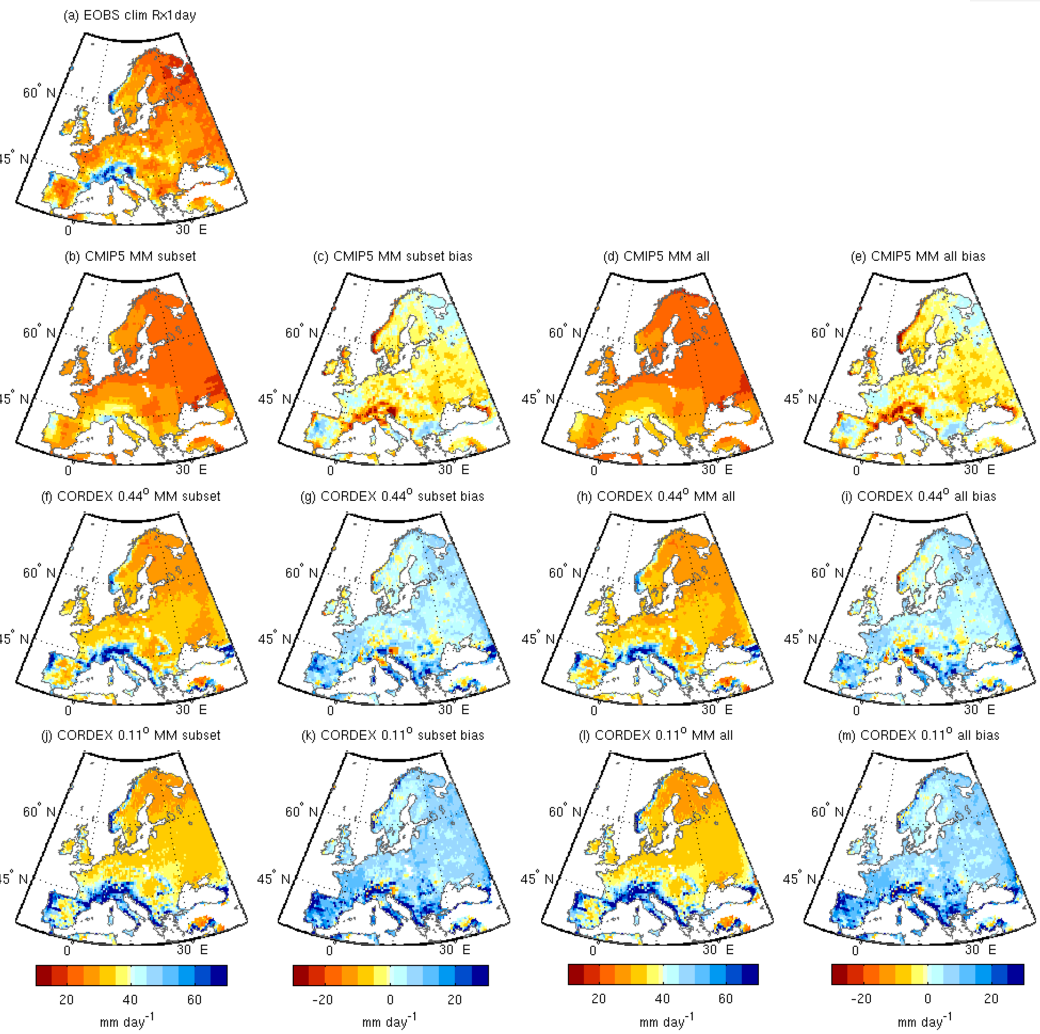

Figure 3: As for Figure 1 but for the climatological mean of Rx1day. Units mm.

Comment [c4]: Replaced with bicubic regridding for cordex 0.11

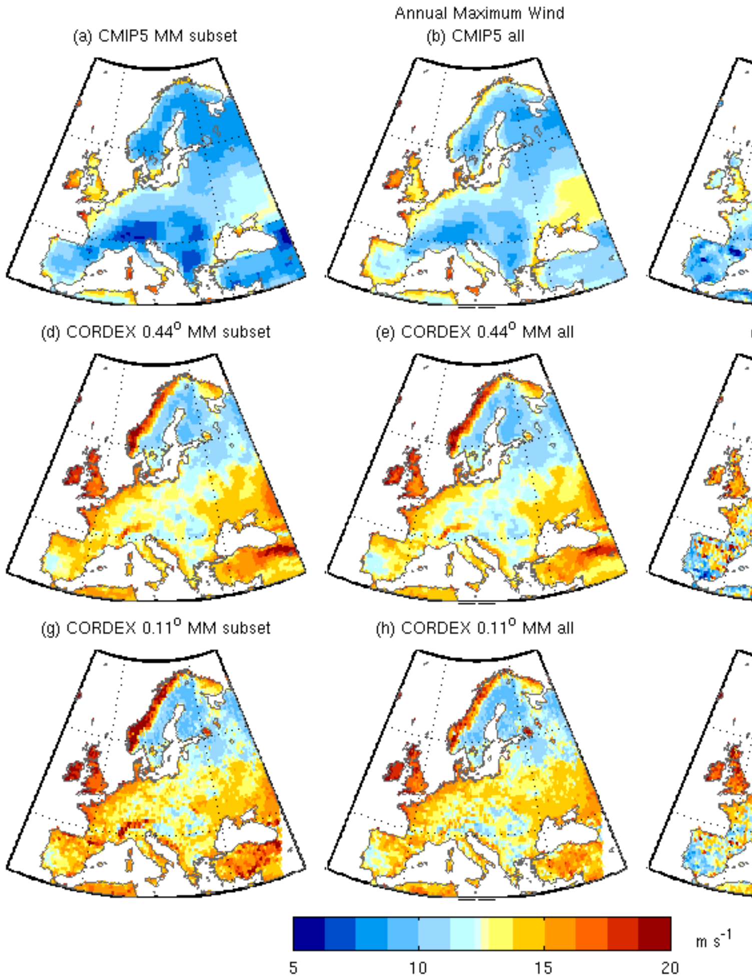

Figure 4: Climatological mean of annual maximum of daily maximum wind for the period 1979-2005 for the multi model mean of the common subset of models for (a) CMIP5, (d) CORDEX 0.44° and (g) CORDEX 0.11°, (b, e, h) the same for the full ensembles of CMIP5 and CORDEX, and the observational datasets (c) ERA5~~WFDEI~~, (f) ~~ECEM noc~~MESCAN (i) DYNAD~~ECEM wbc~~. Units meters per second.

Comment [c5]: Replaced the reanalysis datasets. Changed the landmask to avoid including ocean grid cells. Changed regridding method for CORDEX 0.11 to bicubic.

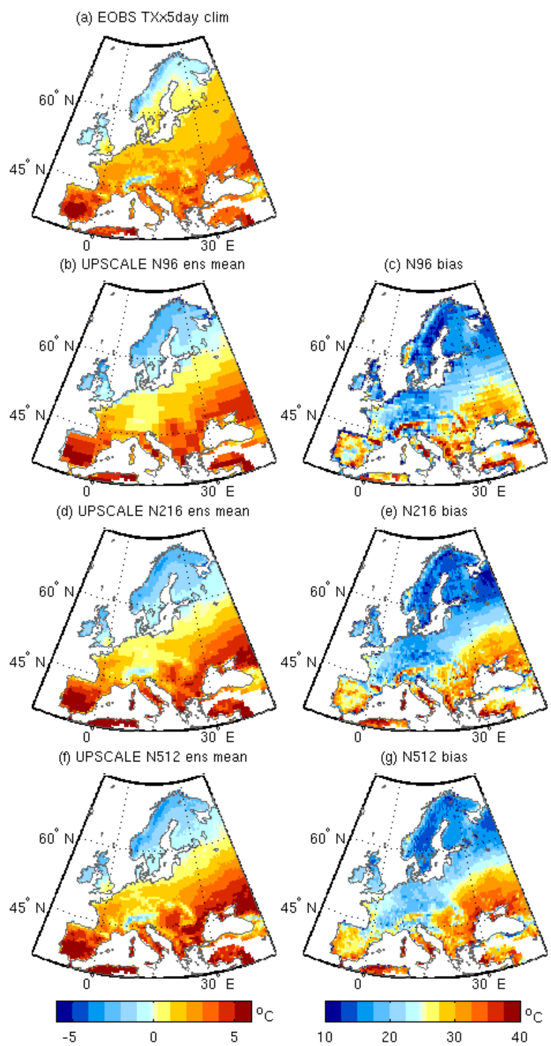

Figure 5: Climatological mean of TXx5day for the ensemble means of three resolutions of HadGEM3-A (UPSCALE) GCM simulations (left) for the period 1985-2011 and their biases with respect to E-OBS (right). (a) EOBS, (b, c) N96 (130 km), (d, e) N216 (60 km), (f, g) N512 (25 km). Units °C.

Comment [c6]: Changed regridding method from bilinear to nearest neighbour

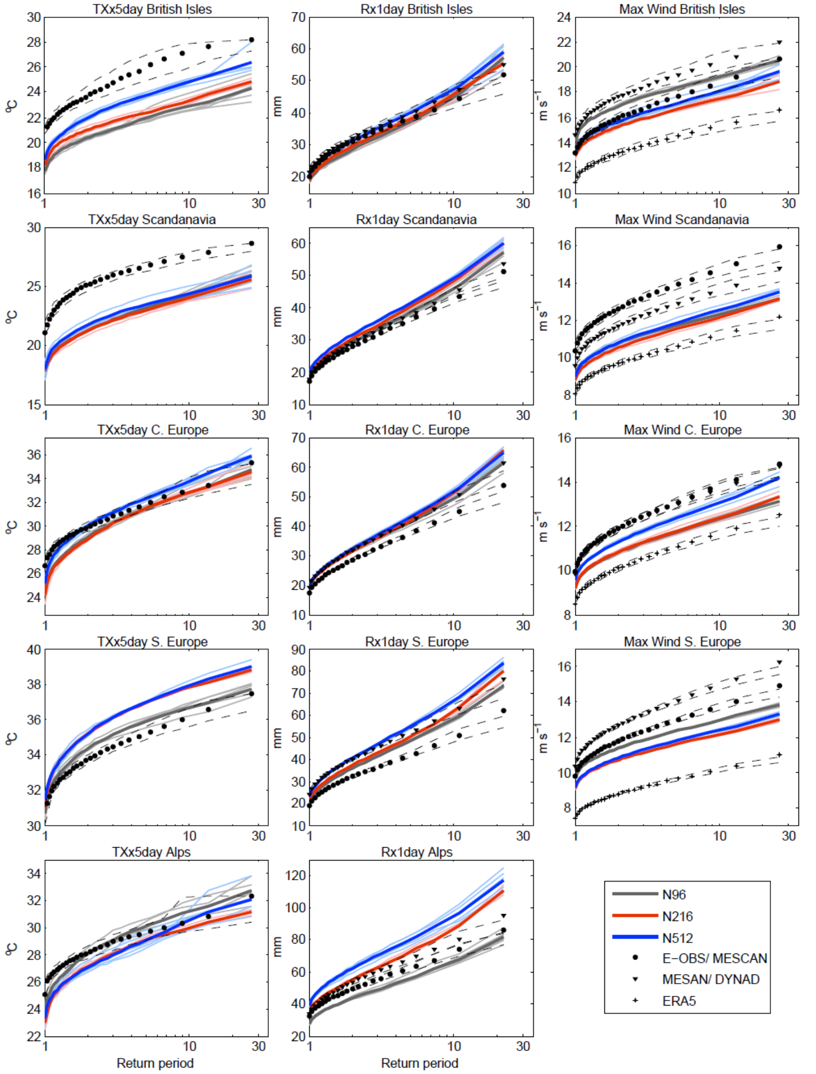

Figure 6: Return period plots for (left) TXx5day, middle column Rx1day and (right) annual maximum wind, for the UPSCALE simulations for (top row) the British Isles, (2nd row) Scandinavia, (3rd row) Central Europe, (4th row) Southern Europe, and (last row) the Alps. N96 is shown in grey, N216 in red and N512 in blue. Thin lines are individual ensemble members, thick lines~~circles~~ represent ensemble means~~the pooled ensembles~~. Observations are shown in black, circles for E-OBS and MESCAN~~WFDEI~~, triangles for MESAN~~–~~ and DYNAD~~and ECEM noc~~, and asterisks for ~~ECEM wbe~~ERA5. Confidence intervals based on

Comment [c7]: Pooling replaced by ensemble means
Regridding method switched to nearest neighbour for temperature
Observational wind datasets replaced and land masking for wind adjusted

bootstrapping are shown with dashed lines for the observations. The time periods considered are 1985-2011 for TXx5day, 1989-2010 for Rx1day, and 1986-2011 for wind. NB: there is no bias adjustment~~correction~~ of the climatology (see methods).

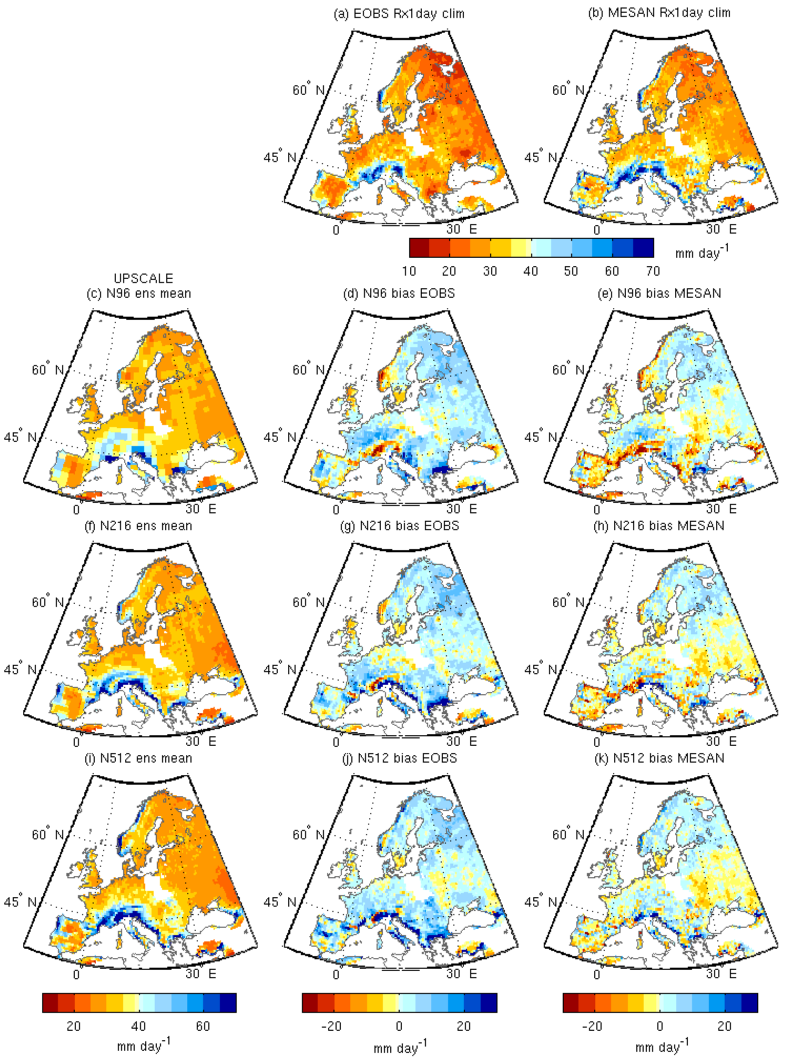

Figure 7: Climatological mean of Rx1day for the ensemble means of three resolutions of UPSCALE (left) simulations for the period 1989-2010 and their biases with respect to E-OBS (middle) and the MESAN reanalysis (right). (a) EOBS, (b) MESAN (c-e) N96, (f-h) N216, (i-k) N512. Units mm.

Comment [c8]: MESAN replaced by 5.5km version regridded bicubically.

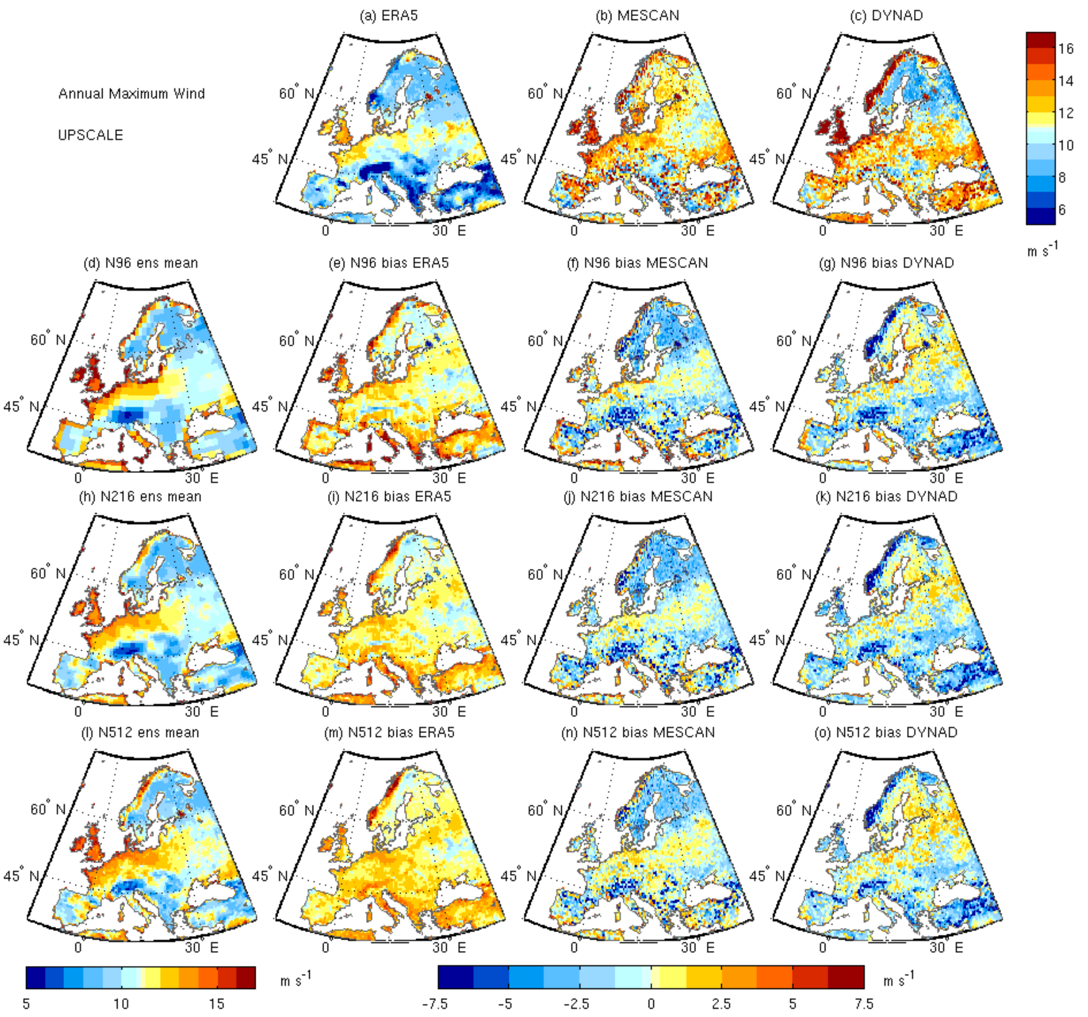

Figure 8: Climatological mean of annual maximum wind for the ensemble means of three resolutions of UPSCALE (left) simulations for the period 1986-2011 and their biases with respect to the observational datasets ~~ECEM wbc~~ERA5 (left), ~~ECEM noc~~MESCAN (middle) and ~~WFDEI~~ MESAN (right). (a) ~~ECEM wbc~~ERA5, (b) ~~ECEM noc~~MESCAN (c) ~~WFDEI~~DYNAD, (d-g) N96, (h-k) N216, (l-o) N512. Units meters per second.

Comment [c9]: Observational datasets replaced. More restrictive land masking applied.

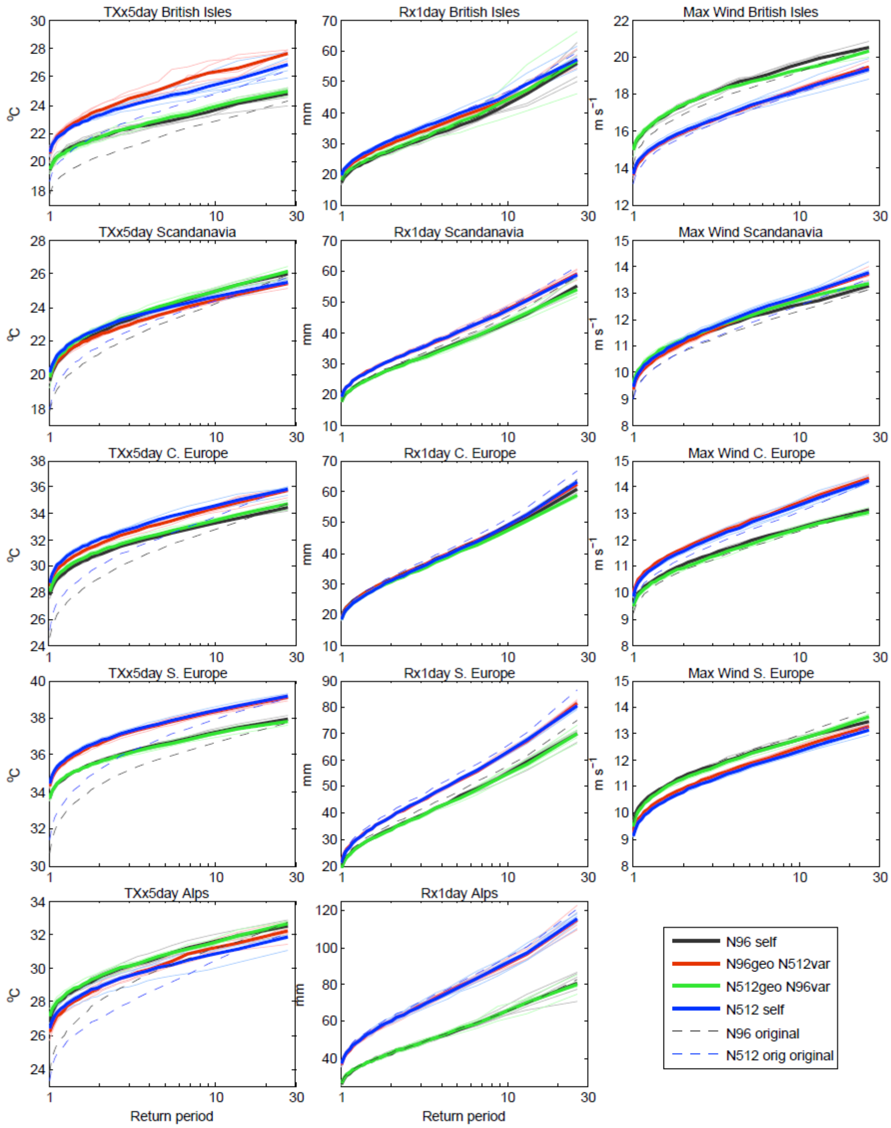

Figure 9: Circulation analogue results. Return period plots for (left) TXx5day, (middle) Rx1day and (right) annual maximum wind for (top) the British Isles, (2nd row) Scandinavia, (3rd row) Central Europe, (4th row) Southern Europe and (5th row) the Alps. Grey represents the N96 self-analogues, blue the N512 self-analogues, red is for N96 circulation with N512 variables (e.g.

Comment [c10]: -Pattern correlation instead of Euclidean distance for the analogues
-Regridding changed to nearest neighbour for temperature.
-More restrictive landmasking applied for wind (to better match the other variables too)
-Pooled results replaced by ensemble means.
-Domain was wrong for temperature

precipitation) and green is for N512 circulation with N96 variables. Thin lines represent individual ensemble members, thick lines~~circles~~ represent the mean across individual ensemble members~~results pooled across ensemble members~~. ~~Dashed lines are 5-95% confidence intervals based on a bootstrapping technique.~~ ~~Thick~~ B~~b~~lue dashed line represents the original ~~pooled~~ N512 ensemble mean results like those shown in Figure 6 (although sometimes based on a different time period), and the ~~thick~~ grey dashed lines represent~~s~~ the equivalent for the N96 simulations. Results for TXx5day are based on the period 1985-2011, Rx1day 1986-2011, and wind 1986-2011.