

Interactive comment on “The importance of model resolution on simulated precipitation in Europe –from global to regional model” by Gustav Strandberg and Petter Lind

Anonymous Referee #1

This study analyses precipitation characteristics over Europe from a wide range of model ensembles, including Global Climate Models (CMIP5, CMIP6, PRIMAVERA) and Regional Climate Models (CORDEX). The precipitation characteristics include daily precipitation distributions based on the ASoP diagnostics developed by Klingaman et al (2017), as well as statistical metrics such as number of wet days, number of heavy precipitation days, intensity of wet days, intensity of heaviest precipitation day. The aim of this study is three-fold: 1) investigate differences between model ensembles, and between models within each ensemble, by using a wide range of ensembles from CMIP5, CMIP6, PRIMAVERA and CORDEX; 2) evaluate model performance against observations, using E-OBS data; 3) investigate the role of resolution in precipitation characteristics over Europe, by selecting only models available at both low and high resolution versions. I have several comments regarding this study, as described below. Some of them would require more analyses and restructuring of the paper, but I think it would also greatly improve it.

We thank you for making the effort of reviewing the paper and for all constructive suggestions. Responses to comments follow below in red. In the markup version of the revised manuscript substantial changes are also marked in red.

1) The authors have made an impressive work by analysing such a huge amount of simulations. This is very complementary to the work by Demory et al (2020), which have analysed daily precipitation over Europe in CMIP5, PRIMAVERA (high-resolution) and CORDEX (low and high resolutions) compared to high-quality observational datasets over Europe. This work has now been revised by focusing more on EUR-11 (which is a newer ensemble than EUR-44), and by including also spatial distribution of precipitation and Taylor diagrams, which confirm the results shown by the precipitation distribution. The paper is now accepted and should appear soon. I suggest to refer to this study already in the introduction. Iles et al (2020) could also be referred to in the introduction as another study evaluating a range of GCMs and RCMs at various resolutions, considering the atmosphere-only UPSCALE simulations. The fact that this study and Demory et al find similar results, despite using slightly different methods, give strength to these two studies and should be discussed further.

The Introduction has been expanded with a paragraph discussing Demory et al., and Iles et al., as well as other similar studies using CORDEX data:

A few studies have been made investigating how model resolution affects the simulated precipitation in the CORDEX ensembles, comparing 50 km and 12.5 km grid spacing. A clear result is that precipitation generally increases with higher resolution, which sometimes means that the bias increases when precipitation is added to already wet models (Kotlarski et al., 2014; Casanueva et al., 2016); something that is also seen in simulations with global models (e.g. Thackeray et al., 2018). An overall improvement of mean precipitation is not seen the high resolution CORDEX simulations, except for regions with complex topography (Kotlarski et al., 2014; Casanueva et al., 2016; Prein et al., 2016). Prein et al. (2016) looked at local precipitation on short time scales. They find that 12.5 km simulations better represents extreme and mean precipitation, also when simulations are aggregated to 50 km. They note,

however, that the results are highly dependent on which observations the simulations are compared with. They also note that improvements are on the ensemble as a whole, and not necessarily for each individual model. In similar studies as the present Iles et al. (2019) and Demory et al. (2020) compare CORDEX simulations with simulations from CMIP5 and Primavera. They see that precipitation increases with resolution so that CMIP5 underestimates precipitation amounts and CORDEX overestimates it, when compared to E-OBS, and that the effect of resolution is largest in complex topography. They also find that Primavera performs similarly to CORDEX when run on the same resolution, which is interesting regarding that the Primavera models are developed for low resolution. Iles et al. (2019) also find considerable inter-model differences meaning that improvements are seen on the ensemble level rather than for individual models.

2) The authors have managed to combine their results into well-designed figures. However, I feel the 3 goals should not be addressed with the same method. The authors have indeed decided to perform the analyses on the model native grids. This is a good choice for showing what each ensemble is able to simulate at its own resolution, and could be used for addressing aim 1) written above, as long as the models are not compared to each other. A clean comparison could only be done on a common coarser grid, as emphasised by Klingaman et al, 2017. Evaluating results on native grid not only shows the potential of the model physics but also includes the technical aspect of doing analyses on a finer grid. This technical aspect can be evaluated by regridding the data on a coarser grid and see how the results are affected by such a regridding. Evaluating results on common grids would show the impact of the model physics, its internal resolution solely (Na et al, 2020), and allows a direct assessment and inter-comparison of the results across resolutions (Demory et al, 2020; Iles et al, 2020; see also Torma et al, 2015 (their Fig 3-6)). I would therefore suggest to redo analyses on a common coarser grid to verify the results shown on native grid. I believe this would strengthen the results. One way to answer all 3 aims of the study could be to split it into two parts: the first part would address 1) and 2) on native grids, considering observations available at various resolutions (such as low-resolution satellite data on grids similar to CMIP); the second part would evaluate the impact of resolution by regridding all data on a common coarser grid.

Thanks for pushing us in this direction. We have now included analyses where all data are regridded to two common grids $0.5^{\circ} \times 0.5^{\circ}$ grid and a $2^{\circ} \times 2^{\circ}$ grid.

3) The models are evaluated against E-OBS. E-OBS is a good product that tries to gather the highest number of stations currently available. This is particularly the case over Scandinavian regions, or Germany. However, there are still many regions where the station density is low (e.g. France, Italy, Spain, Switzerland, Austria). Over these regions, it would be better to use national gridded datasets, available at much higher resolution (see Demory et al, 2020 for details). I understand the authors may not want to go in that direction, as it adds a lot of processing time and the definitions of the regions would be slightly different than in the current study. I would therefore suggest to include a discussion on this (and eventually an intercomparison with observational results of Demory et al if feasible). Moreover, for aim 1) of the study, I would suggest the authors to use another lower resolution dataset, such as satellite observations, using a resolution closer to CMIP models. This would give an additional range of observational uncertainty.

Thank you, this is indeed an important and interesting issue. We agree that it would be valuable if regional and/or national observational datasets (with assumed higher quality than E-OBS) could be included for each of the investigated sub-regions, as for example in Demory et al 2020. As suggested we have now included a separate section with a discussion of observations and their associated uncertainties, including E-OBS. To emphasize the importance of high-quality observations and to partly put our results into perspective, we have also included an ASoP analysis comparing to another high-resolution (1x1 km) dataset covering Scandinavia - called NGCD (Nordic Gridded Climate Dataset). There we can clearly see the impact of including such observations, increasing the confidence in the high-resolution RCM model ensemble. We have not included any satellite data as these often (at least the ones we are aware of) has limited coverage or lower quality over high latitudes.

4) Please verify the use of model resolution when you actually refer to model horizontal grid spacing. The model effective resolution is typically 4 to 8 times the model horizontal grid spacing (Skamarock, 2004; Klaver et al, 2019).

Thanks for reminding us about this. We tried to straighten up the terminology so that we use "grid spacing" when talking about distances in km and "resolution" in more general statements, like comparing high and low resolution models.

5) Most analyses have been performed annually. It would be good to show them seasonally as well (at least DJF and JJA), as the processes driving precipitation are different and RCMs depend more on GCMs in DJF than JJA (e.g. Hall, 2014; Prein et al, 2016; Fernandez et al, 2019).

We have now also included analyses of DJF and JJA. However we could not present results for all regions, seasons and resolutions as this would mean at least a 12 fold increase of the number of figures.

6) The abstract needs to be revised. It writes very general conclusions as it stands. See detailed suggestions below. This is true as well for the entire text. Some sentences are a bit hard to read, and in many places, it reads like general statements or approximative sentences. I provided some suggestions for some of them below, but a careful review of the language would clarify the text and be beneficial to the final paper.

The abstract is rewritten to be more precise, and so is the rest of the text. We hope in a satisfactory way. Thanks for the detailed comments.

7) For reproducibility of the results, it appears important to list the models that were considered for the study.

We have inserted a new Table 1 listing the GCMs and a new Table 2 listing the RCMs.

Detailed comments:

Title: the importance of model 'horizontal' resolution... from global to regional 'models'

Changed as suggested.

L. 10: model 'horizontal' resolution

Changed as suggested.

L. 17-18: I find this conclusion too general. This depends on seasons, and most of the analyses have been performed annually.

The abstract is rewritten to be more precise.

L. 20: I don't agree with this. The authors have shown here that the improvement is systematic across models but that there is a large inter-model variability.

This is rephrased to: "Even though higher resolution improves the simulated precipitation in a systematic way, the inter-model variability is still large. This means that the quality of an ensemble depends also on the models it consists of and not only the average resolution of the ensemble."

L. 21: I agree with this, but I think it cannot be generalised for all resolutions. The authors have shown here that the averaged resolution of CMIP5 and CMIP6 anyway is too low to capture the characteristics of precipitation, at least against E-OBS and other higher resolution ensembles.

We imply that this is valid for the resolutions used in RCMs. To make this clearer we start the section with "Once reaching ~50 km ...".

L. 22: again, this depends on the season and the authors have mostly worked with annual means.

The abstract is rewritten to be more precise.

L. 23: different RCMs driven by the same GCM give different results, but the same RCM driven by different GCMs also give different results (e.g. Vautard et al, 2020).

That's true, and we know this of course. We show it in Fig. 10 and mention it at a few different times. We change the sentence to : "The result of a RCM simulation depends on the driving GCM, but the difference in simulated precipitation between an RCM and the driving GCM depends more on the choice of RCM and less on the down-scaling itself; as different RCMs driven by the same GCM may give different results."

If Vautard et al., 2020 is published before this goes to print we will add a reference to that.

L. 24-25: Given the complementarity to Demory et al (2020), this sentence needs to be rewritten.

This is changed to: "The results presented here are in line with previous similar studies. To these studies we add details about the spread between resolutions and between models."

L. 28: delete 'precipitation extremes' in 'precipitation extremes (heavy precipitation events)' -> heavy precipitation events

Changed as suggested

L. 34: see also Ban et al, 2015

A reference to Ban et al., 2015 is added.

L. 38-39: could the authors add references to support this sentence?

We added references to Champion et al., 2011; Zappa et al., 2013.

L. 40: 'statistically': remove

Changed as suggested

L. 40: 'decreasing' -> 'refining'

Changed as suggested

L. 45: these papers are among many others (e.g. Delworth et al, 2012; Kinter et al, 2013; Roberts et al, 2018 and references therein)

We added these references.

L. 47: please also refer to more recent studies

We added references to Dai 2006; Stratton and Stirling, 2012; Gao et al., 2017

L. 53-54: Please be careful not to suggest that climate change response in RCM versus GCM may be solely due to resolution. They also depend on the forcings. For example, Boe et al, 2020 and Gutierrez et al, 2020 show the impact of different aerosol treatments between GCM and RCM that may explain part of the different climate change response.

Thanks for pointing this out. We added the sentence: "Differences in the treatment of aerosols are also identified as a reason for differences in climate response between RCMs and GCMs (Boé et al., 2020; Gutiérrez et al., 2020)."

L. 61: please add a reference

We added a reference to Iorio et al., 2004

L. 62: check the study by Vergara-Temprado et al, 2019. They show that it is possible to turn off convection scheme at such resolution and get appropriate results.

These are interesting results, but don't change the fact that most simulations on 10 km parameterize convection. We changed the sentence to: "Even at grid spacings of around 10 km convection is usually not resolved by the model dynamics but is instead parameterized (although it might be possible to turn off the parameterization already at this kind of resolution (Vergara-Temprado et al., 2019))."

L. 63: 'certain': which ones?

Mainly the diurnal cycle. We changed the sentence to: "However, models with parameterized convection often exhibit common biases in the diurnal precipitation cycle"

L. 66: 'giving' -> 'simulating'

Changed as suggested

L. 68: that is true for models with parametrised convection, please also refer to Vergara et al, 2019 (also in L. 71).

We change the sentence to: “A deficiency of parameterized convection is that it starts too early (e.g. Dai and Trenberth, 2004; Dai, 2006; Brockhaus et al., 2008; Vergara-Temprado et al., 2019).” And also add a reference to Vergara-Temprado et al., 2019 on L.71.

L. 77-78: 12km is not high resolution for RCMs, it is its new standard resolution within CORDEX

What we refer to here are simulations with “convective permitting resolution” which is <5 km (e.g. Coppola et al., 2018)

L. 79: spell out HighResMIP

Changed as suggested

L. 95: high-resolution PRIMAVERA models are available at higher resolution than 40km at mid-latitude (with is the common referenced latitude), or please specify at which latitude this refers to. I would suggest to use the mid-latitude grid spacing (at 50 degreeN), as it is the mid latitude of the European domain (so comparable to EURO-CORDEX grid spacings). It would be clearer to use the term horizontal grid spacing here.

This was not so much a matter of latitudes, but a writing mistake. Never the less, it's a good suggestion to spell out mid-latitude grid spacing. We change as suggested “The models used in this study are a selection of CMIP5 global models (~100-300 km mid-latitude horizontal grid spacing); the high (~25-50 km mid-latitude) and low (~80-160 km mid latitude) resolution versions of the PRIMAVERA global models and the first models from CMIP6 (~100-300 km); and a selection of CORDEX regional models (at 12.5 and 50 km mid-latitude grid spacing).”

L. 95-97 & Table 1: why not considering the full ensembles? How were the models selected? Why are there 5 PRIMAVERA LR and 4 HR?

We selected the models for which we at the time could get daily precipitation. Since we thought that we got ensembles of reasonable sizes we decided not to track down individual models that were not available in common storages. The Primavera LR and HR ensembles are of different resolutions because HadGEM3-GC31 was run at three resolutions. Only one (25 km) was considered as HR, the other two (60 & 130 km) were considered as LR.

Figures 2-3-5: I refer to the revised figures. What does 'act' mean? Please clarify the x-axis 'precipitation bins' and y-axis 'precipitation contribution' labels.

We have updated these figures, and hopefully the titles and axis annotations are more clear now. The figure labels have also been updated to more clearly describe the figure contents.

Figures 2-3: specify in the caption that the thick lines are for ensemble means, and that the bottom panels are differences with E-OBS.

Thanks. The figure labels have been updated accordingly.

Figures 2-3-4: E-OBS is written in Table 1 to be available at 2 resolutions. Which is shown on these figures?

In these figures we use E-OBS with the highest resolution (0.1 deg). It is now specified in the figure labels.

Figures 6-7-8-9-10: I guess E-OBS is shown here at its 2 available resolutions, which one is which?

Correct. We added: "E-OBS at 0.25° (grey) and 0.1° km (black)."

L. 150: bottom left panel for the Alps. Also, CMIP6 upper end seems to be around 50mm/day and CORDEX HR over 100mm/day.

Correct, bottom left and bottom right was mixed up. The sentence is changed

Figure 3: The spread is much larger in CORDEX than CMIP6 in JJA. It shows that CORDEX is not so sensitive to the GCM boundary conditions but to different parametrisation schemes in JJA. The spread is determined by the min and max values for both EUR-11 and EUR-44. So are these min and max values only represented by 1 RCM, or 1 RCM-GCM simulation? If the spread is represented by min and max values, wouldn't it be better to plot the median instead of the mean?

Indeed, the spread defined by max/min values is very sensitive to possible "outliers" that might not be a good representation of the ensemble spread. It is not entirely clear what the best way would be to indicate the spread of such relatively small ensembles (without the use of more sophisticated statistical techniques like bootstrapping). We have changed from max/min to instead show the 5-95 percentile range. We further agree that median values would be more appropriate than mean values and thus have changed accordingly.

Figure 4: It seems biased to consider EUR-11 as the reference and compare observations to that reference, possibly because, although EUR-11 has a higher resolution, their mean climate seems too wet against high density observations as shown by Demory et al (2020), although I agree observations have undercatch errors. If E-OBS are considered too low resolution and not trustable, considering datasets with higher density stations as the reference would be necessary here. Moreover, the ensembles are clearly compared to each other in this figure, with respect to EUR-11. It would be good to see this analysis performed on a common grid to evaluate how it affects the conclusions. It could be done both at 50km for EUR-11, EUR-44 and PRIMAVERA, and then redone for all datasets at 150 (or even 300km), as done in Torma et al, 2015. Why writing the E-OBS total annual mean in the box if EUR-11 is used as a reference?

We have now included analysis on common grids (at two different resolutions, 0.5°x0.5° and 2°x2°), although not presented in the format as shown in Fig. 4 (see Figures S1 and S2 in Supplementary). The interpolation to common grids of course have an effect but the overall conclusions are not seriously impacted. Further on, as mentioned above, we included another, regional high-quality, data set in an ASoP analysis to emphasize the importance of such data sets and possible impact on the results. Still, we are limited for most regions to the E-OBS data as reference while acknowledging its inherent uncertainties. Regarding Fig.4 your concerns about having EUR-11 as reference is understandable and we have changed to E-OBS as reference instead.

L. 181: more strongly biased lower -> more negatively biased: I suggest not to use the word 'bias' when compared to an ensemble, which is itself biased.

We changed to: "Region total seasonal precipitation (averaged within each ensemble), are either mostly in the range of +/- 20 % from CORDEX HR (e.g. eastern Europe, EA) or with larger negative values..."

L. 193-194: Observations have uncertainties but EUR-11 could also rain too much along coastlines and over topography.

True, we changed to "...both factors contributing to uncertainties in quality and representativeness of observational and simulated data."

L. 211-212: Fig. 5 shows results for the annual mean, so this conclusion may be different at seasonal means (at least between DJF and JJA), so I would suggest to show these seasonally as well. Moreover, the delta in grid spacing between CORDEX LR(50km) and HR (12.5km) is similar for all models (delta=4), so the impact of resolution is potentially more similar (although it depends on models). This is more complex for the PRIMAVERA models that have various deltas between the LR and HR versions. I counted that deltas vary between 2 for most models, 3 for a couple and 5.4 for the HadGEM3 model (<https://www.primavera-h2020.eu/modelling/our-models/>).

Moreover, note that PRIMAVERA HR uses exactly the same tuning parameters as their LR version, so the effect of resolution solely is seen here (this is not the case for the CORDEX ensembles that may use different model versions). Something that could be interesting to show here is whether, depending on their delta in grid spacing, some PRIMAVERA models show larger differences than some others. But I would not generalise, based on ensemble means, that resolution in CORDEX has more effect than resolution in PRIMAVERA. It would be good to see the spread of the ensembles on figure 5.

A good point. We added the sentences: "Some differences between the CORDEX and PRIMAVERA ensembles should be noted. The PRIMAVERA models use the same tuning parameters for both the LR and HR version, but on the other hand the differences in resolution between LR and HR varies between models. The CORDEX ensembles have the same difference in resolution for all models, but the LR and HR simulations may be run with different models versions. Hence, all differences between PRIMAVERA and CORDEX ensembles can't be generalised to be attributed by resolution alone."

We also plotted the absolute difference in the precipitation indices between LR and HR against the ratio LR/HR. It turns out the correlation is weak, e.g. the spread within CORDEX ensemble is large although all models have the same ratio.

In Figure 5 the absolute values for each model have now been included as well (in addition to the ensemble means) showing the ensemble spread.

L. 216-218: I agree with this hypothesis, and yet you found greater differences in CORDEX (driven by same low-resolution GCMs) than in PRIMAVERA (L. 211-212). I think this highlights the need for analyses on a common grid, based on seasonal means, and taking into account the fact that CORDEX and PRIMAVERA have different deltas in grid spacing.

A description of winter and summer is included in the text. Our analysis on common grids and of resolution delta doesn't suggest that this explains the differences. Rather, the conclusion is that for high intensities model resolution and performance is more important

than the driving GCM. We don't know the full answer. This section was also meant to show that there are unresolved issues and to point to possible future studies.

We added the following: "Still, the largest differences are seen in the CORDEX ensemble where the LR and HR models are run with the same coarse resolution GCM. This suggests that (regional) model resolution and performance is what determines high precipitation rates, rather than the driving GCM. "

L. 226-227: this is not a sentence/question: please rephrase.

We changed to: "When do intense precipitation events occur in the high-resolution models? the kind of events that are rarely seen or absent in the low resolutions simulations."

Figures 6-9: I considered the revised figures. I still do not understand why some values are not shown. For example: Fig. 6 top left: For one of the CMIP5, only the 10th and 90th percentiles are shown, nothing else it seems. For some other CMIP5 and CMIP6 models, the boxes are drawn but not the whiskers.

In small regions like the Alps and in models of coarse resolution the number of data points are actually too few to make good statistics. This means that calculation of percentiles can be difficult. Since this only happens in some regions for a small number of models we consider this a major problem.

Figures 6-7: it seems that CMIP5, CORDEX LR and HR have a larger variability, so is the variability of CORDEX driven by the variability of CMIP5? This could be answered by looking at the seasonal means (DJF and JJA).

We don't agree that the variability is large in CMIP5, rather the variability increases with resolution. The signal is the same for the individual seasons, but less pronounced since the potential number of days is smaller when divided over four seasons instead of counted over the whole year.

L. 227-246: Again for these analyses, the metrics can be analysed for each ensemble on their native resolution, but if the ensembles are compared to each other, as written in the text, then the analyses need to be redone on a common grid.

Yes, descriptions of summer and winter are now included and figures of this when relevant. The analysis now also include data on common grids.

L. 232-233: Would it be possible to show this with seasonal means?

Yes, descriptions of summer and winter are now included and figures of this when relevant.

L. 245: isn't it 20 mm/day instead of 10?

Yes, we changed to 20.

L. 271: rephrase 'negative for some models and positive for some' as it reads too vague

We changed to: "The differences are small, mainly within ± 10 days year⁻¹."

Figure 10: This intercomparison needs to be performed on a common grid

This is now done.

L. 281: left -> right

Changed as suggested

L. 283: right -> left

Changed as suggested

L. 283-284: Note that ECMWF HR is 25km grid spacing output at 50km, and LR is 50km output at 100km grid spacing. The delta in grid spacing is therefore 2, and the output are regridded to coarser resolution. This may impact the results.

We added a new Fig 12 showing the correlation between difference and delta.

L. 300-301: Demory et al have revised the manuscript with a focus on EUR-11.

Thanks for pointing that out we change to: "In a similar study Demory et al. (2020) compares PRIMAVERA models with CORDEX LR and CORDEX HR."

L. 306: give extremes that are heavier and more frequent -> simulate more intense and more frequent heavy precipitation. I would avoid the term 'extremes' with such low-resolution models, and refer instead to 'heavy' or 'intense'.

We change to: "They conclude that high resolution models systematically give intense precipitation that is heavier and more frequent."

L. 308: overestimation compared to E-OBS

Changed as suggested

L. 315: CMIP6 and CMIP5

Changed as suggested

L. 318-319: this is probably particularly the case for JJA (as shown in fig 3), but for this conclusion it would be good to see DJF and JJA for fig 6-9.

Information about DJF and JJA for Figs 6-9 are now included in text or in the supplementary.

L. 320: not only. E-OBS is not based on the full network of rain gauges over some other countries, such as France.

We added: "E-OBS is not based on the full network of rain gauges in all countries, which could also lead to undercatch."

L. 332: scale -> grid

Changed as suggested

L. 336: will have -> has

Changed as suggested

L. 340-341: yes but PRIMAVERA tends to be drier than CORDEX in all seasons.

We changed to: "...furthermore GCMs and RCMs of comparable resolution simulate comparable precipitation climates, even though PRIMAVERA is often drier than CORDEX."

L. 343: to -> too

Changed as suggested

L. 344: agree -> agrees

Changed as suggested

L. 345: the quantification can be done if performed on common grids

For the individual models it is possible on common grids, which we now do. On the ensemble level it's more difficult. By difficult we mean that it's not so obvious how resolution influences the ensemble mean because the actual model members used impact the ensemble mean more than the resolution of the members.

L. 348-349: this can depend on seasons

Yes, we added the following to the end of the sentence: "especially for heavy precipitation and particularly in summer."

L. 350-351: this needs to be rephrased, as Demory et al have evaluated CMIP5, CORDEX LR/HR and PRIMAVERA HR

We changed to: "The results presented here are in line with previous similar studies using different methods (Demory et al., 2020; Iles et al., 2020) To these studies details are added about the spread between resolutions and between models."

Proper acknowledgement needs to be given to the PRIMAVERA, CORDEX, and CMIP modelling groups

Changed as suggested

There are several typos in the text, please check carefully (e.g. L.8: effects -> affects; L. 20: in depending -> depends; L. 180: region -> regional (and remove comma afterwards); L. 218: were -> where; many others)

These and others are corrected. We apologise for the lack of proof reading, as a reviewer it's annoying to have to correct typos.

L. 139 and 141: below/above c: are these typos?

We removed the "c" for circa as it only confuses.

References:

*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, <https://doi.org/10.1002/2014GL062588>, 2015.

*Boé, J., Somot, S., Corre, L., and Nabat, P.: Large differences in Summer climate change over Europe as projected by global and regional climate models: causes and consequences, *Clim. Dynam.*, 54, 2981–3002, <https://doi.org/10.1007/s00382-020-05153-1>, 2020.

*Delworth TL, Rosati A, Anderson W, Adcroft AJ, Balaji V, Benson R, Dixon K, Griffies SM, Lee HC, Pacanowski RC, Vecchi GA, Wittenberg AT, Zeng F, Zhang R (2012) Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J Clim* 25:2755–2781. doi:10.1175/JCLI-D-11-00316.1

Fernández, J., Frías, M. D., Cabos, W. D., Cofiño, A. S., Domínguez, M., Fita, L., Gaertner, M. A., García-Díez, M., Gutiérrez, J. M., Jiménez-Guerrero, P., Liguori, G., Montávez, J. P., Romera, R., and Sánchez, E.: Consistency of climate change projections from multiple global and regional model intercomparison projects, *Clim. Dynam.*, 52, 1139–1156, <https://doi.org/10.1007/s00382-018-4181-8>, 2019.

*Gutiérrez, C., Somot, S., Nabat, P., Mallet, M., Corre, L., van Meijgaard, E., Perpiñán, O., and Gaertner, M. A.: Future evolution of surface solar radiation and photovoltaic potential in Europe: investigating the role of aerosols, *Environ. Res. Lett.*, 15, 034035, <https://doi.org/10.1088/1748-9326/ab6666>, 2020.

Hall, A.: Projecting regional change, *Science*, 346, 1460–1462, <https://doi.org/10.1126/science.aaa0629>, 2014.

*Kinter III JL, Cash B, Achuthavarier D, Adams J, Altshuler E, Dirmeyer P, Doty B, Huang B, Jin EK, Marx L, Manganello J, Stan C, Wakefield T, Palmer T, Hamrud M, Jung T, Miller M, Towers P, Wedi N, Satoh M, Tomita H, Kodama C, Nasuno T, Oouchi K, Yamada Y, Taniguchi H, Andrews P, Baer T, Ezell M, Halloy C, John D, Loftis B, Mohr R, Wong K (2013) Revolutionizing climate modeling with project Athena: a multi-institutional, international collaboration. *Bull Am Meteorol Soc* 94:231–245. doi:10.1175/BAMS-D-11-00043.1

Klaver R, Haarsma R, Vidale PL, Hazeleger W. Effective resolution in high resolution global atmospheric models for climate studies. *Atmos Sci Lett*. 2020;21:e952. Na, Y., Fu, Q., & Kodama, C. (2020). Precipitation probability and its future changes from a global cloud-resolving model and CMIP6 simulations. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031926. <https://doi.org/10.1029/2019JD031926>

Prein, A. F., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., FoxMaule, 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, <https://doi.org/10.1007/s00382-015-2589-y>, 2016.

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

*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.

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, <https://doi.org/10.1002/2014JD022781>, 2015.

Vautard, R., Kadyrov, N., Iles, C., Boberg, F., Buonomo, E., Coppola, E., Bülow, K., Corre, L., van Meijgaard, E., Nogherotto, R., et al.: Assessment of the large EURO-CORDEX regional climate simulation ensemble, *J. Geophys. Res. – Atmos.*, in press.

*Vergara-Temprado, J., Ban, N., Panosetti, D., Schlemmer, L., and Schär, C. (2019). Climate models permit convection at much coarser resolutions than previously considered. *J. Clim.*, JCLI-D-19-0286.1. doi:10.1175/JCLI-D-19-0286.1.