Answer to SC#1

General response:

The authors would like to thank Dr. Suarez-Gutierrez for the time and effort put into this review. These comments have been useful in improving our manuscript. We have carefully read the comments and provide a detailed response to each comment.

I write this comment because I want to address some aspects that I believe the authors should consider, and would significantly improve the quality and usefulness of their findings.

1. I find there is some lack of context of how the simulations used in this paper compare to those widely used in previous studies. It would be very beneficial to show some type of evaluation of the simulated climate in the two ensembles used here, for example in terms of mean temperature and precipitation anomalies and changes at 1.5 and 2 degrees with respect to preindustrial. Moreover, the authors should describe how the simulated climates in these ensembles compare to those in previous studies; e.g.., Perkins-Kirkpatrick et al, 2017; Sanderson et al. 2017; King et al. 2017; Suarez-Gutierrez et al., 2018; Wehner et al, 2018. These studies are based on a variety of types of ensembles, from CMIP5 to fully coupled ESM large ensembles, and the paper should include a discussion on how these simulations differ in terms of both climate conditions and fundamental design. In particular highlighting both the advances (i.e. higher resolution, targeted to 1.5 and 2 degrees) and shortcomings (i.e. atm only runs, no fully coupled ocean, SST prescribed from short periods) of these data is in my opinion crucial.

Answer: We agree that the inclusion of an evaluation will improve the paper and put the results into context. We will add a discussion of the model chain performance as a supplement to the paper. This will also allow us to put our results into perspective compared to results from other studies. We think a full discussion on the benefits of coupled vs. uncoupled GCM simulations would be beyond the scope of this paper. In this context, we see ourselves as users of a widely discussed dataset. We compare ourselves to other downscaling activities such as EUROCORDEX.

2. The paper does not address the implications of using atmosphere only runs with prescribed SST based on relatively short time periods sufficiently. A finite set of prescribed SST patterns offers a limited range of climate states that does not completely sample ocean-driven variability (see Sanderson et al., 2017; Fischer et al., 2018). In contrast, large ensembles from fully-coupled climate models sample a wider range of ocean states and include the influence of the ocean-borne variability (Hawkins et al., 2016). Furthermore, fully-coupled large ensembles also offer a more realistic representation of heat extremes over land than atmosphere-only large-ensembles, even if the later offer a larger number of independent simulations (Fischer et al., 2018). These issues should be addressed in the main text.

A: We agree that we should mention these shortcomings and we will include this. A full discussion on coupled vs. uncoupled GCM simulations would be beyond the scope of our paper (see answer above).

3. The authors argue that the improved resolution from using a regional model combined with the large ensemble size are mayor improvements. However, previous studies analyze the changes at 1.5 and 2 degrees using similarly large ensembles of fully coupled ESM (100 members x 250 years; Suarez-Gutierrez et al., 2018), so what are the differences or biases that higher resolution vs. no coupled ocean introduce?

A: We argue that extremes can be better estimated using our downscaled information, which has been shown several times in many studies, especially for precipitation. As mentioned earlier, our focus is on downscaling an existing dataset (in this case HAPPI) with a well-established method (dynamical downscaling with a regional climate model). We think the question raised here, is very much related to the comments above and beyond the scope of our paper.

What do the gray areas over land in figures 2, 3, and 4 represent? I thought maybe the white shading was meant to be transparent but there is white in some parts I think?

A: The grey areas over land represent non-significant changes. This was mentioned in the text and will be added to the figure captions (see also our response to referee#2).

Answers to Referee#1

General response:

The authors would like to thank the referee for the time and effort put into this review. These comments have been useful in improving our manuscript. We have carefully read the comments and provide a detailed response to each comment.

I think Laura Suarez-Gutierrez raises some valid points about how the results presented here relates to previous studies and the impact of the prescribed SSTs (which the authors themselves describe as “unrealistic”). I trust the authors to properly respond to that, so I wont go into that more here.

Answer: We have responded to each of Laura Suarz-Gutierrez’s comments and will revise the text to indicate how our results compare to earlier studies. Please see our responses to her comments. We did not describe the prescribed SST as “unrealistic”. In fact, for the current period they are even based on observations. Due to an interpolation issue with the anomalies used in the future period and a special treatment of sea ice at a given SST in the ECAHM model, unrealistic SST jumps occurred in a very few grid boxes. This was what we were referring to and we hope that the new section 2 (see below) makes this now more clear.

Is this paper a presentation of a data set or a presentation of results? The title suggests results, but the abstract starts “This paper presents a novel data set” and the Discussion “A unique data set has been presented”. It’s of course fine to do both, but a data description paper would require a lot more information about models, time periods, scenarios etc. I’m not sure that I agree that a data set is properly “presented” here.

A: The intention of this paper is to present a new dataset and show some examples of its applications. We propose to extensively rewrite the Methods section (please refer to response to reviewers next two questions, below) to provide a more extensive description of the model experiments, as the referee suggests.

Describe the model experiments in more detail. Why did you use 10 year periods? And why 20 years for the pre-industrial period? How are the specific warmings levels (SWLs) for +1.5 and +2 defined and calculated? You say that you use both RCP2.6 and RCP4.5. Do you mix them in the SWLs? What is the ratio RCP2.6/RCP4.5? I guess that this is described in some HAPPI paper, but it’s worth to spend a few lines on that also here.

A: Indeed, the experiment specification is described in other papers from the HAPPI community. We will update section 2 (see our response to the next comment, below) to better explain the HAPPI protocol.

10 year periods are short in a climatological sense, how is the choice of 10 year periods motivated? One could, of course, argue that with enough ensemble members natural variability will be sampled anyway; however, 10 years with 100 members equals 1000 simulated years which corresponds to 33 members simulating 30 years. 10 years C2 times 25 members equals 30 years and 8.3 members. A 9 or 35 member ensemble does not sound as impressive as a 25 or 100 member ensemble. Don’t make to bold statements about the size of the ensemble. Furthermore, you don’t explicitly say how many members the RCM model consists of. You say that the GCM ensembles have 25 and 100 members respectively, but you don’t actually say that you downscale all of them. Not as far as I can see at least. Overall I think section 2.1 could be rewritten in a clearer way first

presenting the HAPPI project and the GCMs then the RCM and the GCM-RCM combinations, time periods, etc. As it is now it's a bit of a mixture where the reader has to go back and forth to get it all.

A: We will address the choices made in the HAPPI project in a dedicated sub-section, as all referees had questions on the experiment set-up as described in our paper. Therefore, we will restructure the Methods section (2.) in the following manner:

"2 Methods

To create a data set for regional climate impact studies for Europe under 1.5°C and 2.0°C global warming, the regional climate model REMO has been used to dynamically downscale two GCM ensembles following the HAPPI experiment protocol by Mitchell et al. (2017). Several common climate indices are computed, to demonstrate the usefulness of the data.

2.1 Global HAPPI simulations

The HAPPI protocol by Mitchell et al. (2017) has been set up to inform the IPCC Special Report on 1.5°C Warming. Large ensembles (>50 members) of GCM simulations where created that allow extreme events to be studied, even for the small differential warming between a current decade (2006-2015) and two future decades under 1.5°C and 2.0°C global warming. The major aspects of the protocol for the HAPPI experiment are summarized in the following paragraphs, as there are important differences compared to the typical CMIP protocols.

All simulations were conducted in atmosphere-only mode in order to increase ensemble size and provide more accurate regional projections (Mitchell et al., 2017; He and Sodan, 2016). The simulation period for all members is limited to 10 years, because during the current period from 2006-2015 sea-surface temperatures stayed approximately constant. This period forms the basis of the entire experiment and allows for a better estimate of, e.g., return-values from this period compared to periods with a strong warming trend. The experiment design for the current decade follows the DECK AMIP protocol using observed sea ice and SSTs. For the future periods, SSTs are calculated by taking the 2006-2015 observed conditions and adding a SST anomaly representing the future periods.

By chance, the multi-model averaged CMIP5 global mean temperature response between 2091-2100 compared to 1861-1880 under RCP2.6 is 1.55°C. Mitchell et al. (2017) considered this warming as sufficiently close to inform about impacts under 1.5°C and chose this period under RCP2.6 as basis for a 1.5°C warmer period. The SST anomalies for the 1.5°C period were computed using the modelled decade averaged difference between 2091-2100 from RCP2.6 and 2006-2015 from RCP8.5, because RCP8.5 averaged SSTs over this period are closest to observations. Forcing values for anthropogenic greenhouse gases, aerosols and land-use are taken from the year 2095 of RCP2.6 and kept constant during the simulation. Because of the poor representation of sea ice in the CMIP5 models, Mitchell et al. (2017) used a different approach to construct sea ice concentrations for the 1.5° period. A detailed description would be beyond the scope of this paper and can be found in the cited reference.

The SST anomalies for the 2.0°C period cannot be calculated following a similar approach as for the 1.5°C period, because none of the RCPs show a global mean temperature response close to 2.0°C at the end of the century. Therefore, a weighted sum of RCP2.6 and RCP4.5 is calculated with a global mean temperature response of 2.05°C, which is exactly 0.5°C more compared to the chosen 1.5°C period. The calculation of SST anomalies and sea ice extend follows the same methodology as for the 1.5°C period. Mitchell et al. (2017) decided to apply their weighting method only to the well-mixed greenhouse gases, because the land-use changes and aerosols show very different spatial patterns and are therefore kept at the 1.5°C period values.

2.2 Regional HAPPI simulations

In order to create high-resolution climate data for Europe from HAPPI, the RCM REMO has been used for downscaling. REMO is a hydrostatic limited-area model of the atmosphere that has been extensively used and tested in climate change studies over Europe (Jacob et al., 2012; Teichmann et al., 2013; Kotlarski et al., 2014). The simulation domain follows the CORDEX specification for the standard European domain with 0.44° horizontal resolution. The European CORDEX domain for REMO covers 121x129 grid boxes. To exclude the sponge zone, where the REMO simulations are relaxed towards the GCM solutions, a core domain of 106x103 grid boxes, following the CORDEX definition, is used for the analyses. In the vertical, 27 levels are used without nudging except for the boundaries. Boundary conditions are taken from the HAPPI Tier1 experiments (Mitchell et al., 2017), which are carried out with ECHAM6 (Stevens et al., 2013; Lierhammer et al., 2017) (100 members per period) and NorESM (Bentsen et al., 2013) (25 members per period) that provide 6-hourly 3-dimensional data for downscaling. In REMO the same greenhouse gas forcings as for the GCMs were used.

SST and sea ice concentrations were taken directly from the GCM output matching the GCM land-sea mask for NorESM. From ECHAM6 only the sea ice concentrations were taken. Due to the interpolation procedure for the sea ice extent, it could happen that sea ice was artificially created where no ice conditions were present in the original dataset, e.g., during summer in the Baltic Sea. ECHAM6 has a mechanism that as soon as there is a fraction of sea ice greater than zero, the SST is limited to a maximum of 272.5K. This leads to artificial temperature jumps in the SST between adjacent grid boxes as soon as erroneous sea ice appeared in one of the grid boxes. In order to avoid inheriting this issue, the originally provided SST fields from the HAPPI project were used for the REMO simulations, using ECHAM6 as forcing GCM. After testing different temperature and/or sea ice fraction thresholds, the authors decided to keep the original sea ice maps, because in the cases where artificial sea ice was created the fraction was typically well below 1%, and only in rare cases reaches up to 4% (not shown). All other procedures would have removed too much sea ice in other seasons or led to unrealistic gradients of sea ice fractions. With the tile approach of REMO the effect of the artificial sea ice on the averaged near-surface variables is hardly detectable.

For each GCM member only one REMO simulation was carried out, as inter-member variability of an RCM ensemble over Europe on a time scale of 10 years is small compared to the internal variability of a GCM (Sieck et al., 2016). Each simulation covers a period of ten years, and as such, initial conditions for the lower boundary need to be in balance with the RCM's internal climate in order to avoid artificial drifts in the modelled results. To achieve this, for each driving GCM, the first year of a random GCM member was simulated five times with REMO using initial conditions from the end of the previous run, creating one initial soil temperature state for every ensemble member in one period. This was performed for each of the three periods. Tests showed that this minimizes drifts in the deep soil climatology compared to initial conditions taken directly from the GCM (not shown)."

The rest of section 2 in the manuscript will be kept, but section 2.2 will become section 2.3 (except for changes responding to the specific comments).

Lierhammer, L., Mauritsen, T., Legutke, S., Esch, M., Wieners,K.-H., and Saeed, F.: Simulations of HAPPI (Half a degree Additional warming, Prognosis and Projected Impacts) Tier-1 experiments based on the ECHAM6.3 atmospheric model of the Max Planck Institute for Meteorology (MPI-M), http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=HAPPI-MIP-global-ECHAM6.3, 2017.

He, J. and Soden, B. J.: The Impact of SST Biases on Projections of Anthropogenic Climate Change: A Greater Role for Atmosphere-only Models?, Geophys. Res. Lett., 43, 7745–7750, 2016.

Why did you choose these particular indices? And why do you do use different statistical methods to analyse them? The choice of methods and the ways to present the results seem a bit arbitrary.

A: Our intention was to show example applications for the regional HAPPI dataset. Two main aspects of the dataset are the high resolution and the large number of ensemble simulations. These aspects allow one to focus on other indices beyond mean annual warming. We have chosen to cover indices related to extreme temperatures and precipitation. These indices carry direct relevance for several applications in health and water management, two sectors directly affected by weather, and thus these indices are not arbitrary. The presented indices are highly accepted indices for extremes. The choice of the indicators are supported by the mentioned reference from WMO and others. Obviously, many other indices can be chosen, and the dataset also allows further analysis by third parties.

A suggestion for improving section 2.2: Remove the bullet points with indices. It’s a bit strange when the indices are listed together with some kind of motivation or definition, but in a different way for each index. Instead just list the names of the indices. Then, have a sub-heading for each index under which you properly explain the definitions and motivations behind each index.

A: We will adjust the text, and remove the bullets and add separate headings for each indicator.

To what degree would you say that you are showing the value of large ensembles? You mention a smaller sub-ensemble, but I can’t see it in the analyses. Sure, you compare the NorESM and ECHAM6 forced ensembles, but how can you know that all differences between the ensembles are due to the ensemble size and not the models themselves?

A: The value of the large ensemble is clearly given when looking at precipitation. In the manuscript, we compare 25 NorESM members to 100 ECHAM6 members. We find more robust results in terms of spatial patterns when using the 100 ECHAM6 driven members. An analysis of a sub-sample of 25 ECHAM6 driven members confirms this and we will add this analysis as a supplement. Of course there will still be a difference between ECHAM6 and NorESM, which are then model related.

Think about how you want to name the SWLs. “1.5 C period”, “temperature target 1.5 C simulation” (bulky), “2.0 C increase in GMT” or something else. It’s a bit annoying when different names are used at different places in the text.

A: We agree that our style to refer to the periods is misleading. We will harmonize this and change it to “current period”, “1.5°C period” and “2.0°C period”.

SPECIFIC COMMENTS

L20: This sentence (especially the first line) is incomprehensible to me. Think about what you want to say, then write it in two, or even three, sentences. Long sentences with few commas has no intrinsic value.

A: We will rephrase the sentence as follows:

“Identifying regional climate change impacts for different global mean temperature targets is increasingly relevant to both the private and public sectors. In the private sector, investors demand financial disclosure associated with climate change risks and opportunities (Goldstein, et al., 2018). In the public sector, policy makers rely on climate information build on internationally agreed limits to develop national climate action policies.”

L26-28: One could add to this that even though it’s relevant to work with SWLs the choice of RCP can have an impact on the simulated SWL climate (e.g. Bärring & Strandberg, 2018). This should be interesting to you since I suppose that you mix RCPs in you index calculations.

A: Given that the HAPPI protocol is different from time periods associated with RCPs by construction, this should not be relevant to our indices. We will add a sub-section on HAPPI which should make this clear.

L30: “5 to 15 models available”. I had a quick look at ESGF and found ca 70 members from 30 models from 13 model families. I don’t think it’s fair to describe the CMIP5 archive in such a diminutive way.

A: We are only referring to the cited studies. These studies used 5 to 15 models. This is not a general statement on the number of simulations available on ESGF at that time. We hope this re-formulation will make it more clear:

“These studies typically used the 5 to 15 ensemble members which were available in CMIP5 at the time for their global and regional studies.”

L58: How where the +1.5 +2 SWLs calculated, please explain.

A: This will be clarified in an additional sub-section on the HAPPI method.

L59: Why is the pre-industrial period 20 years when the other periods are 10 years?

A: The pre-industrial period is only used as a baseline to define the period with 0°C warming. The definition is coming from the HAPPI protocol that is followed by every group doing global simulations. We are aware that there are several slightly different definitions of this particular period. The IPCC Special Report on 1.5°C lists several of these definitions. We will add a new sub-section on the HAPPI experiments to the manuscript (see previous answerers) to make these definitions more clear. Also, the period of ten years comes from the HAPPI protocol and has been discussed by Mitchell et al. (2017). We will add the motivation for the ten year period to the HAPPI sub-section.

L61: “greenhouse gas forcing is constructed from RCP2.6 and RCP4.5” What do you mean by “constructed”? Don’t you just use the forcing data from the RCPs?

A: This is again related to the HAPPI protocol. We are not using boundary forcing from classical RCP driven global model simulations, but AMIP style GCM simulations following the HAPPI protocol. The sub-section on HAPPI will make this clearer.

L61: “RCP2.6 and RCP4.5” Do you use both and mix them in the SWLs? How many of the 100 (25) members are from RCP2.6 and RCP4.5 respectively?

A: The dataset is not from mixed RCP2.6 and RCP4.5 simulations. See answer above.

L69: The use of “per period” is a bit confusing. Isn’t it enough to just state the number of models?

A: "per period" refers to the three simulation periods (current, 1.5°C, 2.0°C) following the HAPPI protocol mentioned in line 58. Either one has to write 300 and 75 members or the number of members per period. That latter seemed more intuitive to us, especially when we do analysis on simulated slices instead of continues simulations.

L73: "For each GCM member" Are these all of the 125 members?

A: This refers to all 375 simulations done with REMO. We did not construct a regional ensemble using only one GCM member.

L74: "These time scales" What times scales?

A: On a time scale of 10 years. We will change the text to:

"For each GCM member only one REMO simulation was carried out, as inter-member variability of an RCM ensemble over Europe on a time scale of 10 years is small compared to the internal variability of a GCM (Sieck et al., 2016)."

L75: "RCMs" → "RCM's"

A: Will be changed accordingly.

L94: "recommend" → recommended

A: Will be changed accordingly

L97: (or L105-109) Please explain a bit more about apparent temperature. Why is it apparent? Why doesn't it always occur?

A: The term 'apparent temperature', defined by Davis et al., 2016, is a function of both temperature and dew point temperature. The dew point temperature accounts for humidity in the air. The combination of temperature and humidity is more relevant to human health than temperature alone.

L109: "Relative change" I guess you look at the change in all indices, I don't get why you specify this explicitly here.

A: Relative change in L97. The reference to 'relative change' will be removed.

L 103: "and NorESM has" → "and the NorESM driven have"

A: Will be changed accordingly.

L115-116: This sentence is incomprehensible to me. Think about what you want to say, then write that in two, or even three, sentences. Long sentences with few commas has no intrinsic value.

A: We will change the sentence to:

"A change in extreme precipitation directly influences local communities. Such communities have applied design standards for structures to withstand floods with a specified return period. These return standards will no longer be applicable when the extreme value distribution shifts with global warming."

L116: Remove "As such", this is not the correct use of that term.

A: We will delete "as such" accordingly.

L118: "exceedance probability" Isn't rainfall events rather associated with either a probability or a threshold. Maybe I just don't understand what you mean.

A: Exceedance probability for rainfall events is associated with engineering practices. In probability theory, an event can be characterised by either probability of non-exceedance, or exceedance. Here, we use the exceedance probability. We will adjust the sentence, to underline this.

L118: " rainfall intensity for the 50-year return" Do you mean rainfall intensity with a 50-year return period?

A: Yes, we mean a 'rainfall intensity with a 50-year return period'. All references to this have been changed accordingly.

L119: "Such information" What kind of information?

A: This refers to the sentence before where "rainfall intensity with a 50-year return period is computed". We will adjust the sentence to "Information on changes in the rainfall intensity with a 50-year return interval …".

L121: "between 100 and 100 years" I guess that at least on of the "100" should be another number.

A: This is a typo and should state "between 10 and 100 years"

L122: What is your definition of CDD? Is it the longest period of consecutive dry days, or something else? Is it the longest period over the whole 10-year period or is the annual average for all 10 years? What is the threshold for a dry day (1 mm?)?

A: As we explain in the paper, the threshold is less than 1 mm per day (lines 123-124). We calculated the maximum duration for the entire 10 years of each ensemble member. We will add this to the text.

L122 will be rephrased:

"Lastly, the Consecutive Dry Days (CDD), defined as the maximum number of consecutive days with a daily precipitation amount of less than 1mm over a region (Karl et al., 1999; Peterson et al., 2001) is calculated for the entire 10 year period of each ensemble member. The CDD is calculated for each of PRUDENCE regions (Christensen et al., 2007), illustrated in Figure 1, because drought indicators are relevant over large areas."

L122: Why do you analyse CDD for the Prudence regions and not in a map as with the other indices? Or, why don't you do the thorough analysis that you do for CDD for the other indices?

A: The CDD analysis is computed for the Prudence regions and not 'per-grid-box' as for the other indices used in this study, because applications drought indices are relevant over larger areas, whereas in the cases of the other indices considering high temperatures and heavy precipitation, analysis of individual grid-boxes are more relevant as they have more local applications.

L130: "historical" Is this pre-industrial (1861-1880) or current (2006-2015)?

A: This is "current" and will be changed accordingly.

L131-132: "differences of the /.../ percentiles were computed by subtracting the ensemble mean" Isn't the difference in any percentile calculated by computing the difference between the percentile for one period with the same percentile for another period? I think this sentence could be made more understandable.

A: Yes, you are right. We will change this to:

"In case of ATG28, differences of the 5th, 50th, and 95th percentiles were computed by subtracting the respective percentiles of the current decade from the projected periods."

L132: "areas" What kind of areas? Isn't it done per grid point?

A: Yes, the calculation is done on every grid point. The data points are referred to number of exceedances in the current period. The text formulation is changed to read:

"Only grid boxes with more than 20 exceedances over threshold in the current period were included in the analysis of ATG28 in order to allow for confidence interval calculations for the shown percentiles using order statistics."

L135: How is the “percentile confidence range” defined?

A: This is explained in line 134. We use order statistic to compute the confidence range for the percentiles.

L138: Why do you choose the Mann-Whitney-U-test?

A: We chose the Mann-Whitney-U-test, because it is a non-parametric test, which does not require any underlying statistical distribution.

L140: “precipitation intensity of the 50-year period” I think you mean “precipitation intensity with a 50-year return time”.

A: We stated in the paper “intensity of the 50-yr return period”, but in order to be more clear we will rephrase to “precipitation intensity with a 50-year return time”.

L141: “historical” Is this pre-industrial (1861-1880) or current (2006-2015)?

A: This is current and will be changed accordingly.

L141: “NOResm” → “NorESM”

A: Will be changed accordingly.

L141: Why do you explicitly mention the model names here? I expect you to do analyse both ensembles for all indices. It’s implicit that you do.

A: The sentence will be changed to:

“The differences in RI50yr are computed as the relative change in daily precipitation intensity of the 50-yr return period between the 1.5°C and 2.0°C simulations compared to the current period simulations.”

L150: “mean temperature” Please add a “(not shown)” here.

A: This has been added accordingly:

“But also the central and eastern parts of Europe show increases in ATG28, consistent with the increase in mean temperature (not shown).”

L152: “and more in the median around the Mediterranean” Please consider rephrasing to something more understandable.

A: This has been rephrased.

“Around the Mediterranean the increase in ATG28 during the 1.5° C period is mostly moderate with up to 9 days in the median whereas changes in the 2.0°C period are reaching 18 days and more.”

L154: “no change in the distribution of ATG28” Based on figs 2 & 3 I don’t agree. For +2 in central Europe the 5th percentile doesn’t seem to change much, while the 95th percentile increases with around 6 C. Isn’t that a change in the shape of the distribution?

A: Actually, it is a shift of the distribution; we have little reason to think that the shape of the distribution is changing. This is because while we agree that the 5th percentile for ATG28 does not change much in the simulations at 1.5 degree for the NorESM model, there are larger shifts at 2 degree, and also at 1.5 and 2 degree for the ECHAM model (Figure 3).

L154:”spatial resolution allows”. It’s of course better than the GCMs, but is it really true? Isn’t the motivation for EUR-011 that EUR-044 doesn’t resolve complex topography?

A: We agree that EUR-11 would be better to resolve complex topography, however, with the current generation of HPC computers, such a large ensemble of RCM simulations would not have been possible to conduct on much higher resolution than 0.44°.

L156"Mediterranean" → "around the Mediterranean" or "Mediterranean region"

A: This has been changed:

"This is especially important in areas with complex topography such as the Mediterranean region, which is usually only poorly resolved in GCM simulations."

Figs 2 & 3. Please consider the following: Add percentile names in a new top row. Add SWL names in a new left column. Add units by the colour bars. Add letters a-f in the caption. Add ensemble sizes in the caption And it seem like white colours are replaced by grey.

A: We will update the Figures by taking into account the comments of all referees. The grey boxes are masked out areas. On land they refer to grid boxes that do not match our criteria of 20 or more occurrences of ATG28 during the current period. This will be stated explicitly in the text and caption.

L169: "the four REMO ensemble experiments" This is a bit ambiguous. How many REMO ensembles are there, 1, 2, 4, 6? Depends on the definition. Consider erasing "four".

A: Yes, it is better to understand without "four". This will be changed to:

"Figure 4 shows the relative differences of RX5day for the REMO ensemble experiments."

L172: "more coherent" More coherent than what? Not with "larger areas".

A: The ECHAM6 driven ensemble shows a more coherent pattern than the NorESM driven ensemble. The sentence will be changed to:

"It can also be seen that the patterns in the ECHAM6 driven ensemble is more coherent than the NorESM ensemble with larger areas showing a significant change."

L172: "more significant" How is significance calculated, and how is it shown in fig 4?

A: This has been mentioned in lines 137-139. Only results at the 95% significance level are shown using a Mann-Whitney-U-test. It is missing in the caption, though. We will add this to the caption and explicitly state that grey shading refers to areas with non-significant changes.

"Figure 4. Relative difference of RX5day (in percent) between current and the 1.5°C period (left column) respectively the 2.0°C period (right column) for the NorESM with 25 members (top row) and ECHAM6 with 100 members (bottom row) driven REMO simulations. Grey shading show areas with non-significant changes on the 95% significance level."

L173: "difference in ensemble size" Between what?

A: this is referring to the difference in ensemble size between ECHAM6 and NorESM driven simulations. We will update the text.

L176:”the interior of the simulation” What is the “interior of the simulation” if not everything apart from the boundaries? This seems to be an unnecessary complicated way to describe where the largest changes are. Also consider changing “simulation” to “domain”.

A: Rephrased:

“Apart from artificial effects due to the boundary conditions, the strongest signal within the core domain appears over the Baltic Sea, with an increase of up to 15% in RX5day under a 2.0°C increase in GMT.”

L179: It should be easy enough to at least roughly test the effect of SST. Just plot it and see how unrealistic it is. Also check the timing of RX5day, is it in winter or in summer? I guess the SST bias works differently in different seasons. In winter it’s probably too warm, in summer too cold. I suggest that you do some kind of check.

A: The formulation unrealistic is a bit misleading. It should better be formulated as “unresolved SST”. If it is over- or underestimated, depends on the particular sub-basin and the surrounding (resolved) SSTs.

Figure 4. Consider the following: Add SWL names in a new top row. Add model names in a new left column. Add letters a-d to the panels. Add units to the colour bars. Explain grey shading in caption.

A: A general makeover of the plots will be made as suggested. The caption texts will be changed accordingly (see answer above).

L190: “To account for” What do you mean with this¿It doesn’t seem to be the correct use of the term “to account for”.

A: Sentence will be rewritten as proposed below (see response to L190ii).

L190: “the relative change in daily rainfall intensity is presented in Figure 5”. No, it’s not. Figure 5 shows the change in the intensity with a 50 year return time.

A: The values plotted in Figure 5 are the relative change in RI50yr given in percent. Relative change, in the case of the 1.5°C simulation, is computed as the difference in RI50yr between the 1.5°C and current period, divided by the RI50yr of the current period.

The sentence will be rephrased as:

“The relative change (in percent) in RI50yr across Europe are presented in Figure 5.”

L191: “In the both the” → “In both the”

A: This will be changed accordingly.

L193: “precipitation intensity of the 50-year period” I think you mean “precipitation intensity with a 50-year return time”.

A: Yes. The sentence will be reformulated to:

“ECHAM6 driven simulations clearly show increases in the 24-hour rainfall intensity with a 50-yr return time, of up to 20% over continental Europe.”

L190-195: It’s seems like you’re struggling with how to describe the precipitation intensity of events with a 50 year return time. Why don’t you just define RI50yr properly and a bit lengthy, and then just stick to RI50yr? That would save you some trouble in writing, and should avoid some confusion for the reader.

A: We agree, and will update the paragraph defining RI50yr and rewrite this paragraph using the terms accordingly.

Fig 5: Who do you suddenly show results for a different domain? Excluding parts of northern Europe and including parts of northern Africa where you don't have data. For consistency, show the same domain in all plots. This domain should preferably be the same as the model domain, unless you have a good motivation for excluding certain areas. Also, consider the following: Add SWL names in a new left column. Add letters a-d in the caption. Replace "Percent" with "%" in the legend. This is perhaps a matter of taste, but common practice is "%" I think.

A: We will update and harmonize the plots by using only one plotting tool for the horizontal plots.

L204: This sentence should start: "Both the 1.5 C and 2.0 C distributions ..."

A: We will change this accordingly

L205: "historical" Is this pre-industrial (1861-1880) or current (2006-2015)?

A: This should be "current" and will be changed accordingly.

L205: I think you can remove "respectively". It doesn't add anything.

A: We will update this sentence to: "The distributions for the 1.5°C and 2.0°C period are compared to the current CDD distribution".

L207-208: "whereas the /.../ distributions" This goes without saying. Consider removing for brevity.

A: We will delete the sentence ", whereas … distributions" here.

Table 1: Why are the p-values suddenly the most important part of the analysis of an index? And why is CDD analysed for the Prudence regions? Please explain.

A: As explained in the response to L122, we compute CDD over the Prudence regions because drought indicators are more important over large areas than on a grid-box. In addition, we are comparing distributions of CDD and therefore we need to apply a statistical test to determine whether those distributions have any significant changed. We will provide an explanation earlier within the paper as to why we analyse the CDD for the PRUDENCE regions (see answer to Line 122).

L215: "longer period of dry days" Be careful how you interpret changes in CDD. You don't define CDD so I can't be sure if you make the correct interpretation. If your CDD is averaged over your 10- year period it could be correct. If your CDD is the longest dry period over the 10-year period it only tells you that the longest dry period will be longer. That doesn't necessarily mean that dry periods on average will be longer.

A: In our response to L122, our formulation of CDD is the maximum consecutive days in 10 years. In this analysis, we are determining whether the shift in CDD distributions are significant. In Figure 6, the percentage of the entire CDD distribution of a given duration is presented.

We reformulated L215:

"Over this region, one can see an increase in duration of the longest dry period and that they occur more often (Figure 6)."

L217: ". . . indistinguishable in the simulations /.../ (Table 1)." I don't understand the this sentence. Please rewrite.

A: We will rewrite to: "In contrast, regions 6 and 8, the Alps and Eastern European region, have changes in CDD distributions that are statistically indistinguishable under 1.5 C and 2.0C compared to the current simulation (Table 1)."

L218: “more frequent” This is not correct. CDD is the length of the longest dry period (at least this is the common definition). If you want to know if dry periods will be more frequent you should study the number of dry periods.

A: By plotting the CDD distributions in Figure 6, we are studying the number of dry periods. Over region 2 one can see that the length of the longest dry period increases as the reviewer states and one can also see that the number of longer CDD increases. The authors argue ‘more frequent’ in this case is correct.

L221: “1.5_C vs. 2.0_C” I think yo mean “1.5_C instead of 2.0_C”

A: We agree, and will change accordingly.

L222: “adaption” Do you mean “adaptation”?

A: Will be changed accordingly

L221-222: I have a few problems with this, somewhat ambiguous, sentence. First, it is not the +1.5_C or +2_C targets that will have an impact on society, it is the climate change. Second, that the climate will be different in the +1.5_C world compared to the +2_C world is obvious. I guess you mean that the change in CDD is not linear so that the extra 0.5_C will have a large impact. Third, why do you point out changes in CDD in region 7 as a particular motivation for adaptation and mitigation? In my view this whole paper is a motivation for adaptation and mitigation as it shows how climate change may change in the future. Consider rewriting.

A: We do not state “an impact on society”, we mean that there is a (positive) impact of a 1.5° target compared to a 2.0° target. Second, we show that the difference also has a measurable impact, given large natural variability for indicators such as CDD. This is not even obvious in these HAPPI simulations for several other regions, so this is in fact not straightforward.

We will rewrite the sentence to: “Thus, one can conclude for region 7, according to these simulations, a lower target of 1.5°C increase in GMT could reduce the length of the maximum number of dry days in this region, compared to a 2.0°C target.”

L225-226: It’s a poor motivation to exclude regions just because the U-test gives different results. Especially since the results differ also in region 3. Strictly speaking the results differ for all regions since you get different p-values (Table 1). With the possible exception of IP at +2.0 where both ensembles get 0.000.

A: We will delete this sentence, as we already provided motivation for selecting regions 2, 3 and 7 at the start of the paragraph.

Fig 6: It’s very odd to measure the number of days in the unit weeks. Add to the caption something like: “for the ECHAM (top row) and NorESM (bottom row) driven ensembles”.

We will add the suggested sentence to the caption.

L235: “10 x 100 years” I would prefer “100 x 10 years” since it is 100 10-year simulations.

A: This will be changed accordingly.

L250: “smaller sub-ensemble driven by ECHAM6” I don’t see this sub-ensemble anywhere in the text. Should it be added to the analysis?

A: We will add this analysis as supplement material.

Answers to Referee#2

General response:

The authors would like to thank the referee for the time and effort put into this review, which has been useful in improving our manuscript. We have carefully read the comments, and provide a detailed response to all comments, below.

The paper investigates the impacts of 1.5°C and 2.0°C global warming on temperature and precipitation extremes over Europe. With this aim, the authors use an ensemble of dynamically downscaled simulations from the HAPPI project. The analysis focusses on four climate indices from ETCCDI.

The paper covers an interesting and relevant topic that could be a useful addition to the literature. Unfortunately, several aspects in terms of methods, analyses and results are unclear and need further explanation. In fact, the manuscript needs a lot of improvement in order to increase the clarity and readability. Please refer to the main points and specific comments below. In particular, the authors do not provide enough convincing evidence to prove the benefits of their method. Nevertheless, I suppose that there are sufficient arguments for it. Therefore, I suggest a major revision for this manuscript.

Answer: We will revise the introduction and methodology sections to emphasize the paper's objective of introducing a new dataset. In addition, we will increase the clarity and readability, by improving several sentences in response to the comments we received from this referee and the others. The method that the referee mentions, was developed by the HAPPI consortium. We will insert a more extensive description and motivation for this approach, thereby providing a better background. Finally, we also will provide more details on the assumptions and implications of using the HAPPI GCM simulations for regional downscaling using the regional climate model (RCM) REMO. We hope this will satisfy these main concerns of the referee.

MAIN POINTS:

1) Text: The whole manuscript needs a thorough proofreading and language check. Some sentences are incomprehensible; others are just too long and should be split in two to enhance readability. In addition, several phrases/descriptions are not consistent throughout the paper and therefore may confuse the reader. Please also refer to specific comments below.

A: The manuscript will be checked for language and revised to improve the flow, to include more detailed and consistent descriptions, and reduce sentence lengths. We will revise the introduction and methodology sections to emphasize the paper's objective of introducing a new dataset. Also, we use more consistent definitions for simulation periods, and other descriptions, to improve clarity.

2) Figures: The figures and their captions need some general improvement.

A: All Figures and captions will be improved as suggested below.

2a) Why are some land areas in Figures 2-4 grey?

A: The grey boxes are masked out areas. On land they refer to grid boxes that do not match our criteria of 20 or more occurrences of ATG28 during the current period. This will be stated explicitly in the text and caption in the revised paper.

2b) You should use the same format for all spatial plots, e.g. * One colour bar including units * Label individual plots * Same domain * Add more information in the figure itself (warming level, ensemble, . . .).

A: We used different plotting tools and will redo the plots with one tool following the suggestions.

2c) Figure 6: The bars are difficult to distinguish. Non-overlapping bars might be preferable. The distributions for both GCMs are very different. For NorESM, they are quite narrow, while they are much broader for ECHAM6. What are the implications of such large discrepancies? Please discuss. Why do you measure consecutive dry days in weeks?

A: This figure will be redone with non-overlapping bars. With regard to the different distribution widths, there seems to be a dependency on the forcing model. In our investigations, we see NorESM forced ensemble simulate wetter conditions, whereas ECHAM6 forced ensembles simulate warmer and drier conditions. This evaluation will be added in as supplementary material. This likely explains the differences in CDD distributions. The implication is that the absolute values coming from the models should be treated with caution, nevertheless we see a partially significant shift in CDD distributions of both ensembles. This demonstrates a qualitative (not quantitative) change towards longer dry periods. Lastly, CDD is measured in terms of weeks in the plots as it is more intuitive to interpret 8 weeks without precipitation greater that 1mm over a region than it is to read 56 days. In addition, it reduces the ink-to-data ratio in terms of visualization.

3) Is the main objective of this paper to present a new data set or to present mainly new results? Either way, both are not properly presented and need more details.

A: The aim of this paper is to present a new dataset. In addition, we provide 4 examples of how adaptation-relevant information can be derived from this dataset. We will make this clearer by improving the introduction and methodology sections. The latter will have a dedicated sub-section describing the HAPPI protocol with more details (see below).

4) Missing details/explanations: Some points (especially in the Methods section) need a better/more detailed explanation to be understandable. Your descriptions are too short and raise more questions than they answer.

A: Overall, the section on Methods will be substantially improved, as discussed above. We address below the point by point comments made on missing details and explanations also for the other sections.

4a) L58/59: Why did you use 20 years for the pre-industrial period, but only 10 years for all other simulations? From a climatological perspective, 10 years are rather short to enable a climatological view.

A: The pre-industrial period is only used as a baseline to define 0°C global mean warming. The definition is coming from the HAPPI protocol that every group doing global simulations followed. We are aware that there are several slightly different definitions of this particular period. The IPCC Special Report on 1.5°C warming lists several of these definitions. We will add a new sub-section on HAPPI to the manuscript (see below) to make these definitions more clear. Also the ten-year period is coming from the HAPPI protocol and is discussed in Mitchell et al. (2017). We will add the motivation for the ten-year period in the HAPPI sub-section.

4b) L58: Which period did you use for the future climate simulations? There are a few references given, but it should be specifically described which methods have been used here and how they are applied.

A: The method will be explained in more detail in the sub-section about HAPPI (we included a rewritten section in the response to referee#1).

4c) Did you compare the simulation for a current decade to observations and/or reanalyses to check how realistic they are? This would be very important.

A: We agree that comparing simulation results of climate models to observations is always very important to gain trust in models. The model version on this domain has been extensively evaluated using Re-analysis as boundary conditions in many other studies, and the relevant papers are already cited. In this paper, we want to demonstrate the benefits of using a large ensemble when looking at small changes in terms of global mean temperature - therefore an analysis against observations or reanalysis is less important. Especially since we only consider projected changes. We did comparisons in terms of quick views though, and concluded that the results were of comparable quality as, e.g., historical simulations from CORDEX with CMIP5 boundary conditions. However, as this point has also been raised by other referees, we will include a quick analysis on the general performance as supplement material.

4d) L59-61: How are the warming levels (1.5 and 2.0) calculated? Did you use RCP2.6 for 1.5 warming and RCP4.5 for 2.0 C warming?

A: The warming level has been calculated from a CMIP5 ensemble mean global mean temperature response. In case of the 1.5°C period RCP2.6 was used. The 2.0°C period is calculated using a weighted mean between RCP2.6 and RCP4.5. A more extensive explanation is given in the new subsection on the HAPPI experiment (please refer to response to referee#1).

4e) L73: The regional ensemble consists of 125 members, correct?

A: We have 125 members (100 from ECHAM6 and 25 from NorESM) for each of the three periods (current, 1.5° and 2.0° periods).

4f) L122-124: How exactly did you define CDD? Why did you calculate the CDD for the PRUDENCE regions and not on grid-point basis as the other indices? Is CDD the maximum number per year or over the 10-year period?

A: As we explain in the paper, the threshold is less than 1 mm per day (lines 123-124). We calculated the maximum duration for the entire 10 years of each ensemble member. The CDD analysis is computed for the Prudence regions and not 'per-grid-box' as for the other indices used in this study because applications drought indices are relevant over larger areas, whereas in the cases of the other indices considering high temperatures and heavy precipitation, analysis of individual grid-boxes are more relevant as they have more local applications. We will add this to the text.

L122 will be rephrased: "Lastly, the Consecutive Dry Days (CDD), defined as the maximum number of consecutive days with a daily precipitation amount of less than 1mm over a region (Karl et al., 1999; Peterson et al., 2001) is calculated for the entire 10 year period of each ensemble member. The CDD is calculated for each of PRUDENCE regions (Christensen et al., 2007), illustrated in Figure 1, because drought indicators are relevant over large areas.

4g) L130-145: Why did you choose different statistical methods to investigate the individual climate indices?

A: It is a common procedure to employ different statistical methods for different climate indices depending on the physical variables from which it computed from, for example, temperature or precipitation. This is because some statistical tests assume a given underlying distribution shape, for example temperature generally follows a normal distribution, whereas precipitation does not. Thus, different statistical tests ought to be employed for climate indices derived from temperature versus precipitation. In addition, the use and application of the different indices also warrant different approaches and methods.

4h) L130-145: The application of the significance measures is unclear, please rewrite. There are two methods used for two different parameters (ATG28 and RX5day). There seems to be some confusion on what is used for CDD (L143 says CDD similar to ATG28, but the method for CDD seems to be similar to RX5day, namely Mann-Whitney). No information is given for RI50yr. Anyway, the paper provides only information on the significant changes for CDD, but not for the other parameters. This should be remedied.

A: We will rewrite section 2.3 and dedicate one sub-section to each indicator to better distinguish the used methods between them.

4i) Why do you think that all differences between the two ensembles are due to the different ensemble sizes (e.g. L172-175)? They could also result from the driving GCMs. It might be useful to include the results for the smaller sub-sample of ECHAM6 that you mentioned in the text.

A: As we stated already in the paper, analysed this with a random sub-set of 25 ECHAM6 members and found similar, noisier patterns like in the NorESM ensemble. This supports our conclusion regarding ensemble size. However, to provide this actual information to the readers, we will add the ECHAM6 sub-set plots as supplement, in the revised paper.

4j) Table 1: Partly, the smaller ensemble (NorESM) generates more significant results. How does this relate to the hypothesis that a larger ensemble size is beneficial?

A: We disagree with the reviewer, as it cannot be concluded on the basis of the CDD in Table 1 only that the NorESM smaller ensemble provides more robust results. This could be by chance, as the NorESM model could lead to relatively dry projections, for instance, leading to changes that are more significant. We have looked at four different indices (precipitation and temperature), and across the board the larger ensemble has less noise, regardless of the change, at the same level of warming.

4k) Compare your results to previous studies (see interactive comment by Laura Suarez-Gutierrez for more details).

A: We have provided responses to the comments from Laura Suarez-Gutierrez. We will include comparisons to other studies, but many of Laura Suarez-Gutierrez suggestions are beyond the scope of our paper. Please see our responses to her comments.

4l) The authors should not oversell their results (or should argue more convincingly). E.g. Impact of the ATG28 increase (L241-246): What does such a change really mean w.r.t health issues? I assume that the number of days above the threshold is already high around the Mediterranean? Does a change of O(10days) drastically change the base level and/or the potential health impacts in this region? Furthermore, is a resolution of 50km sufficient to derive change estimates for local adaptation measures? The authors do not provide enough convincing evidence for the benefits of their method. Nevertheless, I suppose that there are sufficient arguments for it, but this should be better phrased.

A: The chosen threshold is relevant in particular for sudden cardiac death exposure. The number of days in the baseline is of similar order. Please bare in mind that this indicator is not dependant on temperature alone, but that humidity also plays an important role. We agree that 50km might not be sufficient to inform local adaptation measures, especially if one thinks about cities. But in this regard our results would serve as a lower bound of expected changes, because we have no urban heat island effects in our model. We will rework the discussion section to make it more convincing.

SPECIFIC COMMENTS

Data set, dataset or data-set?

A: We will use the term dataset and change the manuscript accordingly.

NOResm or NorESM?

A: We will use the term NorESM and change the manuscript accordingly.

L17: “measures at a” – Delete “a”.

A: We will correct as recommended.

L20-22: This sentence is incomprehensible, especially the first part.

A: We will rephrase that sentence as follows:

“Identifying regional climate change impacts for different global mean temperature targets is increasingly relevant to both the private and public sectors. In the private sector, investors demand financial disclosure associated with climate change risks and opportunities (Goldstein, et al., 2018). In the public sector, policy makers rely on climate information build on internationally agreed limits to develop national climate action policies.”

L25: “current generation global climate simulations” –> “current generation of global climate simulations”

A: We will change as recommended.

L45-47: Please rephrase sentence.

A: We will merge the sentence with the following sentence:

“Here, we develop two regional climate datasets of 25 and 100 members to create a large ensemble of RCM simulation which are particularly suitable to study extremes. Earlier studies such as Leduc et al. (2019) have successfully demonstrated the usefulness of such an approach.”

L55-56: You cannot downscale an RCM using GCM simulations.

A: We will rephrase that sentence:

“To create a data set for regional climate impact studies for Europe under 1.5°C and 2.0°C global warming the regional climate model REMO has been used to dynamically downscale two GCM ensembles following the HAPPI experiment protocol by Mitchell et al. (2017).”

L59: What does “CMIP5 mean SST anomaly” actually mean? All CMIP5 models and members, or just some, only ECHAM6/NorESM? Is there a reference?

A: This point will be addressed in a dedicated section explaining the HAPPI protocol in more detail (see answer to referee#1). In the HAPPI protocol all CMIP5 models are included in the averaged SST for the current and the projected periods of 1.5°C and 2.0°C respectively. The SST anomalies of the 1.5°C projected period are calculated by subtracting the averaged current SST from the averaged SSTs of the 1.5°C projection. The SST anomalies are then added to the observed SSTs of 2006-2015. The 2.0°C SST anomaly is computed in a similar manner.

L66/67: “from the core domain defined by CORDEX the entire domain has 121x129 grid boxes” – Please reword.

A: We will rephrase:

“The European CORDEX domain for REMO covers 121x129 grid boxes. To exclude the sponge zone, where the REMO simulations are relaxed towards the GCM solutions, a core domain of 106x103 grid boxes, following the CORDEX definition, is used for the analyses.”

L94: “recommend” –> “recommended”

A: We will changed as recommended.

L97-100: You defined an abbreviation for each climate index. Use them more consequently throughout the text.

A: We will make changes made as recommended.

L99: “precipitation intensity at the 50-yr return period” – Do you mean precipitation intensity with a 50-yr return period? Please use a consistent explanation throughout the text.

A: Yes, we mean precipitation intensity that occurs every 50 years. We will provide an explanation, and use the definition more consistently throughout the text.

L111: What do you mean with “annual sum maximum”?

A: This is a spelling mistake. It should be “annual sum”. We will revise:

“The index for the annual maximum of the five-day precipitation sum (RX5day) is used to characterise heavy precipitation events, which can be relevant for flood generation in river basins.

L121: “between 100 and 100 years” – Please correct.

A: This will be corrected to ‘10 and 100 years’.

L130: What do you mean with “historical”? Pre-industrial or current? Same in L141, L205, L215.

A: This is an inconsistency. It should be “current” everywhere. The text will be changed accordingly.

L132: “Only areas with more than 20 non-zero data points” – Isn’t the calculation done at every single grid point?

A: Yes, the calculation is done on every grid point. The data points refer to number of exceedance in the current period. The formulation will be changed:

“Only grid boxes with more than 20 exceedances over threshold in the current period were included in the analysis of ATG28 in order to allow for confidence interval calculations for the shown percentiles using order statistics.”

L152: “in the median around the Mediterranean” – Please reword.

A: This will be rephrased as follows:

“Around the Mediterranean the increase in ATG28 during the 1.5° C period is mostly moderate with up to 9 days in the median whereas changes in the 2.0°C period are reaching 18 days and more.”

L154: Is a spatial resolution of 0.44 high enough to resolve complex topography? What about EUR-11?

A: We agree that EUR-11 would be better to resolve complex topography, however, with the current generation of HPC computers, such a large ensemble of RCM simulations would not have been possible to conduct on much higher resolution than 0.44°.

L176: “interior of the simulation domain” – Please reword.

A: We will rephrase:

“Apart from artificial effects due to the boundary conditions, the strongest signal within the core domain appears over the Baltic Sea, with an increase of up to 15% in RX5day under a 2.0°C increase in GMT.”

L190-191: Please reword sentence.

A: We will rephrase as follows:

“The relative change (in percent) in RI50yr across Europe are presented in Figure 5.”

L204: You never defined/explained p-values.

A: Please see our answer to L204 below

L204: “Both the distributions 1.5C and 2.0C” – Please reword.

A: We will rephrase the entire section to read as follows:

“In this section, the changes in the Consecutive Dry Days (CDD) distributions for the 1.5°C and 2.0°C periods compared to the current period are presented. To distinguish whether these changes in the distributions are statically significant we employ the Mann-Whitney U-Test. Where the resulting p-values of the test are greater or equal to a significance level, alpha, of 0.05 or smaller, the null hypothesis is rejected indicating the distributions differ. The p-values for each of the PRUDENCE regions (Christensen et al., 2007), shown in Fig. 1, are presented in Table 1.”

L215: “shift towards longer periods of dry days” – This depends on your definition of CDD. If CDD is the maximum number of consecutive dry days, your results only show that the longest dry period is getting longer in a warmer climate. That does not mean that all dry periods will be longer.

A: Our formulation of CDD is the maximum consecutive days in 10 years. In this analysis, we are determining whether the shift in CDD distributions are significant. In Figure 6, the percentage of the entire CDD distribution of a given duration is presented.

L215 reformulation:

”Over this region, one can see an increase in duration of the longest dry period and that they occur more often (Figure 6).”

L216: “the Alps and Eastern European region” –> “the Alps and Eastern Europe”

A: We will rephrase accordingly.

L218: “suffer from more frequent and longer drought periods” – Again, this depends on your definition of CDD. If you used the common one (CDD being the length of the longest dry period), you cannot say anything about the frequency of dry periods.

A: In Figure 6, we have plotted the distribution of CDD and are thereby studying the frequency of dry periods. Over region 2 one can see that the length of the longest dry period increases as the reviewer states and one can also see that the number of longer CDD increases. The authors argue ‘more frequent and longer drought periods’ in this case is correct.

L222: “adaption” should be “adaptation” Figure 6: Use ECHAM6 instead of ECHAM.

A: We will change accordingly.

L235: “10 x 100 years” –> “100 x 10 years”

A: We will hange accordingly.

L247: “RX5day” –> “(RX5day)”

A: We will change the sentence to:

“The RX5day shows a general increase over Europe which is more pronounced under higher global mean warming.”

L251: “such precipitation extremes” –> “such as precipitation extremes”

A: “as” will be included.

L263: “historically similar” – Strange wording.

A: See our answer to L263-264, below.

L263-264: “pre-industrial period” – I thought you were calculating differences between future and current climate and not between future and pre-industrial?

A: There was a mix-up in the sentence. It should always be current vs. future. The text will be changed accordingly, as:

“The changes to CDD distributions show that Spain will experience significantly more drought conditions in the future compared to the current period, even at a 1.5°C increase in GMT. For Italy, drought conditions associated with the 1.5°C simulations show non-significant changes, yet those associated with the 2.0°C simulations are significantly different to the current period, thus showing possible consequences of exceeding the 1.5°C GMT target of the Paris agreement.”

Answers to Referee#3

General response:

The authors would like to thank the referee for the time and effort put into this review. These comments have been useful in improving our manuscript. We have carefully read the comments and provide a detailed response to each comment.

This paper presents results from a regional dynamical downscaling of two GCM ensemble from the HAPPI project and compares the model output to 4 extreme event indices.

The large number of ensemble members and important can potentially make a good contribution in how to analyse and estimate the difference between relatively close climate change "targets".

One major discrepancy however is that there are no evaluation of the quality of results with respect to observations or reanalysis. If the model results deviates to much from the "real world" one can not trust the conclusions given for the warm periods. The discussion does not have to be very advanced but I think you should include a point on this.

Answer: We did a quick analysis on this and found that the results are in good agreement with other downscaling activities using REMO for Europe (e.g., CMIP5 downscaling of GCMs). As suggested to referee#2, we will add a short analysis as supplement.

In general both the description of methods and results are not always very clearly defined. Sometimes it is just a matter of language. The sentences are sometimes so long that the reader loose the thread before getting to the end, so please try to be more focused.

A: We re-formulated many sentences that caused confusion in response to Reviewer#1 and #2.

A problem is that it is sometimes unclear whether the assumptions used are your / RCM limitations or inherited from the Happi protocol. Event hough I know the Happi protocol it was sometimes hard to separate. I see that many of the questions from other reviewers on critical assumptions e.g. how does pre-industrial come into play, when is the future time slice .. is often related to the protocol. I think the authors should include a summary of this information, in particular as you rightly point out that the Happi protocol is quite different from the traditional CMIP settings.

A: We came to the same conclusion. Therefore, we separated the description of the HAPPI protocol from the REMO model set-up and extended it (see answer to Reviewer#1).

The defintions/ presentations of the indices and how they are used can be improved (Section 2.2 and 2.3) It is particularly hard to follow the setup of significance tests. "RX5 days are computed ...similar to ATG28, however …

A: We will adjust the text and add sub-sections for each indicator.

CDD is similar to ATG28 but with the method used for RX5 but now however? There are no significance test for RI50yr ? There does not have to be one but I would like to know.

A: No, we did not use a significance test for RI50yr.

With only 4 indices I think you should avoid "similar" as much as possible and just describe the methods for each.

A: Separating the description of the indicators should make it more clear (see answer above).

The discussion on being able to reproduce the more noisy results of the smaller ensemble by picking a smaller number of the largest ensemble was interesting. I wonder if it is any way of presenting this in a cumulative manner, with respect to number of ensemble members. The change rate may also indicate whether even the largest ensemble is too small. I realize however that this is likely beyond the scope of the article.

A: We will include an analysis on the randomly picked 25 member ensemble from ECHAM6 as a supplement as suggested to Reviewer#2. We agree the further analysis on the change rate would be beyond the scope of this article.

To the point on data availability: Is it the data used specifically for this study, or is it general RCM output.

A: We are still struggling with finding a data centre that can host the data. Due to CMIP6 activities our dataset is of low priority. However, we managed to cmorize the most common variables and can provide them by request. This is already ongoing in the framework of other projects we are working on.

Some minor suggestions /corrections

Line 20: “differ” not needed

A: We will rephrase this sentence to (see also answer to Reviewer#1):

“Identifying regional climate change impacts for different global mean temperature targets is increasingly relevant to both the private and public sectors.”

Line 25: Include a line on scenario definition in CMIP6?

A: We will add CMIP6 to this sentence:

“Temperature targets, however, are not directly related to the Representative Concentration Pathways (RCP, Van Vuuren et al., 2011) used in the experimental design of CMIP5 (Taylor et al., 2012) or the Shared Socioeconomic Pathways (SSP, Meinshausen et al., 2019) used in CMIP6 (Eyring et al., 2016).”

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P., Reimann, S., Smith, S. J., van den Berg, M., Velders, G. J. M., Vollmer, M., and Wang, H. J.: The SSP greenhouse gas concentrations and their extensions to 2500, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-222, in review, 2019.

Line 33:skip the word “indeed”

A: We will change the sentence to:

“The high natural variability in models requires the creation of large ensemble datasets (Deser et al., 2013)”

line 70: Green-house-gas -→greenhouse gas

A: This will be changed accordingly.

line 81: Sea-Ice → sea-ice (usually not capital letters)

A: This will be changed accordingly.

line 121: Missing intervall ("100" years on both sides)

A: This is a typo and should read "between 10 and 100 years".

Figure 1. Is the outer map equal to the model domain?

A: Yes, the outer domain is equal to the model domain.

Line 135-136 The explanation is fine, but then you do not need to mask it out either. It would show up as insignificant?

A: We agree that masking "non-significance" and masking for other reasons is confusing. We will change that in an updated version of the Figure. We masked the SST before running the test. Hence, we do not know if it would show significance or not.

Line 141 NOResm → NorESM

A: Will be changed accordingly.

Figure 6. The difference between the two downscaling sets are quite large. Any comments.

A: This is related to the forcing model. On the one hand, the NorESM ensemble is wetter compared to the ECHAM6 ensemble. On the other hand, the ECHAM6 driven ensemble seems to be biased towards warm/dry conditions. This can be examined in more detail when a small evaluation is added to the supplement.

line 260 -265 use significance instead of unlike / similar

A: We will change the text to the following formulation (see also response to referee#2):

"The changes to CDD distributions show that Spain will experience significantly more drought conditions in the future compared to the current period, even at a 1.5°C increase in GMT. For Italy, drought conditions associated with the 1.5°C simulations show non-significant changes, yet those associated with the 2.0°C simulations are significantly different to the current period, thus showing possible consequences of exceeding the 1.5°C GMT target of the Paris agreement."

Weather extremes over Europe under 1.5°C and 2.0°C global warming from HAPPI regional climate ensemble simulations

Kevin Sieck[1], Christine Nam[1], Laurens M. Bouwer[1], Diana Rechid[1], Daniela Jacob[1]

[1]Climate Service Center Germany (GERICS), Helmholtz-Zentrum Geesthacht, Hamburg, 20095, Germany

Correspondence to: Kevin Sieck (kevin.sieck@hzg.de)

Abstract. This paper presents a novel dataset of regional climate model simulations over Europe that significantly improves our ability to detect changes in weather extremes under low and moderate levels of global warming. The dataset provides a unique and physically consistent dataset, as it is derived from a large ensemble of regional climate model simulations. These simulations were driven by two global climate models from the international HAPPI consortium. The set consists of 100 x 10-year simulations and 25 x 10-year simulations, respectively. These large ensembles allow for regional climate change and weather extremes to be investigated with an improved signal-to-noise ratio compared to previous climate simulations. The changes in four climate indices for periods with levels of 1.5°C and 2.0°C global warming are quantified: number of days per year with daily mean near-surface apparent temperature of >28°C (ATG28); the yearly maximum 5-day sum of precipitation (RX5day); the daily precipitation intensity of the 50-yr return period (RI50yr); and the annual Consecutive Dry Days (CDD). This work shows that even for a small signal in projected global mean temperature, changes of extreme temperature and precipitation indices can be robustly estimated. For temperature related indices changes in percentiles can also be estimated with high confidence. Such data can form the basis for tailor-made climate information that can aid adaptive measures at policy-relevant scales, indicating potential impacts at low levels of global warming at steps of 0.5°C.

1 Introduction

Identifying regional climate change impacts for different global mean temperature targets is increasingly relevant to both the private and public sectors. In the private sector, investors demand financial disclosure associated with climate change risks and opportunities (Goldstein, et al., 2018). In the public sector, policy makers rely on climate information build on internationally agreed limits to develop national climate action policies. This is especially true after the adoption of the Paris Agreement of the United Nations, which aims to keep global climate warming well below 2.0°C compared to pre-industrial times (UNFCCC, 2015). Temperature targets, however, are not directly related to the representative concentration pathways (Van Vuuren et al., 2011) used in the generation of global climate simulations (CMIP5, Taylor et al., 2012) and the Shared Socioeconomic Pathways (Meinshausen et al., 2019) used in CMIP6 (Eyring et al., 2016). Therefore, new techniques are being developed to extract information on the possible implications of further global warming. Recent studies using CMIP5 data have shown that climate change indices can be extracted for different warming levels, by identifying specific time periods when a certain global mean temperature (GMT) increase is reached in a general circulation model (GCM) (Schleussner et al., 2016; Vautard et al., 2014; Jacob et al., 2018). These studies typically used the 5 to 15 ensemble members which were available in CMIP5 at the time for their global and regional studies. Mitchell et al. (2016) argue however that a different experiment design is needed to better address the policy-relevant temperature targets with climate simulations, because the relatively small CMIP5 ensemble does not provide the necessary size to quantify changes in weather extremes at low levels of warming. The high natural variability in models requires the creation of large ensemble datasets (Deser et al., 2013). Following the recommendations of Mitchell et al. (2016), the HAPPI consortium ("Half a degree Additional warming, Prognosis and Projected Impacts") designed targeted experiments created for the purpose of extracting the required information on distinct warming levels using 10 state-of-the-art GCMs (Mitchell et al., 2017). The HAPPI experiments include a large number of ensemble members, typically 50 to 100 members per GCM, using AMIP-style integrations (Gates et al., 1992), which significantly improves the signal-to-noise ratios. A better signal-to-noise ratio is essential for differentiating between impacts from 1.5°C and 2.0°C global warming, especially for changes in weather extremes.

To bridge the gap between GCM model output and regional climate impact assessments, which require a much higher resolution than GCMs (Giorgi and Jones, 2009), the Regional Climate Model (RCM) REMO (Jacob et al., 2012) is used to dynamically downscale simulations from two GCMs from the HAPPI consortium. Dynamical downscaling with RCMs is one option to bridge the gap between GCM model output and regional climate impact assessments which provides physically consistent high-resolution climate information (Jacob et al., 2014; Giorgi and Gutowski, 2015; Gutowski et al., 2016). Here, we develop two regional climate datasets of 25 and 100 members to create a large ensemble of RCM simulation which are particularly suitable to study extremes. Earlier studies such as Leduc et al. (2019) have successfully demonstrated the usefulness of such an approach. To demonstrate the potential of this dataset for regional climate impact studies, under 1.5°C and 2.0°C global warming, changes in four climate indices for weather extremes are quantified.

In Section 2, we present the REMO regional climate model, experiment setup and simulations performed. In Section 3, four relevant climate indices for extreme weather which can be derived from the HAPPI dataset are presented, and lastly, conclusions are derived in Section 4.

2 Methods

To create a dataset for regional climate impact studies for Europe under 1.5°C and 2.0°C global warming, the regional climate model REMO has been used to dynamically downscale two GCM ensembles following the HAPPI experiment protocol by Mitchell et al. (2017). Several common climate indices are computed, to demonstrate the usefulness of the data.

2.1 Global HAPPI simulations

The HAPPI protocol by Mitchell et al. (2017) has been set up to inform the IPCC Special Report on 1.5°C Warming. Large ensembles (>50 members) of GCM simulations where created that allow extreme events to be studied, even for the small differential warming between a current decade (2006-2015) and two future decades under 1.5°C and 2.0°C global warming. The major aspects of the protocol for the HAPPI experiment are summarized in the following paragraphs, as there are important differences compared to the typical CMIP protocols.

All simulations were conducted in atmosphere-only mode in order to increase ensemble size and provide more accurate regional projections (Mitchell et al., 2017; He and Soden, 2016). The simulation period for all members is limited to 10 years, because during the current period from 2006-2015 sea-surface temperatures stayed approximately constant. This period forms the basis of the entire experiment and allows for a better estimate of, e.g., return-values from this period compared to periods with a strong warming trend. The experiment design for the current decade follows the DECK AMIP protocol using observed sea ice and SSTs. For the future periods, SSTs are calculated by taking the 2006-2015 observed conditions and adding a SST anomaly representing the future periods.

By chance, the multi-model averaged CMIP5 global mean temperature response between 2091-2100 compared to 1861-1880 under RCP2.6 is 1.55°C. Mitchell et al. (2017) considered this warming as sufficiently close to inform about impacts under 1.5°C and chose this period under RCP2.6 as basis for a 1.5°C warmer period. The SST anomalies for the 1.5°C period were computed using the modelled decade averaged difference between 2091-2100 from RCP2.6 and 2006-2015 from RCP8.5, because RCP8.5 averaged SSTs over this period are closest to observations. Forcing values for anthropogenic greenhouse gases, aerosols and land-use are taken from the year 2095 of RCP2.6 and kept constant during the simulation. Because of the poor representation of sea ice in the CMIP5 models, Mitchell et al. (2017) used a different approach to construct sea ice concentrations for the 1.5° period. A detailed description would be beyond the scope of this paper and can be found in the cited reference.

The SST anomalies for the 2.0°C period cannot be calculated following a similar approach as for the 1.5°C period, because none of the RCPs show a global mean temperature response close to 2.0°C at the end of the century. Therefore, a weighted sum of RCP2.6 and RCP4.5 is calculated with a global mean temperature response of 2.05°C, which is exactly 0.5°C more compared to the chosen 1.5°C period. The calculation of SST anomalies and sea ice extend follows the same methodology as for the 1.5°C period. Mitchell et al. (2017) decided to apply their weighting method only to the well-mixed greenhouse gases, because the land-use changes and aerosols show very different spatial patterns and are therefore kept at the 1.5°C period values.

2.2 Regional HAPPI simulations

In order to create high-resolution climate data for Europe from HAPPI, the RCM REMO has been used for downscaling. REMO is a hydrostatic limited-area model of the atmosphere that has been extensively used and tested in climate change studies over Europe (Jacob et al., 2012; Teichmann et al., 2013; Kotlarski et al., 2014). The simulation domain follows the CORDEX specification for the standard European domain with 0.44° horizontal resolution. The European CORDEX domain for REMO covers 121x129 grid boxes. To exclude the sponge zone, where the REMO simulations are relaxed towards the GCM solutions, a core domain of 106x103 grid boxes, following the CORDEX definition, is used for the analyses. In the vertical, 27 levels are used without nudging except for the boundaries. Boundary conditions are taken from the HAPPI Tier1 experiments (Mitchell et al., 2017), which are carried out with ECHAM6 (Stevens et al., 2013; Lierhammer et al., 2017) (100 members per period) and NorESM (Bentsen et al., 2013) (25 members per period) that provide 6-hourly 3-dimensional data for downscaling. In REMO the same greenhouse gas forcings as for the GCMs were used.

SST and sea ice concentrations were taken directly from the GCM output matching the GCM land-sea mask for NorESM. From ECHAM6 only the sea ice concentrations were taken. Due to the interpolation procedure for the sea ice extent, it could happen that sea ice was artificially created where no ice conditions were present in the original dataset, e.g., during summer in the Baltic Sea. ECHAM6 has a mechanism that as soon as there is a fraction of sea ice greater than zero, the SST is limited to a maximum of 272.5K. This leads to artificial temperature jumps in the SST between adjacent grid boxes as soon as erroneous sea ice appeared in one of the grid boxes. In order to avoid inheriting this issue, the originally provided SST fields from the HAPPI project were used for the REMO simulations, using ECHAM6 as forcing GCM. After testing different temperature and/or sea ice fraction thresholds, the authors decided to keep the original sea ice maps, because in the cases where artificial sea ice was created the fraction was typically well below 1%, and only in rare cases reaches up to 4% (not shown). All other procedures would have removed too much sea ice in other seasons or led to unrealistic gradients of sea ice fractions. With the tile approach of REMO, the effect of the artificial sea ice on the averaged near-surface variables is hardly detectable.

For each GCM member only one REMO simulation was carried out, as inter-member variability of an RCM ensemble over Europe on a time scale of 10 years is small compared to the internal variability of a GCM (Sieck et al., 2016). Each simulation covers a period of ten years, and as such, initial conditions for the lower boundary need to be in balance with the RCM's internal climate in order to avoid artificial drifts in the modelled results. To achieve this, for each driving GCM, the first year of a random GCM member was simulated five times with REMO using initial conditions from the end of the previous run, creating one initial soil temperature state for every ensemble member in one period. This was performed for each of the three periods. Tests showed that this minimizes drifts in the deep soil climatology compared to initial conditions taken directly from the GCM (not shown).

We performed a qualitative analysis of the results compared to observations. In general, the performance of the HAPPI-ensemble is in-line with typical results from the CMIP-type downscaling activities performed in the CORDEX framework with REMO (see the supplement for figures and discussion).

2.3 Climate indices

To demonstrate how adaptation-relevant information can be derived from the HAPPI dataset for two different average global temperature targets, four climate indices used in climate impact studies are presented. The extremes are selected based on recommended indices developed by the joint CCI/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) (Karl et al., 1999; Frich et al., 2002) and other indicators. The selected climate indices are: number of days per year with a daily mean near-surface apparent temperature of more than 28°C (ATG28); annual maximum 5-day sum of precipitation (RX5day); change in daily precipitation intensity at the 50-yr return period (RI50yr); consecutive dry days (CDD) as a measure of meteorological drought.

All four climate indices are calculated from the daily mean precipitation, temperature, and/or dew-point temperature output of the model; for each year and ensemble member. The ECHAM6 driven ensembles yield 1000 data points for each grid box and simulation period, and NorESM driven ensembles have 250 data points, respectively.

Apparent temperature

The ATG28 index is used as an indicator for heat stress, which is relevant for impacts on human health (Davis et al., 2016). The apparent temperature is computed using the same formulation as in Davis et al. (2016):

$$AT = -2.653 + 0.994T + 0.0153T_d^2, \quad (1)$$

with AT being the apparent temperature, T the daily mean near-surface temperature, and T_d the daily mean near-surface dewpoint temperature. Similar formulations exist in the literature showing very similar results (see Anderson et al., 2013 for a review). The threshold of 28°C is based on the definition of Zhao et al. (2015) who set this limit as the lower boundary for human heat stress.

Five-day precipitation sum

The index for the annual maximum of the five-day precipitation sum (RX5day) is used to characterise heavy precipitation events, which can be relevant for flood generation in river basins. The RX5day represents a noisy, i.e., highly spatially and temporally variable parameter. A large ensemble would allow for a better assessment of the signal-to-noise in extreme precipitation.

Daily precipitation intensity

A change in extreme precipitation directly influences local communities. Such communities have applied design standards for structures to withstand floods with a specified return period. These return standards will no longer be applicable when the extreme value distribution shifts with global warming. A Gumbel Type I extreme value distribution is fitted to the annual maxima of daily rainfall amounts. Using this distribution, an estimate is made of the intensity of rainfall events associated with a given exceedance probability, in line with engineering practices. For each ensemble, the daily rainfall intensity with a 50-year return is computed, hereafter called RI50yr. Information on changes in the rainfall intensity with a 50-year return interval is useful for infrastructure design and maintenance. For example, road authorities in Europe typically use between 1 and 10-year return periods for assessing effects of rainwater falling on major roads (highways), and between 10 and 100 years for rainfall beside the road and waters crossing the road (Bless et al., 2018).

Consecutive dry days

Lastly, the Consecutive Dry Days (CDD), defined as the maximum number of consecutive days with a daily precipitation amount of less than 1mm over a region (Karl et al., 1999; Peterson et al., 2001) is calculated for the entire 10-year period of each ensemble member. The CDD is calculated for each of PRUDENCE regions (Christensen and Christensen, 2007), illustrated in Figure 1, because drought indicators are relevant over large areas, as impacts on water resources occur at these scales.

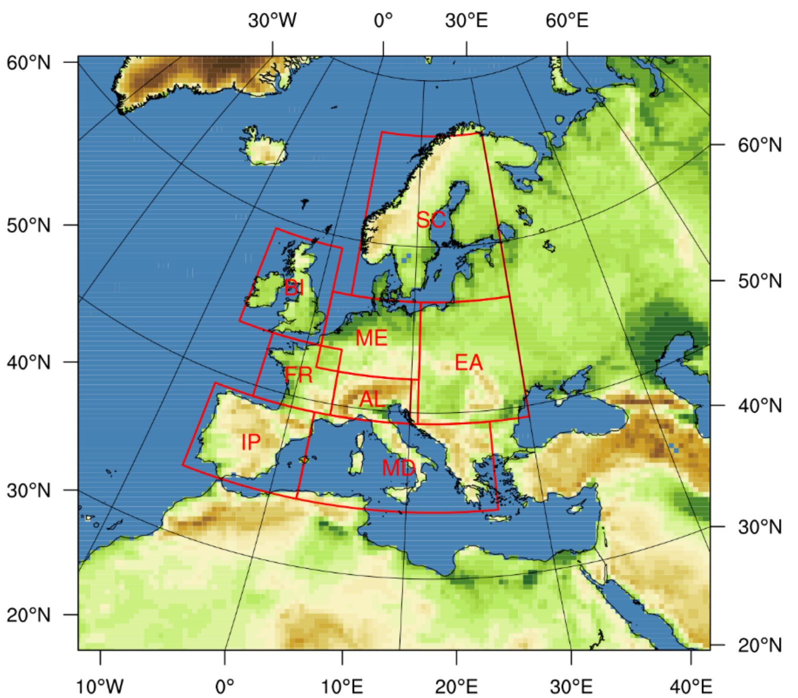


Figure 1. The PRUDENCE regions.

2.4 Changes in climate indices

For each of the climate indices, the differences between simulations of current and future temperature levels of 1.5°C and 2.0°C are computed as follows: in case of ATG28, differences of the 5th, 50th, and 95th percentiles were computed by subtracting the respective percentiles of the current decade from the projected periods. Only grid boxes with more than 20 exceedances over threshold in the current period were included in the analysis of ATG28 in order to allow for confidence interval calculations for the shown percentiles using order statistics. Statistical significance is determined when the calculated percentile in the 1.5°C or 2.0°C period is outside the percentile confidence range of the current period. As ATG28 is temperature-based, changes over the ocean surfaces are masked out, because they are to a large extent determined by the prescribed SST changes.

Differences for RX5day are computed by subtracting the ensemble mean of the current decade from the 1.5°C and 2.0°C periods, similar to the ATG28, however the statistical significance for RX5day was calculated using a Mann-Whitney-U-test and only results are shown with a significance at the 95% level.

The differences in RI50yr are computed as the relative change in daily precipitation intensity of the 50-yr return period between the 1.5°C and 2.0°C simulations compared to the current period simulations.

Similar to the ATG28 analysis, the differences in the distribution of CDD are calculated with a Mann-Whitney U-Test with a significance at the 95% level, determining whether samples from the two periods are drawn from a population with the same distribution.

3. Results

3.1 Apparent temperature

Figures 2 and 3 show the changes in ATG28 for the NorESM and ECHAM6 driven ensembles, respectively. The grey boxes are masked out areas. On land they refer to grid boxes that do not match our criteria of 20 or more occurrences of ATG28 during the current period. In general, the changes are strongest close to warm ocean areas, especially around the Mediterranean. But also the central and eastern parts of Europe show increases in ATG28, consistent with the increase in mean temperature (not shown). The distinct difference between the two warming levels should be noticed. Around the Mediterranean the increase in ATG28 during the 1.5° C period is mostly moderate with up to 9 days in the median whereas changes in the 2.0°C period are reaching 18 days and more. This result is consistent between the ECHAM6 and NorESM driven ensembles. Also, the changes across the percentiles are consistent, i.e., there is no change in the shape of the distribution of ATG28. It should be noted that the spatial resolution of the simulations allows to show the lower level of warming in mountainous areas compared to coastal areas, e.g., over Italy. This is especially important in areas with complex topography such as the Mediterranean region, which is usually only poorly resolved in GCM simulations.

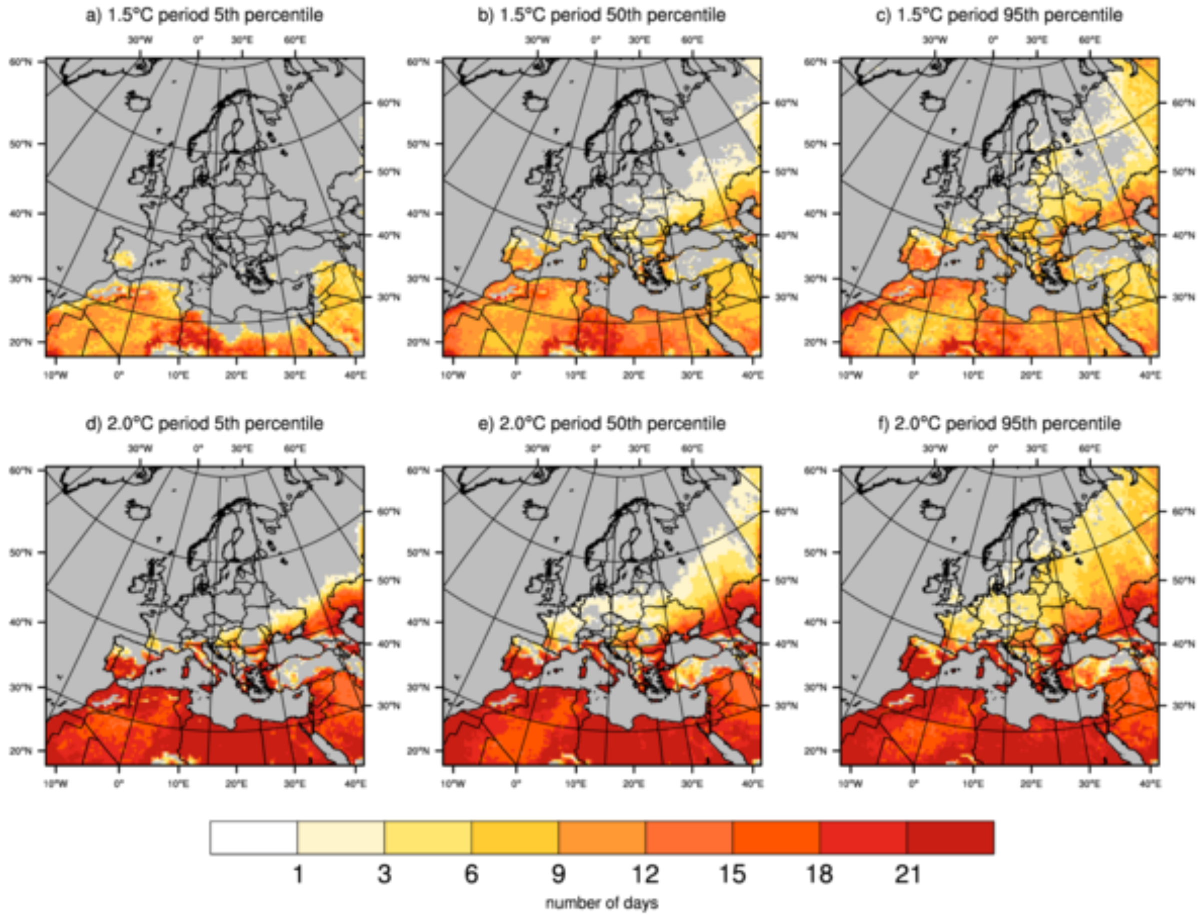


Figure 2. Differences in ATG28 between the current and the 1.5°C period (top row) respectively the 2.0°C period (bottom row) for the NorESM driven REMO simulations in number-of-days. Shown are the Differences in the 5th percentile (left column), median (middle column) and 95th percentile (right column). Differences over Ocean areas are masked out in grey, because they are closely related to the prescribed SST changes.

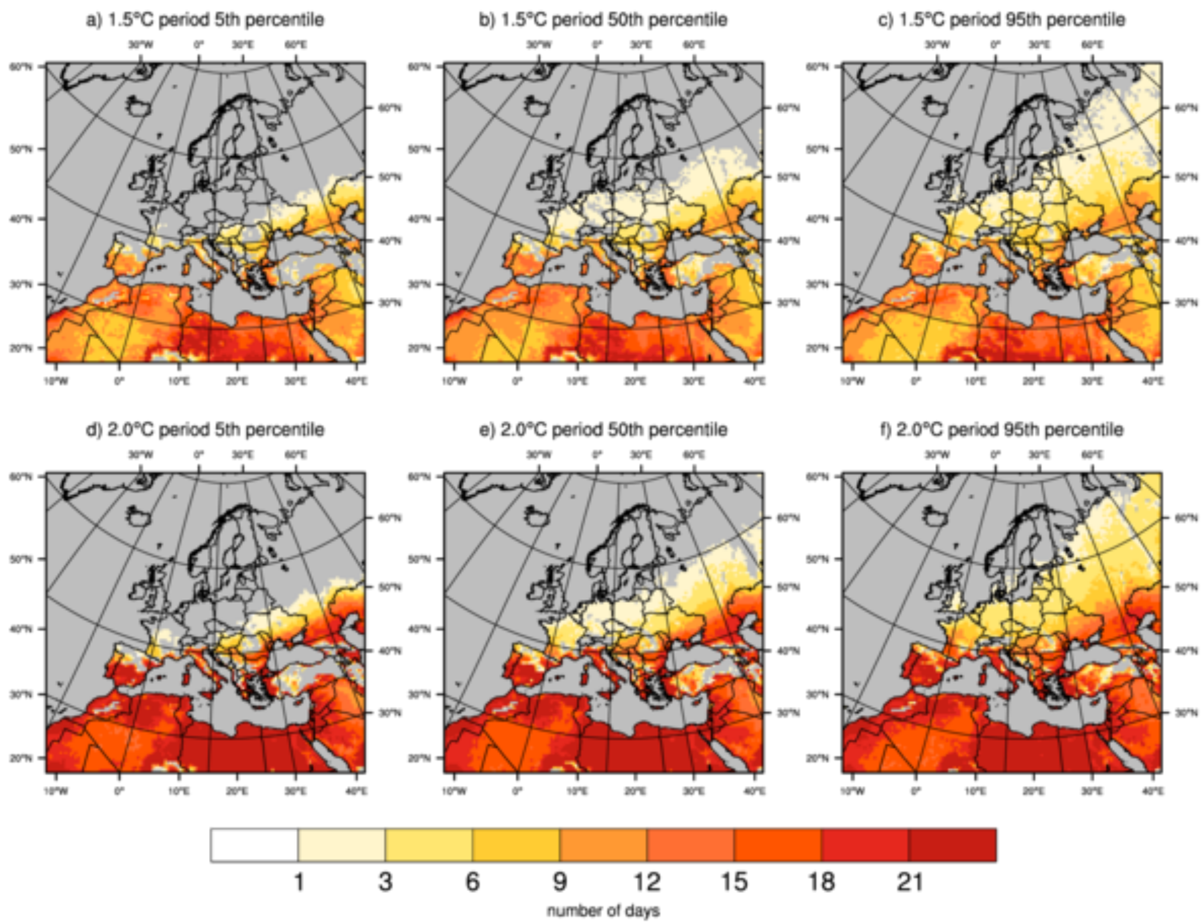
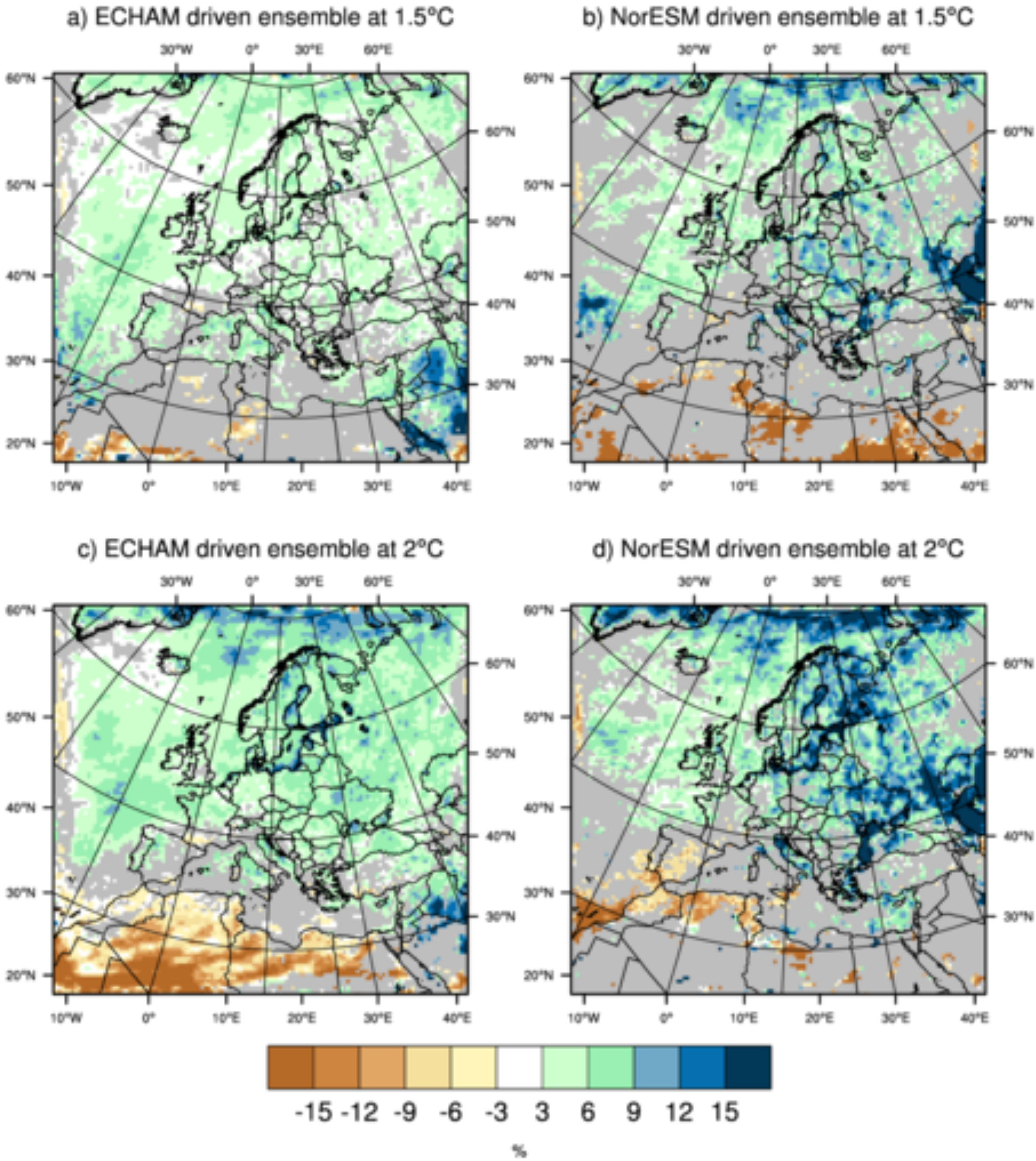


Figure 3. Same as Figure 2, but for the ECHAM6 driven REMO simulations.

3.2 Five-day precipitation sum

Figure 4 shows the relative differences of RX5day for the REMO ensemble experiments. In general, there is an increase in RX5day over the European part of the domain with stronger signals in the 2.0°C period compared to the 1.5°C period. It can also be seen that the patterns in the ECHAM6 driven ensemble is more coherent than the NorESM ensemble with larger areas showing a significant change. This is related to the difference in ensemble size between ECHAM6 and NorESM driven simulations and underlines the necessity for a large ensemble to achieve proper signal-to-noise ratios when looking at the difference in regional changes under small GMT increases in highly variable quantities such as precipitation extremes. Tests with a randomly picked 25-member ensemble from the ECHAM6 driven simulations showed a similar noisy pattern as the NorESM driven runs (see Supplement).

Apart from artificial effects due to the boundary conditions, the strongest signal within the core domain appears over the Baltic Sea, with an increase of up to 15% in RX5day under a 2.0°C increase in GMT. This result is consistent between both ensembles. A similar increase can be seen over the Adriatic Sea, but is not so pronounced in the ECHAM6 driven ensemble. This might be related to feedbacks from the unresolved SSTs, because the GCMs usually do not resolve these small basins. In these locations the SST is interpolated from the nearest SST value of the GCM, which might not be adequate for the region.



220 **Figure 4. Relative difference of RX5day (in percent) between current and the 1.5°C period (left column) respectively the 2.0°C period (right column) for the NorESM with 25 members (top row) and ECHAM6 with 100 members (bottom row) driven REMO simulations.** **Grey shading show areas with non-significant changes on the 95% significance level.**

3.3 Daily rainfall intensity, 50-year return period

To account for the spatial differences in 50-year return period across Europe, the changes (in percent) in RI50yr are presented in Figure 5. In both the NorESM and ECHAM6 driven ensembles, a greater increase in the rainfall intensity is found in the 2.0°C simulations compared to 1.5°C. ECHAM6 driven simulations clearly show increases in RI50yr of up to 20% over continental Europe. The estimated changes in rainfall intensity in the NorESM driven simulations appear to be more extreme but these simulations are also more noisy as they are based on fewer ensemble members.

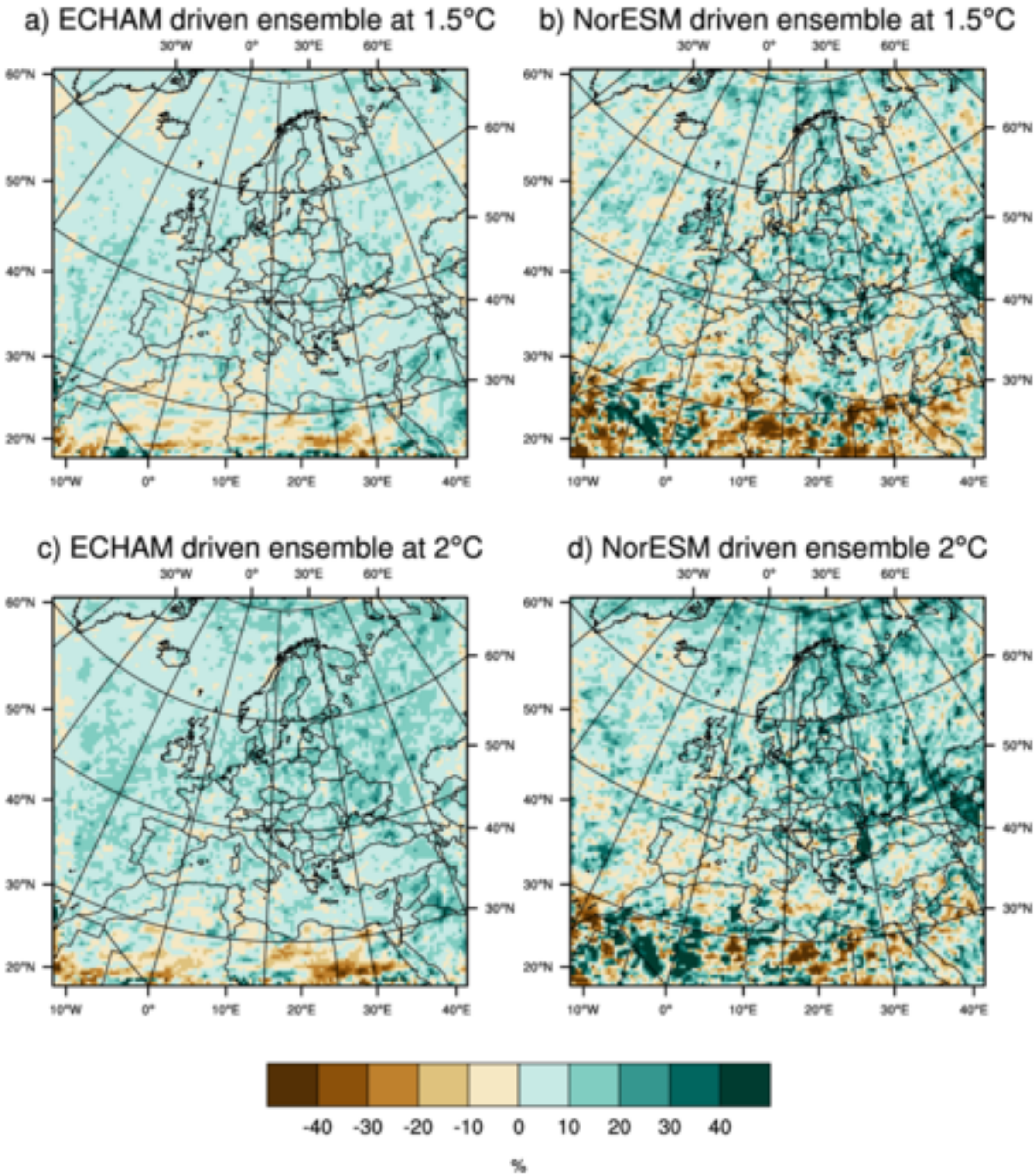


Figure 5. Relative difference of RI50yr (in percent) between current and the 1.5°C period (left column) respectively the 2.0°C period (right column) for the NorESM with 25 members (top row) and ECHAM6 with 100 members (bottom row) driven REMO simulations in %.

3.4 Consecutive dry days

In this section, the changes in the Consecutive Dry Days (CDD) distributions for the 1.5°C and 2.0°C periods compared to the current period are presented. To distinguish whether these changes in the distributions are statically significant we employ the Mann-Whitney U-Test. Where the resulting p-values of the test are greater or equal to a significance level, alpha, of 0.05 or smaller, the null hypothesis is rejected indicating the distributions differ. The p-values for each of the PRUDENCE regions (Christensen and Christensen, 2007), shown in Fig. 1, are presented in Table 1. Both the 1.5°C and 2.0°C distributions are compared to the current CDD distribution. Where the p-value is greater or equal to a significance level, alpha, of 0.05 or smaller, the null hypothesis is rejected indicating the distributions differ. Bold numbers in Table 1 indicate that the distributions differ according to the test.

Table 1. Mann-Whitney U-Test p-values for distributions of Consecutive Dry Days (CDD) for the ECHAM6 and NorESM driven ensembles for different PRUDENCE regions (for locations see Figure 1).

PRUDENCE Region:	**ECHAM6**		**NorESM**	
	Hist vs 1.5°C	Hist vs 2.0°C	Hist vs 1.5°C	Hist vs 2.0°C
1. (British Isles)	0.270	**0.035**	**0.024**	0.053
2. (Iberian Peninsula)	**0.036**	**0.000**	**0.000**	**0.000**
3. (France)	0.391	0.230	**0.015**	**0.002**
4. (Middle Europe)	0.077	**0.015**	0.363	0.212
5. (Scandinavia)	0.238	0.356	**0.046**	0.081
6. (Alps)	0.333	0.105	0.378	0.355
7. (Mediterranean)	0.069	**0.036**	0.348	**0.021**
8. (Eastern Europe)	0.325	0.465	0.419	0.385

We begin by looking at three regions where the Mann-Whitney U-Test provided consistent results across the ensembles. In region 2, the Iberian Peninsula, the CDD distributions in both the 1.5°C vs. 2.0°C simulations differ statistically compared to the simulations for the current period. Over this region, one can see an increase in duration of the longest dry period and that they occur more often (Figure 6). In contrast, regions 6 and 8, the Alps and Eastern Europe, have changes in CDD distributions that are statistically indistinguishable under 1.5 C and 2.0C compared to the current period (Table 1). One can deduce that region 2, will suffer from more frequent and longer drought periods than experienced before compared to regions 6 and 8. Interestingly, for region 7, the Mediterranean, the CDD distributions of the ECHAM6 and NorESM ensembles of 1.5°C do not differ statistically from the current period, yet both ensembles show a statistically different distribution at 2.0°C. Thus, one can conclude for region 7, according to these simulations, a lower target of 1.5°C increase in GMT could reduce the length of the maximum number of dry days in this region, compared to a 2.0°C target. In region 3, France, the results of the two ensembles differ. The ECHAM6 simulations suggest there is no difference in CDD distributions between the warmer climate compared to the current period; whereas the NorESM simulations find the warmer climate has a distinctly different CDD distribution compared to the current period.

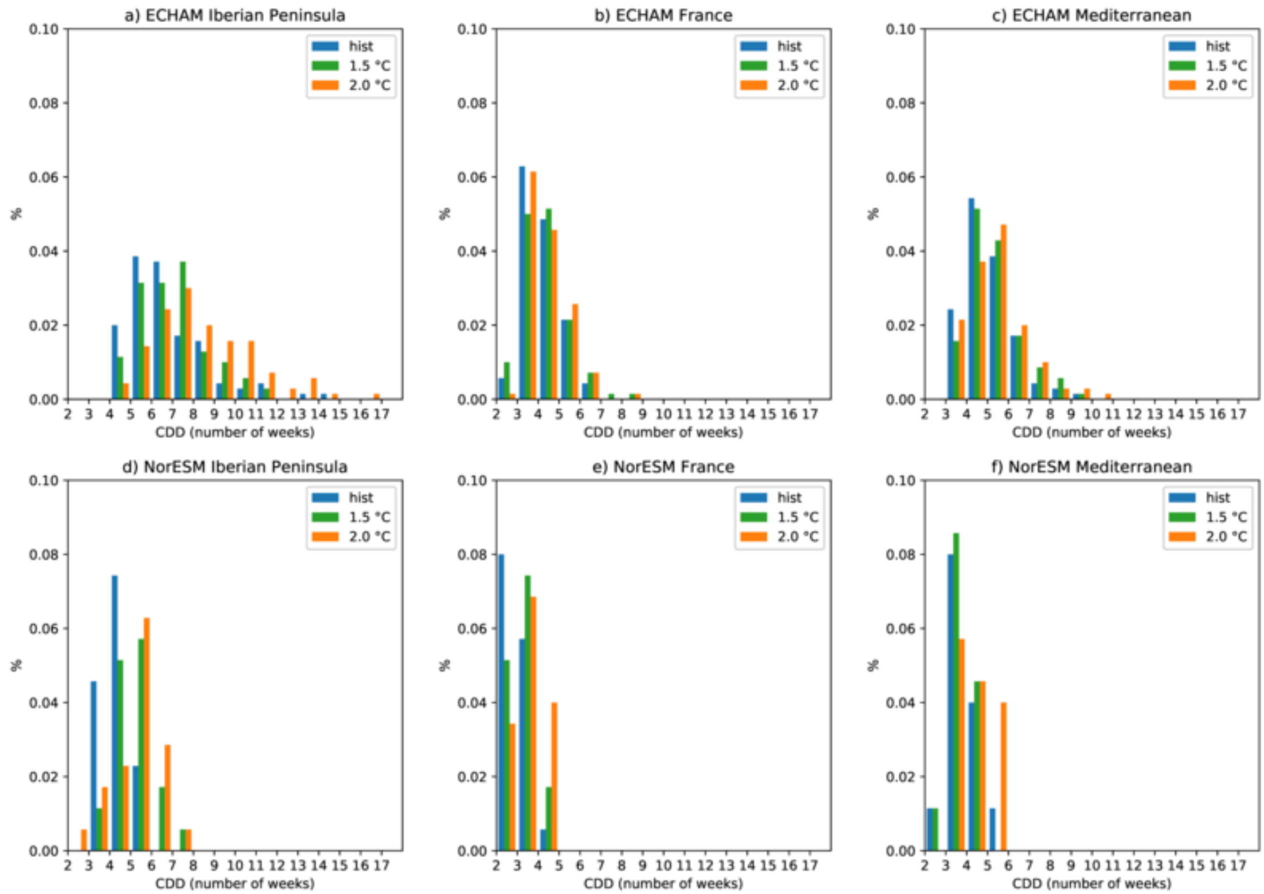


Figure 6. Duration of drought events in three PRUDENCE regions (2=Iberian Peninsula, 3=France, 7=Mediterranean) under 1.5° and 2.0° global warming for the ECHAM (top row) and NorESM (bottom row) driven ensembles. For significance see Table 1.

4. Discussion and conclusions

A unique climate dataset has been presented that enables the quantification of differences between a 1.5°C and 2.0°C warmer world compared to pre-industrial times on a regional level. This dataset can support climate change impact studies on the regional scale with physically consistent data, which is often not possible to achieve with other methods than dynamical downscaling. The use of a large ensemble (100 x 10 years) compared to alternative datasets for analysing changes under different temperature targets is especially beneficial to assess changes in highly variable meteorological parameters, such as extreme temperature and precipitation.

Here the 100 members driven by ECHAM6 provides information of statistically significant changes over relatively large and spatially homogeneous areas. In comparison, the 25-member ensemble driven by NorESM shows a much noisier spatial pattern which lowers confidence in the projected changes.

The significant differences in apparent temperature ATG28 under different global mean warming level show that a 0.5°C larger global mean warming can have considerable consequences for human health. This is especially true around the Mediterranean, where changes towards more hot and humid conditions along the coasts can have negative impacts on the population and may increase mortality due to heat stress. The tourism sector may also be negatively affected by hotter and more humid conditions. Robust estimates of percentiles and changes in percentiles can be derived from the large ensemble.

The changes across the analysed percentiles are consistent, i.e., there is no change in the shape of the distribution of ATG28.

RX5day shows a general increase over Europe which is more pronounced under higher global mean warming. More coherent spatial pattern with larger areas showing significant changes result from the larger ensemble driven by ECHAM6 (100 members) compared to the smaller ensemble driven by NorESM (25 members) and also compared to a smaller sub-ensemble driven by ECHAM6, which underlines the need for big ensemble size to reliably detect changes in highly variable quantities such as precipitation extremes.

With regard to the change in the daily rainfall intensity at the 50-year return period, a greater increase in rainfall intensity was found in the 2.0°C warmer world. Given these changes, information can be derived for local communities, which must consider changes in rainfall intensity when designing hydraulic and water resource infrastructures, as well as transportation infrastructure, including highways and bridges. Cost considerations associated with increasing rain intensity demands can be computed for up-coming design projects to ensure investments remain beneficial. The HAPPI dataset can be used to calculate other return periods, catering to the demands of individual sectors.

Robust high and low percentile changes for precipitation are still difficult to distil on a grid box level from the data because of the high variability of precipitation extremes, but methods such as spatial aggregation might help to achieve robust signals on larger spatial scale.

The changes to CDD distributions show that Spain will experience significantly more drought conditions in the future compared to the current period, even at a 1.5°C increase in GMT. For Italy, drought conditions associated with the 1.5°C simulations show non-significant changes, yet those associated with the 2.0°C simulations are significantly different to the current period, thus showing possible consequences of exceeding the 1.5°C GMT target of the Paris agreement.

The current dataset was created using the only two GCMs available at the time for downscaling, one with a reduced number of ensemble members. As more GCM ensembles become available for downscaling in the future, it will allow for new studies which can provide more robust estimates of inter-model variability/uncertainty. Nevertheless, there is currently a unique dataset targeted to the Paris agreement goals available for further analysis. Future plans include the creation of a similar regional dataset for Africa. A comparison with alternative methods for extracting the warming level is lacking and should be done in future studies.

Data availability. The datasets generated and analysed for this study can soon be found in the long-term archive of DKRZ: http://cera-www.dkrz.de/

Author contributions. KS and DJ designed the study. KS performed the regional climate model simulations. KS and CN did the analysis of the climate indices and created the plots. KS, CN, LMB and DR wrote the initial draft paper. All Authors contributed with discussions and revisions.

Competing interests. The authors declare that they have no conflicts of interest.

Acknowledgements. The authors thank three anonymous referees and Laura Suarez-Gutierrez for providing comments that helped to improve this paper. We also thank the German Climate Computing Center (DKRZ) for supplying computing time and the HAPPI consortium for providing the boundary conditions for the downscaling. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). The authors thank their co-workers for the valuable discussions. This work is part of the research project HAPPI-DE supported by the German Federal Ministry of Education and Research (BMBF) under grant number 01LS1613C. It was also supported by the Digital Earth project of the Helmholtz Association, funding code ZT-0025.

References

Anderson, G. B., Bell, M. L., Peng, R. D.: Methods to calculate the heat index as an exposure metric in environmental health research, Environ. Health Persp., 121, 1111-1119, 2013.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1-M - Part 1: Description and basic evaluation of the physical climate, Geoscie. Model Dev., 6, 687-720, 2013. https://www.geosci-model-dev.net/6/687/2013/

Bles, T., De Lange, D., Foucher, L., Axelsen, C., Leahy, C., Rooney, J.P.: Water management for road authorities in the face of climate Change: Country Comparison Report, Conference of European Directors of Roads (CEDR), Brussels, 2018. Available at https://www.cedr.eu/strategic-plan-tasks/research/cedr-call-2015/call-2015-climate-change-desk-road/ Last accessed on 13-12-2019.

Christensen, J. H., 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, 2007.

Davis, R. E., McGregor, G. R., Eneld, K. B.: Humidity: A review and primer on atmospheric moisture and human health, Environ. Res., 144, 106-116, 2016.

Deser, C., Phillips, A. S., Alexander, M. A., Smoliak, B. V.: Projecting North American climate over the next 50 years: Uncertainty due to internal variability, J. Climate, 27, 2271-2296, 2014. https://doi.org/10.1175/JCLI-D-13-00451.1

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Frich, P., Alexander, L. V., Della-Marta, P., Gleason, B., Haylock, M., Klein Tank, A. M. G., Peterson, T.: Observed coherent changes un climatic extremes during the second half of the twentieth century, Clim. Res., 18, 193-212, 2002.

Gates, W. L.: AMIP: The Atmospheric Model Intercomparison Project, B. Am. Meteorol. Soc., 73, 1962-1970, 1992. https://doi.org/10.1175/1520-0477(1992)073<1962:ATAMIP>2.0.CO;2

Giorgi, F., Jones, C. A. G.: Addressing climate information needs at the regional level: The CORDEX framework, WMO Bulletin, 58, 175-183, 2009.

Giorgi, F., Gutowski, W. J.: Regional dynamical downscaling and the CORDEX initiative, Annu. Rev. Env. Resour., 40, 467-490, 2015. https://doi.org/10.1146/annurev-environ-102014-021217

Goldstein, A., Turner, W. R., Gladstone, J., Hole, D. G.: The private sector's climate change risk and adaptation blind spots, Nat. Clim. Change, 9, 18-25, 2018.

Gutowski, W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., Nikulin, G., O'Rourke, E., Rixen, M., Solman, S., Stephenson, T., Tangang, F.: WCRP coordinated regional downscaling experiment (CORDEX): a diagnostic MIP for CMIP6, Geoscie. Model Dev., 9, 4087-4095, 2016. https://www.geosci-model-dev.net/9/4087/2016/

He, J., Soden, B. J.: The Impact of SST Biases on Projections of Anthropogenic Climate Change: A Greater Role for Atmosphere-only Models?, Geophys. Res. Lett., 43, 7745-7750, 2016.

Jacob, D., Elizalde, A., Haensler, A., Hagemann, S., Kumar, P., Podzun, R., Rechid, D., Remedio, A. R., Saeed, F., Sieck, K., Teichmann, C., Wilhelm, C.: Assessing the transferability of the regional climate model REMO to different coordinated regional climate downscaling experiment (CORDEX) regions, Atmosphere, 3, 181-199, 2012.

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., Kovats, S.,Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., 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., Yiou, P.: Euro-CORDEX: New high-resolution climate change projections for European impact research, Reg. Environ. Change, 14, 563-578, 2014. https://doi.org/10.1007/s10113-013-0499-2

Jacob, D., Kotova, L., Teichmann, C., Sobolowski, S. P., Vautard, R., Donnelly, C., Koutroulis, A. G., Grillakis, M. G., Tsanis, I. K., Damm, A., Sakalli, A., Van Vliet, M. T. H.: Climate impacts in Europe under +1.5°C global warming, Earth's Future, 6, 264-285, 2018. https://doi.org/10.1002/2017EF000710

Karl, T. R., Nicholls, N., Ghazi, A.: CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes Workshop summary, Climatic Change, 42, 3-7, 1999.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Deque, M., Gobiet, A., Goergen, K., Jacob, D., Luethi, D., van Meijgaard, E., Nikulin, G., Schaer, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., Wulfmeyer, V., Regional climate modeling on European scales: a joint standard evaluation of the Euro-CORDEX RCM ensemble, Geoscie. Model Dev. 7, 1297-1333, 2014.

Leduc, M., Mailhot, A., Frigon, A., Martel, J., Ludwig, R., Brietzke, G. B., Giguère, M., Brissette, F., Turcotte, R., Braun, M., Scinocca, J.: The ClimEx Project: A 50-Member ensemble of climate change projections at 12-km resolution over Europe and northeastern North America with the Canadian Regional Climate Model (CRCM5), J. Appl. Meteorol. Clim., 58, 663-693, 2019. https://doi.org/10.1175/JAMC-D-18-0021.1

Lierhammer, L., Mauritsen, T., Legutke, S., Esch, M., Wieners, K. H., Saeed, F.: Simulations of HAPPI (Half a degree Additional warming, Prognosis and Projected Impacts) Tier-1 experiments based on the ECHAM6.3 atmospheric model of the Max Planck Institute for Meteorology (MPI-M), World Data Center for Climate (WDCC) at DKRZ, https://doi.org/10.1594/WDCC/HAPPI-MIP-global-ECHAM6.3, 2017.

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P., Reimann, S., Smith, S. J., van den Berg, M., Velders, G. J. M., Vollmer, M., Wang, H. J.: The SSP greenhouse gas concentrations and their extensions to 2500, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-222, in review, 2019.

Mitchell, D., AchutaRao, K., Allen, M., Bethke, I., Beyerle, U., Ciavarella, A., Forster, P. M., Fuglestvedt, J., Gillett, N., Haustein, K., Ingram, W., Iversen, T., Kharin, V., Klingaman, N., Massey, N., Fischer, E., Schleussner, C.-F., Scinocca, J., Seland, Ø., Shiogama, H., Shuckburgh, E., Sparrow, S., Stone, D., Uhe, P., Wallom, D., Wehner, M., Zaaboul, R.: Half a degree additional warming, prognosis and projected impacts (HAPPI): background and experimental design. Geoscie. Model Dev., 10, 571-583, 2017. http://www.geosci-model-dev.net/10/571/2017/

Mitchell, D., James, R., Forster, P. M., Betts, R. A., Shiogama, H., Allen, M.: Realizing the impacts of a 1.5°C warmer world. Nat. Clim. Change, 6, 735-267, 2016. http://dx.doi.org/10.1038/nclimate3055

Peterson, T. C., Folland, C., Gruza, G., Hogg, W., Mokssit, A., Plummer, N.: Report on the Activities of the Working Group on Climate Change Detection and Related Rapporteurs 1998-2001. Report WCDMP-47, WMO-TD 1071, World Meteorological Organization, Geneva, 143pp, 2001.

Schleussner, C.-F., Lissner, T. K., Fischer, E. M., Wohland, J., Perrette, M., Golly, A., Rogelj, J., Childers, K., Schewe, J., Frieler, K., Mengel, M., Hare, W., Schär, M.: Differential climate impacts for policy-relevant limits to global warming: the case of 1.5°C and 2°C, Earth Syst. Dynam., 7(2), 327-351, 2016. https://www.earth-syst-dynam.net/7/327/2016/

Sieck, K., Jacob, D.: Influence of the boundary forcing on the internal variability of a regional climate model. Am. J. Clim. Change, 5, 373-382, 2016. http://dx.doi.org/10.4236/ajcc.2016.53028

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., Roeckner, E.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, J. Adv. Model Earth Sy., 5(2), 146-172, 2013. http://dx.doi.org/10.1002/jame.20015

Taylor, K. E., Stouffer, R. J., Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93(4), 485-498, 2012. https://doi.org/10.1175/BAMS-D-11-00094.1

Teichmann, C., Eggert, B., Elizalde, A., Haensler, A., Jacob, D., Kumar, P., Moseley, C., Pfeifer, S., Rechid, D., Remedio, A. R., Ries, H., Petersen, J., Preuschmann, S., Raub, T., Saeed, F., Sieck, K., Weber, T.: How does a regional climate model modify the projected climate change signal of the driving GCM: A study over different CORDEX regions using REMO, Atmosphere, 4, 214-236, 2013.

UNFCCC: Paris Agreement. United Nations Framework Convention on Climate Change, Bonn, 2015. Available at: https://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., Rose, S. K.: The representative concentration pathways: An overview, Climatic Change, 109, 5-31, 2011. http://dx.doi.org/10.1007/s10584-011-0148-z

Vautard, R., Gobiet, A., Sobolowski, S., Kjellström, E., Stegehuis, A., Watkiss, P., Mendlik, T., Landgren, O., Nikulin, G., Teichmann, C., Jacob, D.: The European climate under a 2°C global warming, Environ. Res. Lett., 9, 034006, 2014. http://stacks.iop.org/1748-9326/9/i=3/a=034006

Zhao, Y., Ducharne, A., Sultan, B., Braconnot, P., Vautard, R.: Estimating heat stress from climate-based indicators: present-day biases and future spreads in the CMIP5 global climate model ensemble, Environ. Res. Lett., 10, 084013, 2015. http://stacks.iop.org/1748-9326/10/i=8/a=084013