

## 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 were 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 extent 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 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. Barring & 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 one 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.