This paper applies the regional scaling concept (e.g. Tebaldi and Arblaster, 2014) to the distributions of annual mean precipitation (P) and precipitation minus evaporation (P-E) anomalies. The rationale behind this methodology is based on the work by Seneviratne et al., 2016, in which it is shown that local temperature and precipitation extremes scale linearly with global mean temperatures in CMIP5 CO2-increasing scenarios projections. The mean and uncertainty ranges of P and P-E local responses scale with the global mean temperature in the same way, independently of the emission scenario. This paper extends Seneviratne et al., 2016, considering annual mean local precipitation response scaling with global mean temperatures over all the SREX regions (Seneviratne et al., 2012). The impacts of uncertainty due to models internal variability, to the inter-model spread and to the scenario spread are also separately accounted for.

1 GENERAL COMMENTS

The aim of the paper is focused and clear, and it well suits in the current debate about the regional impact of global temperature change. Exploring the ranges of applicability of the pattern scaling approach allows improving the capability to communicate the impact of climate change to the stakeholders and public opinion. In this sense the assessment of regional changes in P and P-E is of outmost importance for the adaptation to future changes in local water resources.
The regional pattern scaling is here assessed in terms of a basic least squares fit, regressing annual mean P and P-E anomalies over annual mean global temperature anomalies at every grid-point. The uncertainties related to internal variability, model and scenario-related uncertainties for each model are obtained by resampling the residuals 1000-times over each gridpoint. The empirical probability distribution provides thus a way to characterize the range encompassing the median values of the regression slopes, allowing the distinction between very likely (90-100%), likely (66-100%) increase/decrease in P or P-E. The use of a basic linear scaling is justified by the lack of a-priori information about the shape of the annual mean P and P-E distributions over the various regions is not available. This shifts the focus from the choice of suitable downscaling techniques to the evaluation of uncertainty ranges attributable to the regression coefficients. In this respect, I think that the manuscript partially fails in discussing the impact of models' choice. Seneviratne et al., 2016 outlined limitations to the regional scaling pattern approach in this context. Particularly, point 4) of their discussion emphasized the risk of common biases through models for some regional phenomena. They point out that a careful model evaluation against appropriate observations would be necessary to deal with this problem. The internal variability of considered model and the multi-model ensemble uncertainty is addressed in the manuscript, but some more effort should be devoted to the evaluation of each model. Particularly, the biases induced by the imbalance in the water mass budget and the impact of different choices of the model ensembles should be carefully addressed, in order to assess the applicability of the method and the robustness of the findings.

We thank the reviewer for his/her thorough comments, which will help to improve the manuscript substantially. Based on his/her comments, we will provide additional material/figures to support our model choice. We will address all comments in a detailed point-by-point response below.

2 SPECIFIC COMMENTS

l. 22-23 p. 3: if not argued the choice of the models may look a bit arbitrary. On one hand the authors rely on Fischer et al., 2014 to choose only one model for each modelling centre. On the Env. Res. Lett.) and this may in principle prevent from consistent estimates of regional changes in P and P-E. Evaluating the long-term mean atmospheric moisture budget in control runs, identifying the regions where climate models diverge from available observations is thus a pre-requisite to this analysis. An inconsistent global mean moisture budget is a potential source of biases and the impact of adding/removing individual models should be carefully evaluated. In the framework of the regional pattern scaling, it would also be relevant to compare the atmospheric moisture budget separately over continents and oceans with the total runoff from the continents (which is a standard output in climate models, if I am not wrong is named as “mrro”), in order to provide a complete description of the hydrological cycle consistency in the model; other hand they do not consider the impact of biases in the atmospheric moisture budget. Models are known to show diverse estimates of the global mean water budget (cfr. Liepert and Lo, 2012,

As already stated in our initial response, based on the work of Liepert and Lo (ERL, 2013) in which they update their previous work (Liepert and Lo, ERL, 2012) for all CMIP5 models, we identified only MIROC5 among our subset of models that potentially exhibits a large drift. This probably further applies to FGOALS-g2, even though Liepert and Lo (2013) used FGOALS-s2. From our own experience, we are further aware that the IPSL model is associated with a drying bias over land areas. In order to assess the impact of these models on our results, we have added additional figures to the Supplementary showing the results of Figs. 5 and 6 after excluding these models. It is our assessment, that there are qualitatively no substantial differences between the different ensembles that would alter our overall conclusions.
We further agree that the identified drift is of great importance and potentially induces spurious changes in hydroclimatological storage components over long time scales. However, the global mean changes identified in Liepert and Lo (ERL, 2013) are equivalent to a maximum of only ca. 0.02mm/day. We further assess the multimodel ensemble in a probabilistic approach, providing median estimates and quantiles, thereby following the recommendation provided by Liepert and Lo (i.e.~avoiding the ensemble mean). Hence, potentially biased models within the ensemble will not substantially affect the median response provided here.

l. 23-24 p. 3: as also mentioned in l. 23-24 p. 6 the choice of ensembles with different numerosity is inherently a considerable source of uncertainty, unless one considers the 14-member and 7 (in the case of RCP6.0) and 11 (in the case of RCP2.6) members ensembles having the same statistical properties. For the same reasons motivating the previous comments, the impact of adding/removing a model from the ensembles should be carefully evaluated. To be on the safe side, I would suggest to reconsider the RCP2.6, RCP4.5 and RCP8.5 scenarios only using those models that are available in the RCP6.0 and discuss about the presence/absence of significance differences in the results. In the occurrence of significant differences I would try to identify and describe those models significantly reshaping the ensemble distribution;

We added supplementary figures showing the results of Figs. 5 and 6 after excluding all models that are not available in RCP6.0. It is here also our assessment, that there are qualitatively no substantial differences between the different ensembles that would alter our overall conclusions. We, therefore, assume that all ensembles (14-member – RCP4.5, RCP8.5; 11-member – RCP2.6; 7-member – RCP6) have similar statistical characteristics. However, please note that all results in the main text are already provided separately for the individual emission scenarios (Figs. 5,6,7,8).

l. 12-18 p. 4: the definition of variances might be clarified by labelling each sigma with a different subscript, either referring to internal variability, model uncertainty, scenario uncertainty;

Changed.

l. 15 p. 4: following above comment, it should be specified how to deal with the model uncertainty when the ensemble numerosity is lower than 14, e.g. in the RCP6.0 n=7?

See above.

l. 30-31 p. 4: to me it is not clear how the authors deal with uncertainty ranges including the zero value for the slope. Could you please expand this statement?

We rephrased this part and hope it is more understandable now.

l. 5-6 p. 5: the authors might want to comment on the fact that spatial averaging over northern high latitudes is not the same as spatial averaging at lower latitudes, and this shall be considered when discussing the significance of results at different latitudes. I wonder if one could compare circles of latitude somewhat weighting the likelihood of the changes with the cosine of latitude or the surface area covered by each circle.

It is not completely clear to us, what the reviewer requests us to do here. However, we can assure the reviewer that the spatial averages presented in this work are weighted by latitudes. Regression slopes are, however, computed at grid point scale, i.e. that the scaling coefficient at low latitudes represents a larger area than those at high latitudes. We added this information to Sec. 2. Regression slopes for the SREX regions are computed by using weighted averages of P and P-E from the particular regions.

l. 19-22 p. 6: the authors mention the different shapes of the uncertainty distributions for different SREX regions in P and P-E regression slopes. Could you please specify whether you refer to the P, P-E or both variables. Otherwise these statements appear a bit arbitrary and one might want to consider removing them;

We changed this and now clearly state, separately for P and P-E, which regions are affected.

l.1 p. 7 (and l. 25 p. 8): please specify the meaning of “significantly”;

We rephrased this part and now specifically explain what we mean here.

l. 4-6 p. 7: the authors list here a number of SREX regions characterized by larger/smaller internal variability, model uncertainty, scenario uncertainty compared to other regions. I think some more explanation might be welcomed here, rather than just listing the findings over the various regions. Why these regions, rather than others? For instance, the large model uncertainty over northern high latitudes might be related to the more relevant signal (“very likely increase” in precipitation), whereas the large internal variability over the two sides of the Tropical-Northern Pacific might reflect some relatively well understood mechanisms of inter-annual variabilities, such as the QBO (cfr. Labat et al., 2004, Geophys. Res.Lett.);

Thanks. We expanded this part and added the reference.

l. 14-16 p. 7: repetition of l. 2-4 p. 3, consider removing;

We rephrased this part to avoid any repetition.

3 TECHNICAL COMMENTS

l. 25 p. 3: remove one “in”;
l. 17-18 p. 4: replace “coefficient” with “coefficients”;
l.23 p. 6: replace “causes” with “cause”;

l. 14 p. 8: replace “extent” with “extend;
Table 2: the acronym for Northern Australia should be NAU (instead of NAS);

Thanks! Changed.

In this manuscript, Greve, Gudmundsson, and Seneviratne examine the scaling of local and regional precipitation and P-E with global mean surface temperature in climate change projections. They diagnose the likelihood of increases or decreases with warming in both quantities, and characterize and identify uncertainty due to internal variability, structural model differences, and differences in emissions scenario. To address the impacts of P and P-E on the 1.5 and 2C warming targets, they quantify the P and P-E responses and their uncertainty in each of a variety of land regions in response to the two targets. They find that the mean changes in P and P-E are indistinguishable for 1.5 and 2C, but that the two warming targets do differ in the tail of risk estimates, with a higher risk of the largest changes for 2C warming compared to 1.5C
This work makes a useful contribution to the literature, as regional changes in mean precipitation scaling have not yet been diagnosed. The maps and violin plots for individual regions are particularly useful. There are a few issues I think should be addressed to improve the manuscript.

We sincerely thank the reviewer for the positive evaluation of the manuscript and his/her comments, which substantially helped to improve the manuscript. We provide a detailed point-by-point response below.

Scientific issues

P3 line 26-27: Why omit locations where P-E<0?

We here focus on global land. Locations where P-E<0 are omitted since such conditions are generally not present over land at yearly or longer time scales. We will now further mention throughout the manuscript that our focus is on global land areas. Hence, we also decided against providing maps for ocean regions, since the underlying mechanisms and drivers are potentially very different from those over land areas. It is now our assessment that this should be analyzed individually and properly in future studies.

Figures 1, 3, and 4: In all dP versus T plots with the exception of the top panel of Fig. 1, the regressions cross through the origin. The uncertainty in the regression slope is shown as occurring entirely at the upper end of the temperature change axis. These are in conflict with the top panel of Fig. 1, where the regression slopes do not pass through the origin. Internal variability is always present, so we would expect small changes in dP even when dT=0. Is there a better way to visualize the range of regression slopes and their uncertainty? The violin plots are quite useful and do not contain these distortions.

As already stated in our initial response, in all dP versus T plots the main assumption is that (initial) P is known in case global mean temperature change dT=0. We understand that this might be unrealistic. However, we focus on the relative changes in P (dP) with changes in T, such that dP=0 when dT=0 and hence, the lines cross the origin. This approach provides an option to illustrate the uncertainty distribution as a function of temperature change. The violin plot nicely illustrates the uncertainty distribution basically for dT=1K, whereas the dP vs. dT plots illustrate the uncertainty distribution for every dT between 0K and 6K, which, in our assessment, makes it easier to assess probabilities/risks as a function of dT.

P4 line 28/30: I believe the 10th-90th percentile confidence corresponds to p=0.2, rather than p=0.1. In addition, why do you choose 10th and 90th percentile – since these are wider bounds than is customary? Why not 5 and 95 (p=0.1), or 2.5 and 97.5

(p=0.05)?

Thanks! We rephrased this part to also emphasize why we chose the 10th and 90th percentile. If the zero coefficient is within the range, it means that the probability of experiencing a scaling response of different sign compared to the median response, is, at least, 10% or higher. This corresponds to the "very likely" definition of the IPCC, which is used throughout the manuscript.

P4 line 10, P6 line 28: The methodology of Hawkins and Sutton (2009) assumes that variance is constant over the course of simulations. They only examined temperature, for which this assumption is more or less valid. It seems to me that resampling residuals would rely on the same assumption. For precipitation, it is not the case that precipitation variability is generally constant – instead, it increases in most regions (e.g., Räisänen, 2002). Do you think increasing precipitation variability would affect your uncertainty decomposition, and if so, how?

As already stated in the initial response, if the variance increases over time, the uncertainty of the sensitivity coefficient (estimated through resampling residuals) consequently also increases. However, this will not necessarily influence the decomposition of the uncertainties unless changes in precipitation variability are different between scenarios or models.

Typos and grammatical comments

P2 line 31: “comprehensive subset”: This is contradictory, since a subset is by definition not comprehensive.

We now use "representative".

P5 line 9: “A very likely decrease is rarely found only in South Africa.” I think what you mean is that a decrease with very likely confidence is found only in South Africa, and therefore it is rare; your wording means something else: that very likely decrease is often found in many places, rarely only in South Africa.

Rephrased.

P6 line 12-14: “the higher emission scenarios are usually enclosed by the low emission scenarios and the uncertainty is narrowing down”; “partly huge differences”: These phrases are not quite grammatically correct.

We rephrased the sentence.

Fig. 1: “Global mean Temperature” should probably be “Global mean Temperature Change”

Changed.

The authors investigate regional changes in precipitation (P) and water availability (expressed in terms of precipitation minus evaporation, P-E) as a function of global temperature changes in a sub-set of the CMIP5 simulations. They further decompose the uncertainties by sources related to climate variability, scenario, and model choice. They find robust changes towards wetting in northern high-latitude regions, and tendencies towards drying in subtropical regions, however associated with larger uncertainties. In particular, they also discuss changes related to political global warming limits of 1.5K and 2K.
This study is a worthwhile contribution to the literature, addressing the relevant topic of regional impact-relevant responses related to different amounts of global warming. The manuscript is mostly well written, but some clarification is needed at a few places. I also have a few more major questions related to the methodology, but think that it should be possible to clarify these with some revisions.

We sincerely thank the reviewer for the thorough and overall positive evaluation of the manuscript. His/her comments will help to improve the manuscript and we provide a detailed point-by-point response in the following.

Major comments:

(1) The authors use resampling to estimate the effects of internal climate variability. They mention that this leads to similar results as using different realisations of one model but do not show results. As estimation of different uncertainty sources, including variability, is one of the main goals of this paper, I think the authors should provide evidence that their approach by just resampling results from one run does actually lead to comparable results to analysing different runs. This seems important as usually effects of variability are estimated from a number of runs started from different climate states with respect to internal variability.

The only model from the chosen subset that was available to us providing a sufficient number of realisations (10 different realisations) was CSIRO-Mk3-6-0. We have now added plots for P similar to those provided in Figs. 3 for each SREX region comparing the slope estimates from (i) the 10 different realisations of CSIRO-Mk3-6-0 against (ii) those estimated from the resampling approach. These results are now provided as a supplementary figure in the final response. We further performed a Kolmogorov-Smirnov test to determine if the two samples, (i) 10 slope estimates from CSIRO-Mk3-6-0 and (ii) 1000 slope estimates from the resampling, are from the same parent distribution. For each SREX region, the null hypothesis, that the two samples are from the same parent distribution, can not be rejected. Respective p-Values are also provided in the new supplementary figure.

(2) The authors document some larger differences in the response between different scenarios, and seem to discuss these differences in the context of different strength of the GHG forcing. However, also the aerosol concentrations differ between the different RCP scenarios, and I wonder to which extent these scenario differences of P and P-E changes could be attributed to differences in aerosols?

We are aware of the potential influence of different aerosol concentrations on mean precipitation. We referenced in particular the work of Pendergrass and Hartmann (GRL, 2012) and Pendergrass et al. (GRL, 2015) to clarify that mean precipitation scaling depends on the emission scenario (whereas the scaling of extreme precipitation is independent of the scenario). We did, however, not explicitly mention that the differences in mean precipitation scaling can be attributed to differences in the prevailing aerosol concentration, which is now briefly discussed. Nonetheless, quantitatively assessing the extent to which the scenario-specific differences in aerosols relate to the scenario differences in P and P-E requires additional work that goes beyond the rather simplistic approach that is used here, which aims to attribute relative uncertainty contributions from different sources.

Specific and technical comments:

- Abstract, line 3: I'd remove "large" as I don't think 14 model simulations is a "large" sub set of the total number of runs available in CMIP5

We romved "large" here.

- Abstract, line 6: (Please also check throughout the text!) I suggest avoiding "dependency" when discussing the relationship of regional climate with global mean temperature. Better just say "linear relationship" here.

We now avoid the word "dependency" throughout the manuscript.

- page 1, line 21: suggest adding "public and political debate"

Changed.

- page 2, line 14: I wonder if the assumption of a linear relationship is justified when investigating changes at individual grid cells from individual ensemble members. Especially P and P-E can be rather noisy variables, strongly affected by low-frequency variability, so it might help to justify the robustness of the approach if the authors provided some tests whether linearity is a reasonable assumption in this context.

The linearity assumption is potentially valid if e.g. the residuals are (i) normally distributed and (ii) not autocorrelated. We now provide supplementary information on these statistical properties. We show for each gridpoint the number of models for which (i) the Kolmogorov-Smirnov test does not reject the null hypothesis (i.e. residuals are normally distributed) and (ii) there is no significant lag1 autocorrelation of the residuals. However, it is important to note that a further and more thorough assessment of the linearity assumption is aggravated through the large amount of data (which does e.g. not allow for any visual inspection at grid point scale, however, we added this for the SREX regions).

Nonetheless, in this study we aim to assess the scaling relationship between dT and dP or d(P-E) in terms of a single number (the scaling coefficient), which might be expressed through the slope of the regression line between the two variables. This is a simple approach, not accounting for any nonlinearities, which would ultimately lead to a more complex scaling relationship (not a single number, but potentially a more complex function). In this sense, if the linear assumption is valid, the slope estimate itself is a good representation (or a good model) of the relationship, if the linear assumption does not apply, the slope provides a bad representation (or a bad model) of the relationship. However, we do not provide deterministic estimates of the scaling coefficient alone, but we also thoroughly assess the uncertainty of the slope through resampling the residuals and repeating the regression analysis (1000 times at each grid point for each model and for each scenario). This provides us with an uncertainty estimate (even an uncertainty distribution) of the slope estimate that corresponds to the validity of linearity. In regions with high uncertainties related to the slope estimates (evidently in many arid and hyper-arid regions), both the tests for normality and autocorrelation fail for most models. In most world regions, however, residuals are normally distributed and not autocorrelated for the majority of models. This is now shown in a supplementary figure.

- page 2, line 20: This sentence is ambiguous, it seems like “this work” and “this analysis” refer to different studies, but it is not really clear what refers to what.

We rephrased the sentence.

- page 3, line 1: insert “global” warming-degree targets

Changed.

- page 3, line 23: many readers may not be aware what exactly the “prerequisites provided in Fischer et al.” are – for better readability please briefly summarise

We added this information.

- page 3, line 25: the historical runs include the year 2005, therefore if starting in 1980 this should be “26 years” and “from 2006 onwards”.
- page 3, line 26: remove duplicate word “in”

Changed.

- page 3, line 26/27: Sippel et al (https://doi.org/10.5194/hess-21-441-2017) discuss that assessing changes relative to a short reference period may lead to bias in the out-of-reference period. As the authors chose here to use only 20 years as baseline, I am wondering whether their quantifications of changes would be affected by such biases?

Regarding the work of Sippel et al. it is important to mention that our reference period (1980-1999) lies outside the study period (2000-2099) and values from the reference period are hence not used to estimate the scaling factors.

- page 3, line 27: Sentence not clear, does “majority of models” suggest that some models are treated different than others?

We rephrased this sentence.

- page 4, line 1: What kind of least squares fit did you use, e.g. ordinary or orthogonal (i.e. minimising squared differences only in y-direction or in both x and y-directions)? I think there may also be some error in the T values, so orthogonal least squares might be most appropriate?

We use here ordinary least squares and mention this in the text now. Since we want to focus on uncertainties in P and P-E, we decided against using orthogonal squares. Nonetheless, you are right and there might be also some errors in dT, but it is our assessment that these errors are not necessarily well captured through using orthogonal least squares.

- page 4, line 24: Please check, is it 10th and 90th quantile, or 25th and 75th quantile as written on Figure 2?

We now provide the correct figure showing the 10th and 90th quantile.

- page 4, line 25: Very confusing use of parentheses. If I follow your logic that the text in parentheses indicates some opposite results/statements then this sentence seems to suggest negative values always relate to (P-E) – which of cause is nonsense. Please also see this text by A. Robock (https://eos.org/opinions/parentheses-are-are-not-for-references-and-clarification-saving-space), and consider rewriting this sentence in a more readable (and clearer!) way.

You are right.We removed this sentence and rephrased this part. Thanks!

- page 5, line 5: based on only 14 models, the 90-100

We are not sure what the reviewer intends us to do here? The "very likely" decrease presented here is not just based on 14 models, but on the uncertainty distribution consisting of several thousand slope estimates that were computed through resampling the residuals.

- Page 5, line 21: Figure 4 shows the scaling relationships as regression lines, it does not explicitly show the “coefficients” as claimed in this sentence.

We changed this. However, please note that in all dP versus T plots the main assumption is that (initial) P is known in case global mean temperature change dT=0, which might be unrealistic. However, we focus on the relative changes in P (dP) with dT, such that dP=0 when dT=0 and hence, the lines cross the orign. This approach provides an option to illustrate the uncertainty distribution as a function of temperature change. The dP vs. dT plots provided here illustrate the uncertainty distribution for every dT between 0K and 6K, which, in our assessment, makes it easier to assess probalities/risks as a function of dT.

- page 5, line 26: replace “/” by “or”
- page 6, line 1: remove “much”
- page 6, line 3: replace “within” with “in”, and “parts” with “individual grid cells”
- page 6, line 4: replace “within” with “in”
- page 6, line 7: consider adding the clarification “(very) likely decrease [across all scenarios]”.
- page 6, line 21: replace “many” with “some”

Changed. Thanks!

- page 6, line 25: The structure of the Results section seems somewhat confusing. Section 3.2 is named “sources of uncertainty” – but didn’t already section 3.1 discuss one specific source of uncertainty? Please consider restructuring Section 3 more logically, or at least use more suitable sub-section names, e.g. “3.2 Comparing different sources of uncertainty”.

We now introduced different sub-section titles.

- page 7, line 26: add “with stronger global warming” (or similar) at the end of the sentence after “P-E”.

Added.

- page 9, line 1: This sentence is a literal repetition of page 7, line 30-32. Please consider rephrasing one of these instances. Otherwise this is a very nice conclusion.

Thanks! We rephrased the part on page 7.

- Figure 2: It looks like ocean areas and Antarctica were removed. This is not explicitly mentioned in the text – are these regions removed due to P-E<0 here (page 3, line 27) ? As mentioned above, please also check for consistency whether it is 10th/90th percentile (or if you wish to express in quantiles: 0.1 and 0.9), or 25th/75th

We now provide the correct figure showing the 10th and 90th quantile. We will now further mention throughout the manuscript that our focus is on global land areas. Hence, we also decided against providing maps for ocean regions, since the underlying mechanisms and drivers are potentially very different from those over land areas. It is our assessment that this should be analyzed individually in future studies.

- Figure 3/4: The text and labelling of the T-P scaling plots surrounding the map is impossible to read and should be larger. To save space you may consider to axis labels only on one plot (assuming it is equal for all), and then minimise the white space.

Changed.

- Figure 10 caption: replace “all SREX regions” with “each SREX region” – I assume this is what you actually wanted to say (having an average for each region rather than one average over all)?

Changed. Also in the other figure captions.

Regional scaling of annual mean precipitation and water availability with global temperature change

Peter Greve *[1,2], Lukas Gudmundsson[1], and Sonia I. Seneviratne[1]

[1]Institute of Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[2]International Institute for Applied Systems Analysis, Laxenburg, Austria

Correspondence to: Peter Greve (greve@iiasa.ac.at)

Abstract. Changes in regional water availability belong to the most crucial potential impacts of anthropogenic climate change, but are highly uncertain. It is thus of key importance for stakeholders to assess the possible implications of different global temperature thresholds on these quantities. Using a ~~large~~ subset of climate model simulations from the 5th phase of the Coupled Modeling Intercomparison Project (CMIP5), we derive here the sensitivity of regional changes in precipitation and precipitation minus evapotranspiration to global temperature changes. The simulations span the full range of available emissions scenarios and the sensitivities are derived using a modified pattern scaling approach. The applied approach assumes linear ~~dependencies~~ relationships on global temperature changes while thoroughly addressing associated uncertainties via resampling methods. This allows us to assess the full distribution of the simulations in a probabilistic sense. Northern high latitude regions display robust responses towards a wetting, while subtropical regions display a tendency towards drying but with a large range of responses. Even though both internal variability and the scenario choice play an important role in the overall spread of the simulations, the uncertainty stemming from the climate model choice usually accounts for about half of the total uncertainty in most regions. We additionally assess the implications of limiting global mean temperature warming to values below (i) $2K$ or (ii) $1.5K$ (as stated within the 2015 Paris Agreement). We show that opting for the $1.5K$-target might just slightly influence the mean response, but could substantially reduce the risk of experiencing extreme changes in regional water availability.

1 Introduction

Assessing regional changes in mean-annual precipitation, P, and precipitation minus evapotranspiration, $P-E$ (often also referred to as water availability), in the context of on-going global warming is of high relevance for a wide-range of socio-economic sectors. Regional differences in P and $P-E$ pose important challenges to farmers, water resources managers, stakeholders and decision-makers and a comprehensive, easily accessible communication and visualization of complex climate model output is necessary to allow for targeted adaptation and mitigation strategies.

The public and political debate on climate change is usually limited to a debate about global temperature change, which is, however, an abstract measure and does not enable end-users to infer direct implications for regional to local climate change, especially also with respect to hydroclimatological variables (Victor and Kennel, 2014; Seneviratne et al., 2016). However,

*E-Mail: greve@iiasa.ac.at

due to its omnipresence in popular climate communication, global mean temperature T could be used as a general measure of climate change and thereby enable a different communication of regional climate impacts to the public: 'The regional change of a climate variable as a function of global warming'. Many studies use approaches following this guideline, with one of the most common techniques used being summarized as 'pattern scaling'.

In this study we follow the tradition of pattern scaling, but introduce a more rigorous, probabilistic assessment of the underlying uncertainties. Common pattern scaling approaches originally have the goal to use a spatial response pattern in a certain variable (e.g. regional temperature, precipitation) that is derived from observational or (usually) from climate model data with respect to global mean temperature or CO_2-changes in order to create a large number of additional scenarios (Santer et al., 1990; Mitchell, 2003). In this study we use the "large number of additional scenarios" created by the utilized pattern scaling technique to estimate the uncertainty distribution of the response pattern in a probabilistic approach. Pattern scaling approaches have been employed in a large number of studies (see e.g. Tebaldi and Arblaster (2014) for an overview), but many common approaches to estimate the response pattern are also subject to an ongoing debate (Tebaldi and Arblaster, 2014; Herger et al., 2015; Kravitz et al., 2016). To estimate the spatial response pattern in mean-annual P and $P-E$, we adapt here a technique based on the assumption that the scaling relationship between local temperature at each gridpoint and global mean temperature is linear and that the resulting maps of regression slopes could be used as the response pattern (Solomon et al., 2009). Following a more empirical approach without a priori assumptions on the ~~dependency~~ relationship of regional variables on global temperature, it was recently shown that these findings also hold for extreme temperatures and extreme precipitation, mostly independent of emission scenarios (Seneviratne et al., 2016). This approach was further applied and extended in Wartenburger et al. (2017) by using a comprehensive set of hydroclimatological variables, including both mean-annual P and $P-E$. The assessment presented in this work builds upon ~~this analysis~~ the analysis presented in Wartenburger et al. (2017) by utilizing a similar data collection to quantify the associated response pattern.

The scaling relationship between global mean P and global warming has also been analysed in previous studies (Andrews et al., 2009; Frieler et al., 2011; Pendergrass and Hartmann, 2012; Fischer et al., 2014; Pendergrass et al., 2015). At global scales, mean precipitation scales positively with global temperature increase (Knutti et al., 2016), but the associated scaling coefficient is still subject to an ongoing debate and might not necessarily follow a linear relationship (Good et al., 2016). It was further shown that the magnitude of the scaling relationship depends on the emission scenario (Andrews et al., 2009; Frieler et al., 2011; Pendergrass and Hartmann, 2012; Pendergrass et al., 2015), whereas the scaling relationship of extreme precipitation is independent of the emission scenario (Pendergrass et al., 2015; Seneviratne et al., 2016). The scaling relationship between global mean $P-E$ and global warming was, to our knowledge, only assessed in Wartenburger et al. (2017) and more research is needed to evaluate the full range of potential impacts of regional water availability change.

We aim here to develop a methodological framework in order to assess regional changes in mean-annual P and $P-E$ over global land areas with respect to global warming by using a ~~comprehensive~~ representative subset of climate models and considering different emission scenarios. We further account for the internal variability of each projection by considering the year-to-year variability of P and $P-E$. This enables us to generate conservative estimates of the uncertainty distribution of the scaling coefficient for P and $P-E$ at every gridpoint and within specific regions.

Another issue addressed within this study is related to the implications of different global warming-degree targets on regional P and $P-E$. At the United Nations Climate Change Conference held in Paris in 2015 (COP21), most nations agreed to limit the increase in global mean temperature to values "well below $2K$" and to ideally not surpass a warming of $1.5K$ above pre-industrial conditions. Thereby, previous goals to limit global warming to “only” $2K$ global warming are significantly intensified. However, this raises the question of potential implications and differences between these “warming-degree targets” with respect to changes in many other climate variables besides the (rather abstract) value of global mean temperature and especially at regional scales (Seneviratne et al., 2016; Schleussner et al., 2016; Guiot and Cramer, 2016; James et al., 2017). The framework developed within this study allows us to directly assess regional changes in P and $P-E$ in the context of these warming-degree targets, thereby providing important and useful information to decision makers, farmers, water resources managers, stakeholders and the general public within a specific region.

First, we introduce the climate model data that is utilized within this study before describing the methodological approach that is used to estimate the uncertainty distribution of the scaling coefficients of P and $P-E$ with respect to global warming (Sec. 2). We provide in the following illustrations of the median and the range of the scaling coefficients (Section 3). Next, we comprehensively assess the uncertainty that is stemming from the choice of emission scenario (Section 3.1) and how other sources of uncertainty contribute to the total uncertainty (Section 3.2). We further apply the new framework to analyse changes between different warming degree targets (Section 4) and summarize and discuss our results also within the context of previous assessments (Sec. 5).

2 Scaling - Data and Methodology

The Coupled Model Intercomparison Project, version 5 (CMIP5) ensemble (Taylor et al., 2012) includes climate model projections forced by four Representative Concentration Pathway (RCP) emissions scenarios (Moss et al., 2010). These scenarios correspond to their relative radiative forcings reached by the end of the 21st century with respect to the preindustrial period: 2.6, 4.5, 6.0, and $8.5Wm^{-2}$ (from hereon referred to as RCP2.6, RCP4.5, RCP6.0 and RCP8.5). We use a total of 14 climate models selected based on prerequisites provided in Fischer et al. (2014), which are only one model from each modeling center and thereof the newest with the highest resolution. Please note that not all climate models provide data for all emission scenarios (see Table 1). We consider a time period of 120 years beginning in 1980 and ending in 2099 comprising historical simulations for the first ~~25~~ 26 years which emerge into simulations of the respective emission scenarios from ~~2005~~ 2006 onwards. The first 20 years (1980-1999) are used as a common baseline period and values in ~~in~~ mean annual P and $P-E$ are assessed in relative terms [%] with respect to the baseline period. ~~Please~~ Since we focus here on global land areas, please note that in case $P-E<0$ for ~~the majority of~~ all models and scenarios, those ~~gridpoints~~ grid points were neglected.

For each model m and each emission scenario s within the 100-year period 2000-2099 (yr), ~~these~~ the relative values of precipitation, $P_{m,s,yr}$, and precipitation minus evapotranspiration, $(P-E)_{m,s,yr}$, are regressed at each gridpoint (or averaged over a certain region) against mean-annual global temperature, $T_{m,s,yr}$). We use ~~a~~ an ordinary least squares fit to estimate the parameters of the linear equation:

$$P'_{m,s,yr} = r_{m,s} \cdot T_{m,s,yr} + I_{m,s} \tag{1}$$

with $r_{m,s}$ denoting the regression slope and $I_{m,s}$ the intercept (and likewise for $(P-E)_{m,s,yr}$). The slope itself provides us with an estimate of the regional scaling coefficient of P against global changes in T.

Given the annual residuals $R_{m,s,yr} = P_{m,s,yr} - P'_{m,s,yr}$, the uncertainty of the regression slope $r_{m,s}$ is assessed by resampling years yr' of the residuals ($R_{m,s,yr'}$) and fitting the regression slope against the new pairs ($T_{m,s,yr}$,$P'_{m,s,yr} + R_{m,s,yr'}$). Repeating this approach 1000-times at each gridpoint (or within each specific region) provides us with a comprehensive uncertainty measure $\epsilon_{m,s}$ of each model- and scenario-specific regression slope $r_{m,s}$ for both P and $P-E$. We like to point out that the uncertainty estimated through resampling residuals results in very similar results as computing the uncertainty through using different realisations of a single model (~~not shown).~~ this is shown for CSIRO-Mk3-6-0 in the supplementary information). We further note, that scaling coefficients in low latitude regions represent the response within a larger area compared to those in high latitude regions. To also test the validity of the linearity assumption, we assessed (i) if the Kolmogorov-Smirnov test does not reject the null hypothesis that the annual residuals $R_{m,s,yr}$ are normally distributed and (ii) if there is no significant lag-1yr autocorrelation of all $R_{m,s,yr}$. Following this approach, in the majority of world regions and for most models, the linearity assumption is potentially valid. However, please note for the following sections, that in many hyper-arid regions both tests fail for the majority of models.

This approach allows us to distinguish between three different sources of uncertainty. As illustrated in the conceptual Fig. 1, these are (i) internal variability, σ_i, representing the uncertainty stemming from interannual variability for each model under each scenario, (ii) the model uncertainty, σ_m, related to the uncertainty across all models and for a specific scenario (e.g. in terms of variance: ~~$\sigma = 1/n\sum_{m=1}^{n}(r_{m,rx} - \mu_r)^2$~~ $\sigma_m = 1/n\sum_{m=1}^{n}(r_{m,rx} - \mu_r)^2$, with $n = 14$ models and μ_r denoting the average of all scaling coefficient) and (iii) the scenario uncertainty, σ_s, related to the uncertainty across all scenarios and for a specific model or the multi-model mean (e.g. in terms of variance: ~~$\sigma = 1/n\sum_{rx=1}^{n}(r_{m,rx} - \mu_r)^2$~~ $\sigma_s = 1/n\sum_{rx=1}^{n}(r_{m,rx} - \mu_r)^2$, with $n = 4$ scenarios and μ_r denoting the average of all scaling ~~coefficient~~coefficients). The total uncertainty denotes the uncertainty across all models, all scenarios and considering the internal uncertainty. Please note, that this approach of attributing uncertainties is very simplistic and neglects any potential relationship between the individual sources of uncertainty, but is suitable and useful to provide a general measure of the underlying uncertainty sources.

3 Scaling - Results

Considering the total uncertainty across all models and scenarios and by additionally including the internal variability, we are able to estimate the uncertainty distribution of the regional scaling coefficient of P ($P-E$) against globally-averaged T. Displayed in Fig. 2 are the median and the 10th and 90th quantile of the uncertainty distribution of the scaling coefficient for both P and $P-E$ at each gridpoint. ~~Positive (negative) valuesdenote an increase (decrease) in~~ The median scaling coefficient shows positive values, and hence an increase in both P ~~(~~and $P-E$ ~~)~~ with increasing T~~. The median scaling coefficient shows positive values~~, in most parts of the northern high latitudes and Asia, but also in eastern Africa for both P and $P-E$. Negative values, and hence a decrease in both P and $P-E$ with increasing T, are found in the Mediterranean region, southern Africa, Australia and in parts of West Africa, as well as Central and South America. Comparing the 10th and 90th quantiles of the uncertainty distribution shows the range of possible scaling coefficients. This range does, in most regions and especially for $P-E$, include the zero coefficient, which means that the ~~median response is not significant in classical sense ($p=0.1$) and that there is a non-negligible chance of~~ probability of experiencing a scaling response of different sign compared to the median response~~switching signs~~, is, at least, 10% or higher. The range is further generally much larger for $P-E$, pointing towards overall higher uncertainties in the estimation of the scaling relationship. The range is especially large in most subtropical regions (e.g. Sahara, Arab peninsula, India, Australia, etc.). Regions showing a significant increase of P and $P-E$ with global warming are located mainly in the northern high latitudes.

Fig. 3 summarizes these findings by qualitatively showing the probability of experiencing either a positive or negative scaling response in P with respect to global warming. A very likely increase ($90-100\%$ probability) in regional P with ongoing global warming is hence found only within regions of the northern high latitudes, whereas a likely increase ($66-100\%$ probability) is located also in many parts of Asia and North America and to a minor extent also in some regions of South America and Africa. A likely decrease is located in most parts of the Mediterranean region, southern Africa, northeastern South America, Central America and along the Australian coastal regions. A ~~very likely decrease is rarely found only~~ decrease that is very likely is only found in South Africa. Most other regions show either uncertain or no change. Fig. 3 also ~~shows illustrations~~ illustrates selected quantiles of the uncertainty distribution of the scaling coefficient of P as a function of global temperature increase for a comprehensive subset of SREX regions (Seneviratne et al., 2012) as outlined in the map (see also Table 2 for more information). Very certain responses within the SREX-regions are only found for those in the northern high latitudes (ALA, NAS, NEU), while most other regions show a large spread of the uncertainty distribution (especially e.g. in NEB, NAU).

Similarly for $P-E$, Fig. 4 displays a very likely increase ($90-100\%$ probability) in regional $P-E$ with ongoing global warming for an even smaller portion of land in the northern high latitudes, whereas a likely increase ($66-100\%$ probability) is located throughout the northern high latitudes and similarly to P in many parts of Asia and North America and to a minor extent also in some regions in South America and Africa. A likely decrease is located in parts of the Mediterranean region, southern Africa, northeastern South America, Central America and some parts of Australia. A very likely decrease is only found for single gridpoints, primarily in Central America. Most other (and when compared to P an even higher number of) regions show either uncertain or no change. Fig. 4 also ~~shows illustrations~~ illustrates selected quantiles of the uncertainty distribution of the scaling coefficient of $P-E$ as a function of global temperature increase for the same set of SREX regions as shown in Fig. 3. Very certain responses are again only found in the northern high latitudes (ALA, NAS) and southern Asia (SAS), while most other regions show a very large and even larger spread of the uncertainty distribution when compared to estimates of P.

Please note that the results for both P and $P-E$ are neither substantially influenced by the unequal number of available models per scenario nor individual climate models of the ensemble that potentially exhibit a large hydroclimatological drift (Liepert and Previdi, 2012; Liepert and Lo, 2013) (refer to the supplementary material for more information).

3.1 Scenario uncertainty

The probability of experiencing an increase ~~/~~or decrease in regional P and $P-E$ with global warming depends on the emission scenario. At global scales, mean precipitation scaling was shown to depend on the emission scenario (Andrews et al., 2009; Frieler et al., 2011; Pendergrass and Hartmann, 2012; Pendergrass et al., 2015), whereas the scaling of extreme precipitation is independent of the emission scenario (Pendergrass et al., 2015; Seneviratne et al., 2016). Here we assess the ~~dependence~~ relationship of regional changes in P and $P-E$ on the emission scenario by analysing the uncertainty distributions of the scaling coefficient for each scenario individually. A conceptual representation of the probability of the scaling coefficient being postive/negative is displayed for P in Fig. 5 and for $P-E$ in Fig. 6 (similar to the total uncertainty as shown in Fig. 3 and Fig. 4). In general, the fraction of regions showing either likely or very likely changes is increasing with the emission scenario for both P and $P-E$, pointing towards a larger uncertainty in the estimation of the scaling coefficient in case the climate change forcing is weak (RCP2.6, RCP4.5). Further, regions showing very likely changes are ~~much~~ more common under high emission scenarios (RCP6.0, RCP8.5). The drying response in the Mediterranean region is e.g. not evident when considering the RCP2.6 scenario alone, whereas a very likely decrease is found ~~within~~ in the RCP8.5 scenario. In fact, ~~parts~~ individual grid points of the Mediterranean region (e.g. in central Spain) even show a likely increase in P and $P-E$ ~~within~~ in the RCP2.6 scenario. On the other hand, a likely drying response in parts of central and northen Australia found in RCP2.6 disappears for higher emission scenarios for P, or even turns into a wetting response for $P-E$. Robust signals are, again, found in most parts of the northern high latitudes, showing a (very) likely increase across all emission scenarios. A (very) likely decrease across all scenarios is further found for parts of southern Africa and parts of the Amazon region. Also note here that the overall conclusions for both P and $P-E$ are not influenced by the unequal number of available models per scenario (see supplementary information).

A more detailed look on the underlying uncertainty distributions for P ($P-E$) within each SREX-region is provided in Fig. 7 (Fig. 8). It is clearly evident that the uncertainty is largest for low emission scenarios throughout all regions and in most cases lowest for the RCP8.5 scenario (with overall larger uncertainties in $P-E$, please note the different y-axis scales). Additionally, the uncertainty distribution of the higher emission scenarios ~~are usually enclosed by the~~ is usually found within the uncertainty range of the low emission scenarios~~and the uncertainty is narrowing down~~, pointing towards a more certain signal. However, there are ~~partly~~ often huge differences regarding the median response and the location and shape of the uncertainty distribution of a particular emission scenario with respect to other emission scenarios. This is especially evident when comparing low to high emission scenarios. Most prominently, for many regions (e.g. WNA, CAM, MED, WAS, CAS, please see Table 2 for more information on the acronyms) the uncertainty distribution of the RCP6.0 or RCP8.5 scenario is located mainly within the lowest tercile of the RCP2.6 scenario, leading to a dryer response in P ($P-E$) with global warming for high emission scenarios. This finding is, however, reversed in a few other regions (especially NAU and for $P-E$, to a certain extent also in ALA, EAF, SEA). The shapes of the uncertainty distributions for both P and $P-E$ are also different between regions and emission scenarios. While the distributions for the low emission scenarios are, in most cases, unimodal, there are bimodal distributions in ~~many~~ some (e.g. for P: NEU, WAS, CAS, for $P-E$: MED, EAS, SEA) and even multimodal distributions in a few other regions (e.g. for P: AMZ, NAS, for $P-E$: CNA, WAF, EAF) for the higher emissions scenerios (especially RCP8.5). Please note, however, that not all models provide data for the RCP2.6 and RCP6.0 emission scenarios, which might also ~~causes~~ cause differences between those and the other scenarios (see Table 1 for more information).

3.2 Sources of uncertainty

Besides the scenario uncertainty we also introduced two other sources of uncertainty in Sec. 2, the internal variability and the model uncertainty which contribute to the total uncertainty. Here we assess the fraction of uncertainty which each source contributes to the total uncertainty. We follow the approach of Hawkins and Sutton (2009), which was also adapted in Orlowsky and Seneviratne (2013). Therefore we compare (i) the average over the variances of the uncertainty distributions of each model under each emission scenario (internal uncertainty, , σ_i), (ii) the average of the variances of scenario-specific uncertainty distributions of each model (model uncertainty, , σ_m) and (iii) the variance of the averages of all uncertainty distributions within a specific scenario (scenario uncertainty, σ_s). Even though this approach of attributing uncertainties is very simplistic (see section 2), it provides basic information on the composition of different uncertainty sources within the total uncertainty. The percentage of the total uncertainty that stems from a particular source is illustrated for both P and $P-E$ in Fig. 9. For all SREX regions, there is generally no ~~uncertainty source which is significantly dominating~~ single source of uncertainty contributing more than 80% to the total uncertainty. However, in most regions the largest source of uncertainty stems from model uncertainty, which is contributing up to ca. 3/4 of the total uncertainty in some regions and is especially large in most northern high latitude regions (CGI, NEU, NAS, except ALA), possibly related to the relatively large and certain scaling response in these regions. Internal variability contributes between $20-40\%$ to the total uncertainty, with highest values found ~~for contrasting regions such as ALA, CEU and SAU~~in particular for several regions adjacent to the Pacific Ocean (e.g. WNA, SEA, SAU), probably related to previously assessed modes of inter-annual variability, such as the Quasibiennial Oscillation, QBO (e.g. Labat et al., 2004). Internal variability seems to be rather low in other tropical to subtropical regions, such as AMZ, EAS and NAU for P and ~~in~~ NEB, WAF and SAS for $P-E$. Scenario uncertainty contributes between ca. $5-30\%$ to the total uncertainty with those regions reaching highest values that have differing locations of the uncertainty distributions between low and high emission scenarios as shown in Fig. 7 and Fig. 8 (~~especially~~ e.g. WNA, ENA, CAM, MED, WAS, CAS, and NAU). Differences in P between emission scenarios are further not solely caused by varying radiative forcing due to differing greenhouse gas emissions, but also due to differences in the black carbon forcing (Pendergrass and Hartmann, 2012). It is further interesting to note that scenario uncertainty is generally lower and internal variability generally larger for $P-E$ when compared to P. However, even though the scenario uncertainty is the overall weakest source of uncertainty in most regions, it is by no means negligible. Please note, again, that the scenario uncertainty interferes especially with the rather large model uncertainty and we do not account for such relationships in this approach.

4 Application: Assessing warming degree limits

~~At the United Nations Climate Change Conference held in Paris in 2015 (~~As agreed in COP21~~), most nations agreed to limit~~ , the increase in global mean temperature should be limited to values well below the previously set goal of $2K$~~and to limit warming to~~ , preferably to not more than $1.5K$ above pre-industrial conditions. However, global mean temperature is an abstract value and provides no information about direct implications at regional scales and with respect to other climate variables such as regional, mean-annual P or $P-E$. The framework developed within this study enables us to directly assess the regional response of P and $P-E$ to these targets and to study differences between them. Using the uncertainty distributions of each SREX region and scaling them to either $1.5K$ or $2K$ (as illustrated in Fig. 10 for P and in Fig. 11 for $P-E$) allows us to study differences both in the median response as well as in the tails of the distribution. It is, however, naturally evident that in regions with a weak median scaling response the difference between the warming degree targets is small regarding the median response itself (e.g. AMZ, CEU, WAS), whereas in regions with a stronger median scaling (e.g. ALA, NEU, NAS) an additional $0.5K$ warming could lead to substantial differences. Nonetheless, even though the difference in the median might be small, the difference in the tails of the uncertainty distributions are in most cases significant and stress an increased risk of experiencing strong changes in P and $P-E$. As an example for the Mediterranean region (MED): The median responses of P to $1.5K$ global warming vs $2K$ global warming are not strongly different, while there are stronger differences at the tails, showing that the 1.5K limit would avoid a decrease of P of more than 20%, which can on the other hand not be excluded with the 2K limit. This behavior is even more evident for $P-E$ and also occurs in regions with almost no median response (e.g. CEU). Also the irregularity of the distribution further amplifies this behavior in some regions. In summary, opting for a low warming degree target (such as $1.5K$) might just slightly influence the mean response but could substantially reduce the risk of experiencing more extreme changes in regional P and $P-E$.

5 Conclusions

We developed here a framework building upon the pattern scaling approach to assess regional changes in mean-annual P and $P-E$ with respect to global mean T-increase by utilizing a comprehensive subset of climate models and considering all available emission scenarios. We further took into account internal variability from each projection by accounting for the year-to-year variability of P ($P-E$). This enabled us to assess a conservative estimate of the uncertainty distribution of the scaling coefficient of P ($P-E$) to global warming at every gridpoint or within SREX regions.

Analysing maps of the median response and the responses in the 10th and 90th quantile of the gridpoint-specific uncertainty distributions showed low uncertainties and positive scaling coefficients (thereby a certain increase in P and $P-E$ with global warming) within most northern high latitude regions. Slight decreases in the median response together with large uncertainties (and thereby an uncertain decrease in P and $P-E$ with global warming) are found for most subtropical regions. Uncertainties are, however, larger for estimates of $P-E$ and do hence not permit robust conclusions for many regions. Our results support previous findings of hydroclimatological changes (Greve and Seneviratne, 2015), but provide a new, probabilistic and rigorous perspective on the assessment of uncertainties in regional hydroclimatological changes under conditions of ongoing global warming and ~~extent~~ extend the wealth of studies investigating pattern scaling approaches of climate variables (Tebaldi and Arblaster, 2014; Herger et al., 2015).

Assessing scenario-specific uncertainty distributions revealed strong regional differences between different emission scenarios. It is evident that weaker climate change signals within the low-emission scenarios (RCP2.6, RCP4.5) lead to high uncertainties in the estimation of scaling coefficients. A very likely change in regional P only emerges under high emission scenarios (RCP6.0, RCP8.5) and is even less likely to occur for $P-E$. In some regions low emission scenarios further show likely opposite changes compared to changes identified in higher emission scenarios (both switching from a likely wetting to a very likely drying in parts of MED, or from a likely drying to no change in P or even a likely wetting in $P-E$ in NAU). A closer look a the uncertainty distributions shows large differences both in location and shape across regions and emission scenario. However, in most cases higher emission scenarios point towards a dryer response than low emission scenarios (with a few regions showing, however, the opposite behavior).

This led us to the analysis of the relative contribution of single sources of uncertainty to the overall uncertainty. It is shown that model uncertainty is largest in most regions, but is~~not significantly dominating~~, however, in no case contributing more than 80% to the overall uncertainty. It is therefore important that both internal variability and scenario uncertainty are considered as well in order to get a complete picture of the total uncertainty. Comparing mean annual P and $P-E$ shows that scenario uncertainty is generally lower and internal variability generally larger for $P-E$.

We further assessed the implications of different warming degree limits on changes in regional P and $P-E$. At the COP21, most nations agreed to limit the increase in global mean temperature to values well below the previously set goal of $2K$ and to consider limiting warming to not more than $1.5K$ above pre-industrial conditions. Comparing these two targets reveals naturally little differences in the mean response in regions where the mean response is small anyway. However, since uncertainties are large, especially for $P-E$, there is a nonlinear increase in the risk of experiencing more extreme changes. Therefore, opting for a low warming degree target (such as $1.5K$) might just slightly influence the mean response but could substantially reduce the risk of experiencing extreme changes in regional P and $P-E$. This means that even though the discussion about the implications of $1.5K$ vs. $2K$ global warming might be moot for the mean response, it is, given the underlying, large uncertainties of climate projections, absolutely necessary to more closely investigate the potentially large increase in the risk of experiencing extreme change. This is especially important in order to enable robust decision making to ensure adequate development pathways and to avoid the risk of maladaptation; in the specific case of changes in mean-annual P and $P-E$ this is e.g. of high relevance for water resources managers and farmers.

Data availability. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organisation for Earth System Science Portals. The data used in this study are available through the Coupled Model Intercomparison Project Phase 5 at http://pcmdi9.llnl.gov/esgf-web-fe/

Competing interests. The authors declare no competing financial interests.

Acknowledgements. We thank Jakob Zscheischler for providing useful input regarding methodological aspects of this study and Urs Beyerle and Jan Sedlacek for processing the CMIP5 data.

References

Andrews, T., Forster, P. M., and Gregory, J. M.: A Surface Energy Perspective on Climate Change, J. Climate, 22, 2557–2570, doi:10.1175/2008JCLI2759.1, http://journals.ametsoc.org/doi/full/10.1175/2008JCLI2759.1, 2009.

Fischer, E. M., Sedláček, J., Hawkins, E., and Knutti, R.: Models agree on forced response pattern of precipitation and temperature extremes, Geophysical Research Letters, 41, 2014GL062 018, doi:10.1002/2014GL062018, http://onlinelibrary.wiley.com/doi/10.1002/2014GL062018/abstract, 2014.

Frieler, K., Meinshausen, M., Schneider von Deimling, T., Andrews, T., and Forster, P.: Changes in global-mean precipitation in response to warming, greenhouse gas forcing and black carbon, Geophys. Res. Lett., 38, L04 702, doi:10.1029/2010GL045953, http://onlinelibrary.wiley.com/doi/10.1029/2010GL045953/abstract, 2011.

Good, P., Booth, B. B. B., Chadwick, R., Hawkins, E., Jonko, A., and Lowe, J. A.: Large differences in regional precipitation change between a first and second 2 K of global warming, Nature Communications, 7, 13 667, doi:10.1038/ncomms13667, http://www.nature.com/ncomms/2016/161206/ncomms13667/full/ncomms13667.html, 2016.

Greve, P. and Seneviratne, S. I.: Assessment of future changes in water availability and aridity, Geophysical Research Letters, 42, 5493–5499, doi:10.1002/2015GL064127, http://onlinelibrary.wiley.com/doi/10.1002/2015GL064127/abstract, 2015.

Guiot, J. and Cramer, W.: Climate change: The 2015 Paris Agreement thresholds and Mediterranean basin ecosystems, Science, 354, 465–468, doi:10.1126/science.aah5015, http://science.sciencemag.org/content/354/6311/465, 2016.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Predictions, Bulletin of the American Meteorological Society, 90, 1095–1107, doi:10.1175/2009BAMS2607.1, http://journals.ametsoc.org/doi/abs/10.1175/2009BAMS2607.1, 2009.

Herger, N., Sanderson, B. M., and Knutti, R.: Improved pattern scaling approaches for the use in climate impact studies, Geophys. Res. Lett., 42, 2015GL063 569, doi:10.1002/2015GL063569, http://onlinelibrary.wiley.com/doi/10.1002/2015GL063569/abstract, 2015.

James, R., Washington, R., Schleussner, C.-F., Rogelj, J., and Conway, D.: Characterizing half-a-degree difference: a review of methods for identifying regional climate responses to global warming targets, WIREs Clim Change, 8, n/a–n/a, doi:10.1002/wcc.457, http://onlinelibrary.wiley.com/doi/10.1002/wcc.457/abstract, 2017.

Knutti, R., Rogelj, J., Sedláček, J., and Fischer, E. M.: A scientific critique of the two-degree climate change target, Nature Geoscience, 9, 13–18, doi:10.1038/ngeo2595, http://www.nature.com/ngeo/journal/v9/n1/full/ngeo2595.html, 2016.

Kravitz, B., Lynch, C., Hartin, C., and Bond-Lamberty, B.: Exploring precipitation pattern scaling methodologies and robustness among CMIP5 models, Geosci. Model Dev. Discuss., 2016, 1–37, doi:10.5194/gmd-2016-258, http://www.geosci-model-dev-discuss.net/gmd-2016-258/, 2016.

Labat, D., Ronchail, J., Callede, J., Guyot, J. L., De Oliveira, E., and Guimarães, W.: Wavelet analysis of Amazon hydrological regime variability, Geophysical Research Letters, 31, doi:10.1029/2003GL018741, http://onlinelibrary.wiley.com/doi/10.1029/2003GL018741/full, 2004.

Liepert, B. G. and Lo, F.: CMIP5 update of ‘Inter-model variability and biases of the global water cycle in CMIP3 coupled climate models’, Environ. Res. Lett., 8, 029 401, doi:10.1088/1748-9326/8/2/029401, http://stacks.iop.org/1748-9326/8/i=2/a=029401, 2013.

Liepert, B. G. and Previdi, M.: Inter-model variability and biases of the global water cycle in CMIP3 coupled climate models, Environ. Res. Lett., 7, 014 006, doi:10.1088/1748-9326/7/1/014006, http://stacks.iop.org/1748-9326/7/i=1/a=014006, 2012.

Mitchell, T. D.: Pattern Scaling: An Examination of the Accuracy of the Technique for Describing Future Climates, Climatic Change, 60, 217–242, doi:10.1023/A:1026035305597, http://link.springer.com/article/10.1023/A:1026035305597, 2003.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J.: The next generation of scenarios for climate change research and assessment, Nature, 463, 747–756, doi:10.1038/nature08823, 2010.

Orlowsky, B. and Seneviratne, S. I.: Elusive drought: uncertainty in observed trends and short- and long-term CMIP5 projections, Hydrol. Earth Syst. Sci., 17, 1765–1781, doi:10.5194/hess-17-1765-2013, http://www.hydrol-earth-syst-sci.net/17/1765/2013/, 2013.

Pendergrass, A. G. and Hartmann, D. L.: Global-mean precipitation and black carbon in AR4 simulations, Geophys. Res. Lett., 39, L01 703, doi:10.1029/2011GL050067, http://onlinelibrary.wiley.com/doi/10.1029/2011GL050067/abstract, 2012.

Pendergrass, A. G., Lehner, F., Sanderson, B. M., and Xu, Y.: Does extreme precipitation intensity depend on the emissions scenario?, Geophysical Research Letters, 42, 2015GL065 854, doi:10.1002/2015GL065854, http://onlinelibrary.wiley.com/doi/10.1002/2015GL065854/abstract, 2015.

Santer, B., Wigley, M., Schlesinger, M., and Mitchell, J.: Developing climate scenarios from equilibrium GCM results, Tech. Rep., pp. Hamburg, Germany, 1990.

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., and Schaeffer, M.: Differential climate impacts for policy-relevant limits to global warming: the case of 1.5 °C and 2 °C, Earth Syst. Dynam., 7, 327–351, doi:10.5194/esd-7-327-2016, http://www.earth-syst-dynam.net/7/327/2016/, 2016.

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

Seneviratne, S. I., Donat, M. G., Pitman, A. J., Knutti, R., and Wilby, R. L.: Allowable CO2 emissions based on regional and impact-related climate targets, Nature, 529, 477–483, doi:10.1038/nature16542, http://www.nature.com/nature/journal/v529/n7587/full/nature16542.html, 2016.

Solomon, S., Plattner, G.-K., Knutti, R., and Friedlingstein, P.: Irreversible climate change due to carbon dioxide emissions, PNAS, 106, 1704–1709, doi:10.1073/pnas.0812721106, http://www.pnas.org/content/106/6/1704, 2009.

Taylor, K., Stouffer, R., and Meehl, G.: An Overview of CMIP5 and the Experiment Design, Bull. Amer. Meteor. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Tebaldi, C. and Arblaster, J. M.: Pattern scaling: Its strengths and limitations, and an update on the latest model simulations, Climatic Change, 122, 459–471, doi:10.1007/s10584-013-1032-9, http://link.springer.com/article/10.1007/s10584-013-1032-9, 2014.

Victor, D. G. and Kennel, C. F.: Climate policy: Ditch the 2 °C warming goal, Nature News, 514, 30, doi:10.1038/514030a, http://www.nature.com/news/climate-policy-ditch-the-2-c-warming-goal-1.16018, 2014.

Wartenburger, R., Hirschi, M., Donat, M. G., Greve, P., Pitman, A. J., and Seneviratne, S. I.: Changes in regional climate extremes as a function of global mean temperature: an interactive plotting framework, Geosci. Model Dev. Discuss., 2017, 1–30, doi:10.5194/gmd-2017-33, http://www.geosci-model-dev-discuss.net/gmd-2017-33/, 2017.

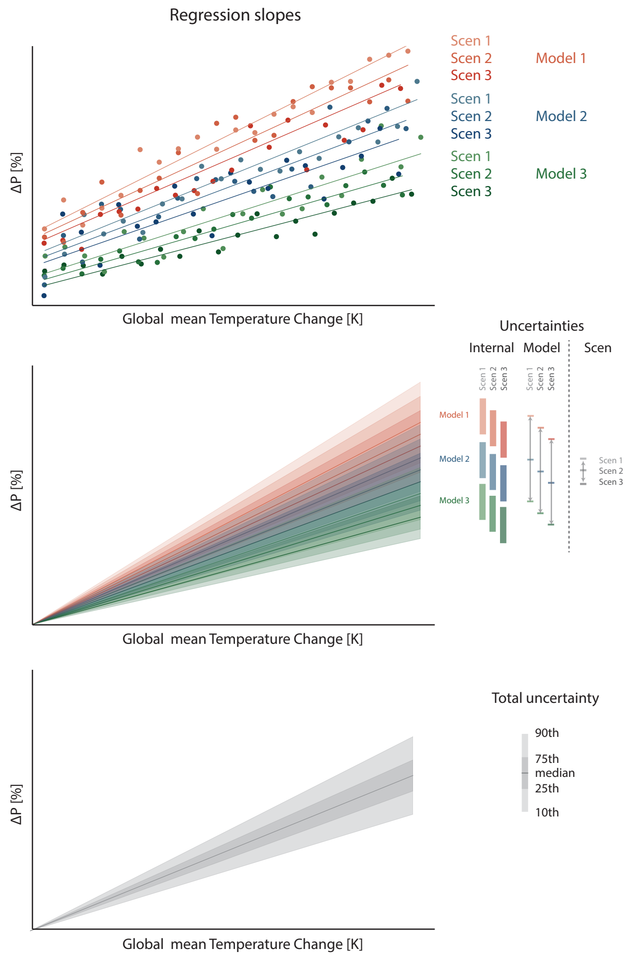

Figure 1. Conceptual illustration of deriving the uncertainty distribution of the scaling coefficient of P with respect to global warming from multiple models forced by multiple emissions scenarios. (Top) For each model under each scenario we regress the relative change in mean-annual P (ΔP, with respect to a baseline period) against global mean T. We thereby obtain the regression slope which is the scaling coefficient of P to global warming. The year-to-year variability further causes the estimate of the slope to be uncertain. We account for this uncertainty by numerically estimating the uncertainty distribution of each model- and scenario-specific regression slope through resampling the residuals in a bootstrapping approach. (Middle) This uncertainty is associated with every model run and represents the internal variability. The average of the uncertainties stemming from the range of all individual models within a certain scenario represents the model uncertainty and the uncertainty associated with the range of all scenario-specific multi-model means represents the scenario uncertainty. The uncertainty distribution is illustrated here as a function of global temperature increase. (Bottom) The Total uncertainty combines all sources of uncertainty and provides a conservative estimate of regional ΔP as a function of global warming, that can be used to assess either the median response or to study changes in any other quantile of the uncertainty distribution.

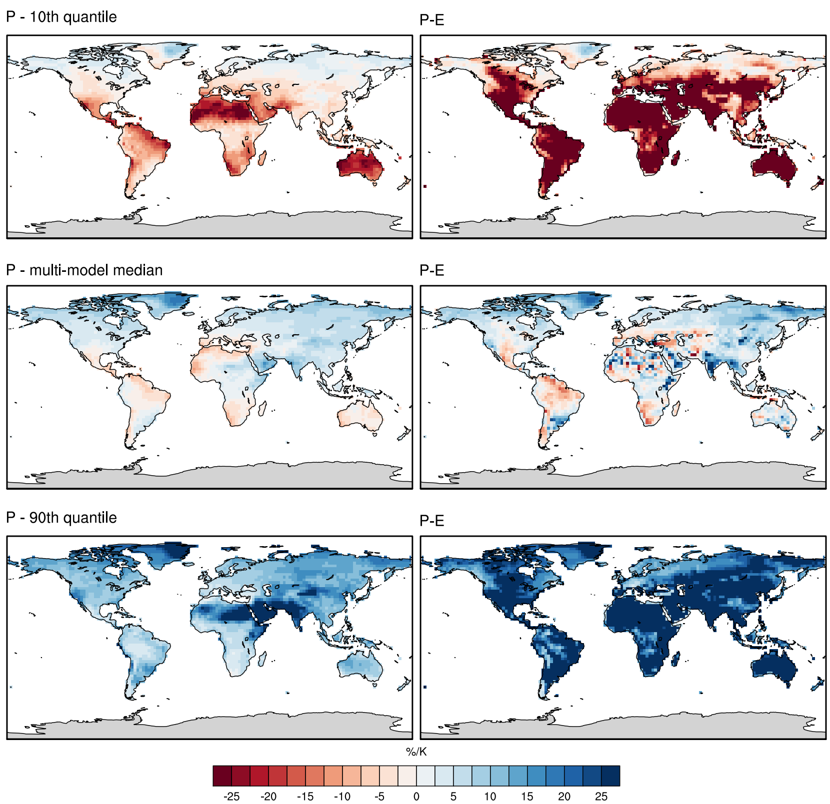

Figure 2. Median (middle), 10th (top) and 90th quantile (bottom) of the sensitivity of P (left) and $P - E$ (right) to changes in global mean temperature $[\%/K]$. A total of 14 CMIP5 models and all scenarios (RCP2.6, RCP4.5, RCP6.0, RCP8.5) are considered.

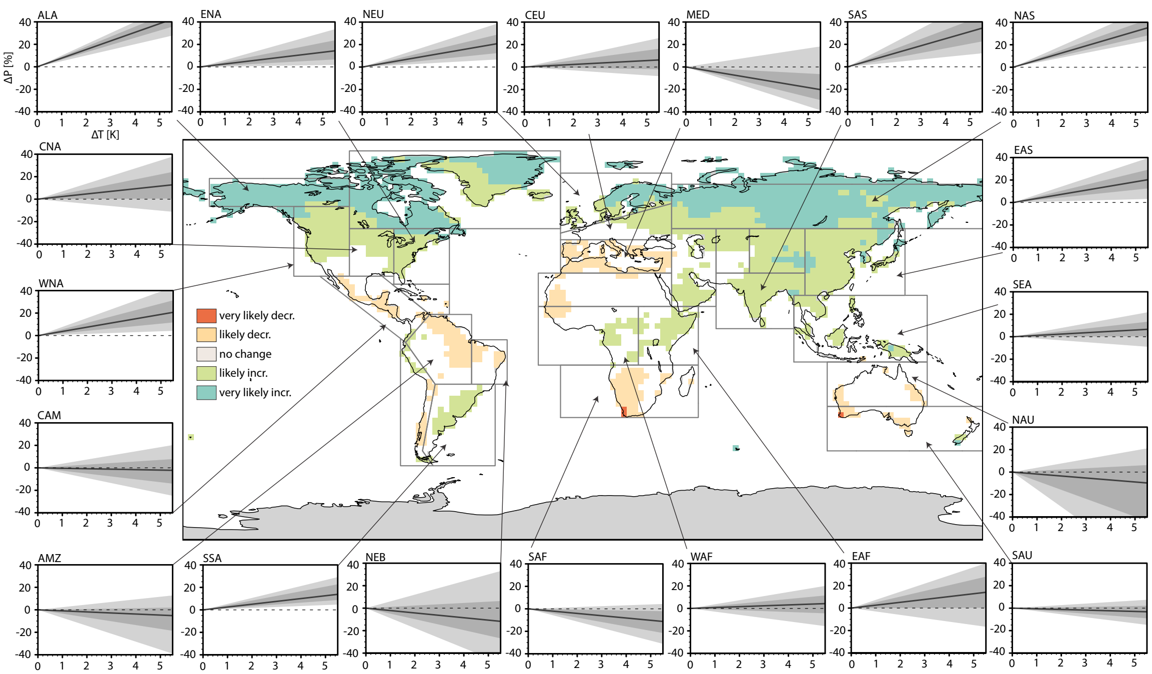

Figure 3. Conceptual summary of the probability that the slope of P is negatively/positively different from zero considering all climate models and all scenarios. Panel plots illustrates the uncertainty distribution of the sensitivity of P to global temperature change as a function of global mean temperature change averaged for each SREX regions outlined in the map (the shading in each panel plot corresponds to those illustrated in Fig. 1).

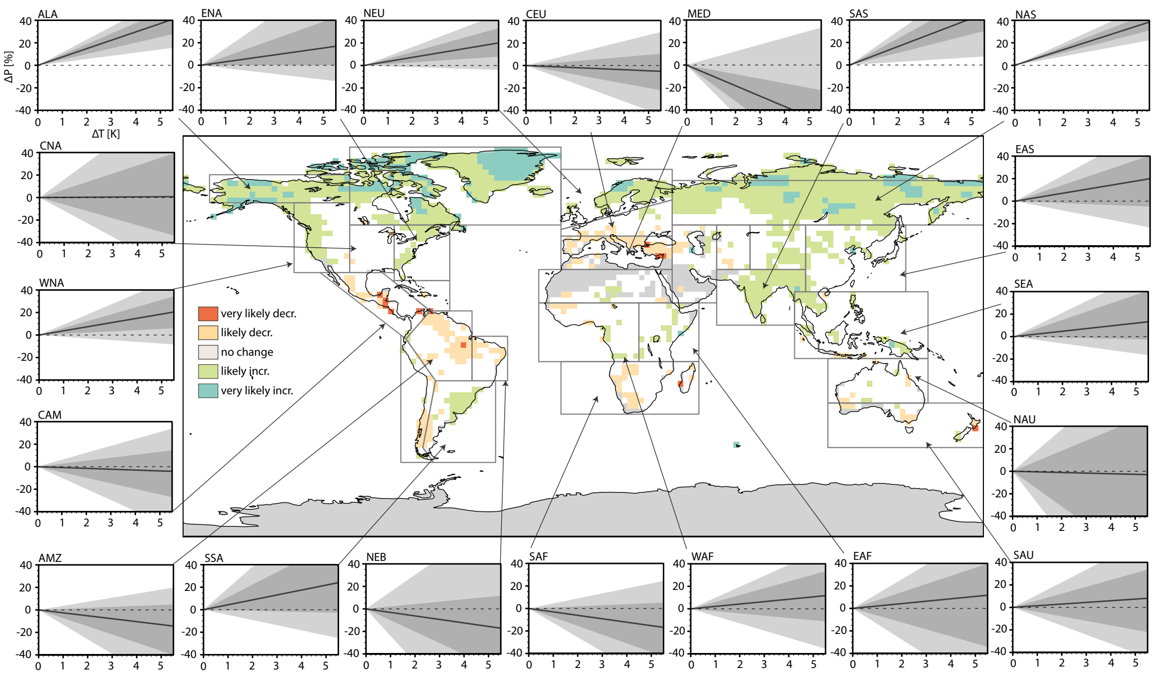

Figure 4. Conceptual summary of the probability that the slope of $P-E$ is negatively/positively different from zero considering all climate models and all scenarios. Panel plots show the uncertainty distribution of the sensitivity of $P-E$ to global temperature change as a function of global mean temperature change averaged for each SREX regions outlined in the map (the shading in each panel plot corresponds to those illustrated in Fig. 1).

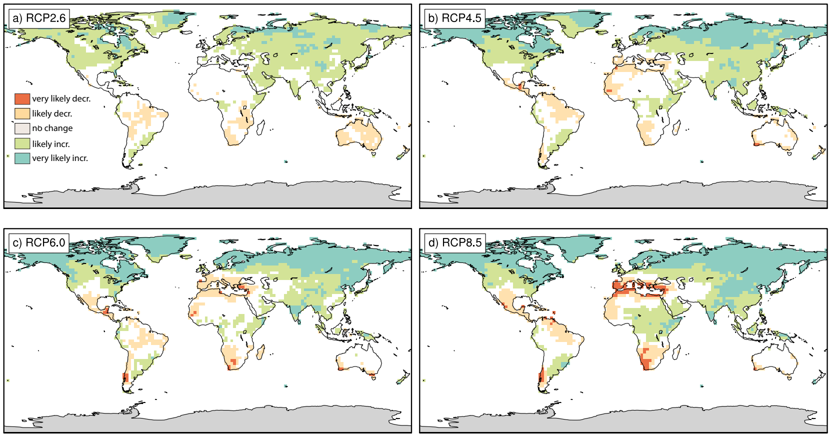

Figure 5. Conceptual summary of the probability that the slope of P is negatively/positively different from zero considering all climate models and a) the RCP2.6, b) the RCP4.5, c) RCP6.0 and d) RCP8.5 emission scenario only. See Fig. 3 for comparison.

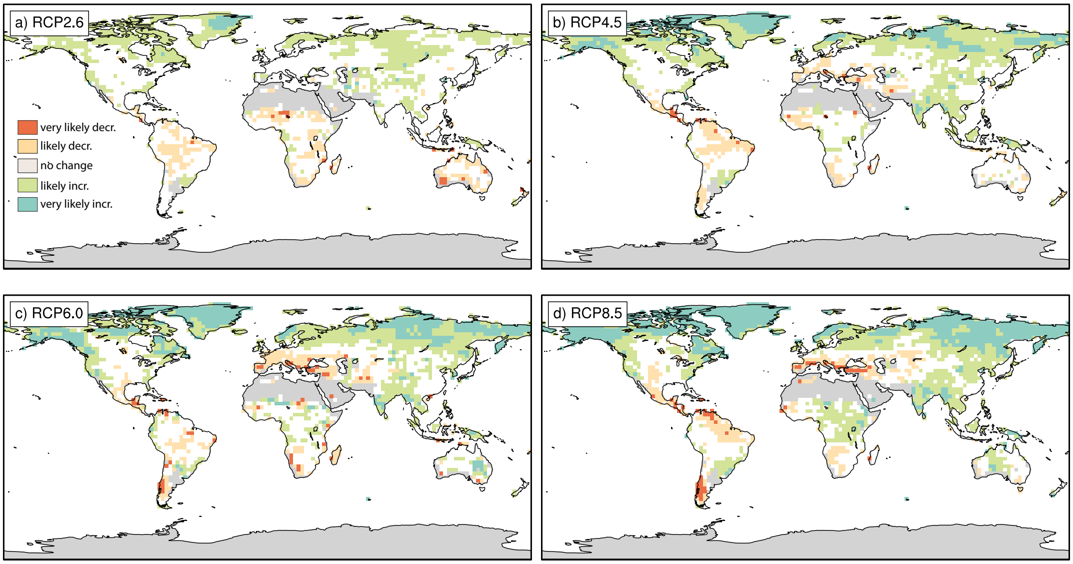

Figure 6. Conceptual summary of the probability that the slope of $P - E$ is negatively/positively different from zero considering all climate models and a) the RCP2.6, b) the RCP4.5, c) RCP6.0 and d) RCP8.5 emission scenario only. See Fig. 4 for comparison.

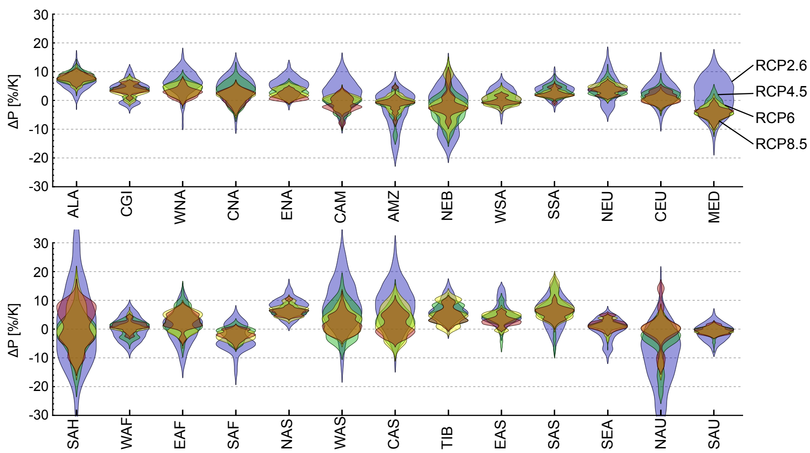

Figure 7. Uncertainty distributions (shown as violin plots) of the sensitivity of P to global mean temperature change for each emission scenario averaged over all SREX regions (as outlined in Table 2 and Fig. 3).

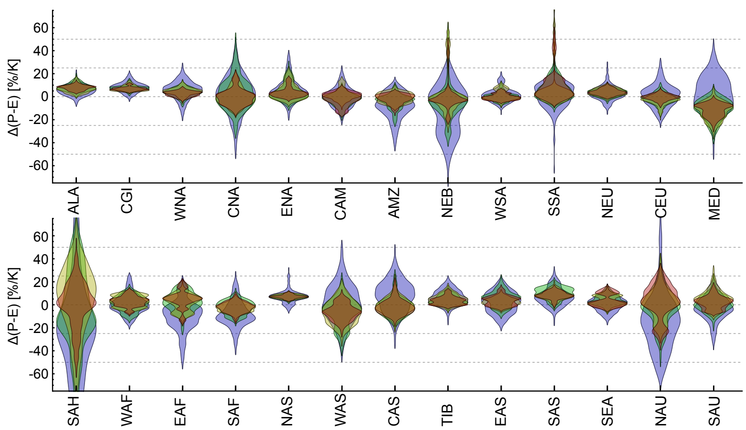

Figure 8. Uncertainty distributions (shown as violin plots) of the sensitivity of $P - E$ to global mean temperature change for each emission scenario averaged over all SREX regions (as outlined in Table 2 and Fig. 4). It is important to note that the data considered to estimate the area average is scarce in some regions (e.g. SAH). Please also refer to Fig. 7 for more information.

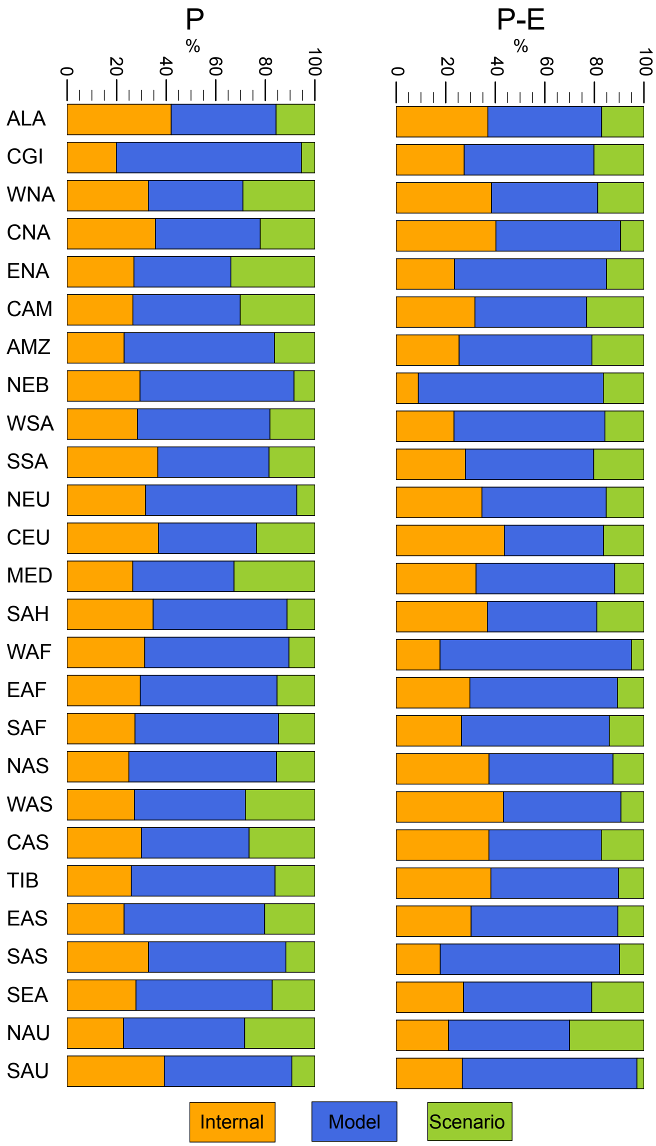

Figure 9. Sources of uncertainty in the sensitivity of P (left) and $P-E$ (right) to global mean temperature change averaged over each SREX regions as outlined in Fig. 3.

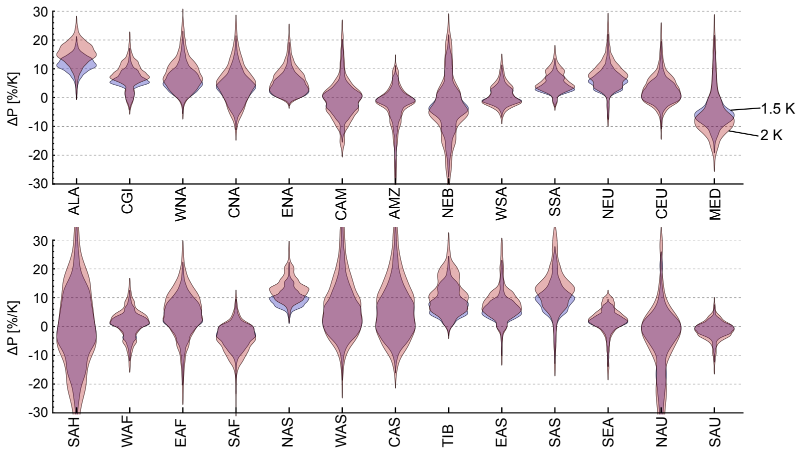

Figure 10. Uncertainty distributions (shown as violin plots) of the sensitivity of P to global mean temperature change for two different degrees of global mean temperature change, which correspond to the widely used warming degree limits of $1.5K$ and $2K$. The estimates are averaged over all SREX regions (as outlined in Fig. 3 and Table 2).

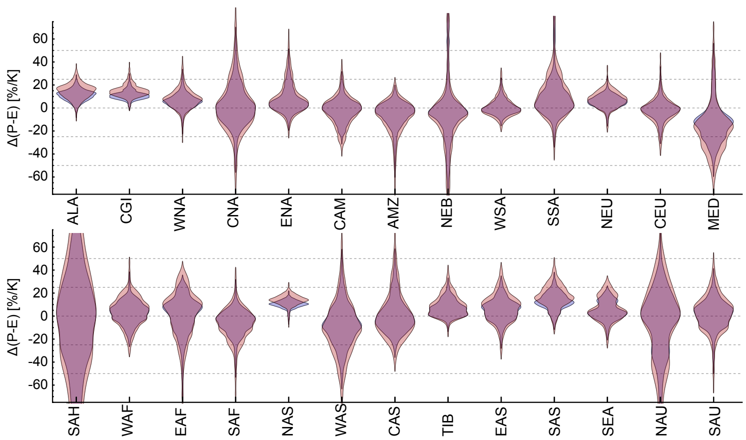

Figure 11. Uncertainty distributions (shown as violin plots) of the sensitivity of $P - E$ to global mean temperature change for two different degrees of global mean temperature change, which correspond to the widely used warming degree limits of $1.5K$ and $2K$. The estimates are averaged over all SREX regions (as outlined in Fig. 3 and Table 2). It is important to note that the data considered to estimate the area average is scarce in some regions (e.g. SAH). Please also refer to Fig. 10 for more information.

Table 1. List of models and availability under each emission scenario.

Model	RCP2.6	RCP4.5	RCP6.0	RCP8.5
ACCESS1-3		x		x
bcc-csm1-1	x	x	x	x
CanESM2	x	x		x
CESM1-BGC		x		x
CMCC-CMS		x		x
CNRM-CM5	x	x		x
CSIRO-Mk3-6-0	x	x	x	x
FGOALS-g2	x	x		x
GISS-E2-R	x	x	x	x
HadGEM2-ES	x	x		x
IPSL-CM5A-MR	x	x	x	x
MIROC5	x	x	x	x
MRI-GCM3	x	x	x	x
NorESM1-M	x	x	x	x

Table 2. List of Acronyms for all 26 SREX-regions (Seneviratne et al., 2012)

Region	SREX-Acronym
Alaska/NW Canada	ALA
Eastern Canada/Greenland/Iceland	CGI
Western North America	WNA
Central North America	CNA
Eastern North America	ENA
Central America/Mexico	CAM
Amazon	AMZ
NE Brazil	NEB
West Coast South America	WSA
South- Eastern South America	SSA
Northern Europe	NEU
Central Europe	CEU
Southern Europe/the Mediterranean	MED
Sahara	SAH
Western Africa	WAF
Eastern Africa	EAF
Southern Africa	SAF
Northern Asia	NAS
Western Asia	WAS
Central Asia	CAS
Tibetan Plateau	TIB
Eastern Asia	EAS
Southern Asia	SAS
Southeast Asia	SEA
Northern Australia	NAU
Southern Australia/New Zealand	SAU