

## **REPLY TO REVIEWER #3 for the revised manuscript**

*We want to thank the reviewer for her/his valuable comments and suggestions. Suggestions that were directly adapted are marked in green, while additional replies from the authors are marked in blue.*

This paper compares projections of seasonal temperature, precipitation, heatwaves and droughts over Europe on three SMILEs based on regional climate models. The paper has some really interesting results, such as the general agreement of the sign of the change, although not the magnitude across the three SMILEs used. The paper provides a significant contribution to the literature as possibly the first intercomparison of regional climate model SMILEs. The scientific results are generally sound and well founded, however the paper itself often lacks clarity and precision, making the job of the reader difficult at times. I recommend revisions to the text in the manuscript before publication.

This is my first review of the manuscript. I note that the following issue pointed out by the original reviewers is not completely resolved.

1. There are still some instances where the term internal variability is confusing and it is unclear which metric is being talked about and whether you refer to inter-annual variability or general variability. *We went through the text and clarified all occurrences of “internal variability” and “variability”*

Examples are:  
lines 224/234

Comments on the manuscript:

Section 3.2 – can you do a statistical test on the means and spread to be able to say for certain whether the different models are truly different? *We added some information to section 3.2 concerning the statistical tests for mean (T-test) and variance (Brown-Forsythe Test)*

91-110 – should this explanation be in the introduction or methods? *This is a valid comment as it would also fit into the methods section. We chose to leave it in the introduction, since it sets the frame for all coming sections and there is no proper place for these paragraphs in the methods section*

164/174 – is this really all models? Or all ensemble members? This is an important distinction *The figures clearly show that it is both*

184 why do we expect this to be normally distributed? *Because the “perfect” model would totally capture the IAV as IMV, and IAV of precipitation should be normally distributed in Europe*

210 – you say ‘looks like’ can you test this? *It would be possible to test the observational distribution against each member individually, using a KS test for example. If all individual tests would show no similarity, one could state that it is rather unlikely that the observations could be part of the ensemble of distributions from the respective model. It would, however, still only allow us to make a probabilistic statement – we cannot say for sure if the model is able to represent observational IAV satisfactorily (see discussion on outlier members). The general problem that we only have one realization of historical climate (variability) remains, why we chose to leave this a qualitative analysis.*

In general I am confused about how you calculate some of the variability metrics, in some places you refer to standard deviations and other places variance. Can you clarify this?

*Lines 91, 234,235,239,246 are places where I was unsure what metric you were using IAV is calculated as standard deviations, while the BF Test checks for equal variances. Since std is the square root of var, and they are therefore strongly tied, we use both metrics. We went through the text and made changes where necessary.*

253 – can you be more specific about the ‘period around this year’ statement? [We added some information](#)

Conclusions: These are mostly context and future work. Can you make some stronger conclusions from this manuscript. What did you find? Why is it important? [We do not agree. We clearly state what we found \(e.g. new lines 458-463\) and say why this is important \(454-457, 463-464, last two paragraphs\)](#)

The following set of textual suggestions are meant as just suggestions. These are some examples of where the choice of wording is confusing or the grammar is incorrect. This is not an extensive list and I strongly suggest the authors revisit the text for clarity and to make sure all sentences are grammatically correct and make sense.

Textual suggestions (with line number first):

Abstract

10 Define IAV in its first instance

11 get rid of ‘often’

12 might be better to say something like: ‘represent a combination of external forcing and IAV’

14 powerful ‘tools’

14 ‘many’ instead of ‘multiple’

16 ‘summer and winter’

18 what do you mean by the two extreme indicators? How are they biased? [The two extreme indicators are the number of heatwaves and the maximum length of dry periods – in contrast to seasonal tas and pr we think it is clear that these can be considered extreme indicators. The biases do not really show a systematic behavior, why we did not specify this in the abstract. A new wording tries to make it clearer what we found without making this information too long for the abstract, as it is not one of the main findings of the paper.](#)

24 what do you mean ‘mostly follow’? [we changed the wording to ‘also show’](#)

19 ‘some of the individual realisations show’ [the current wording better emphasizes the differences between realizations and is therefore not changed](#)

21 ‘the’ significance

23 dont need both ‘further’ and ‘also’

Introduction

28 remove the words ‘Next to’ replace with ‘In addition to’ and replace ‘another’ with ‘an’

29 ‘on’ different timescales

30 replace ‘like’ with ‘such as’

33 ‘utmost’

36 replace ‘especially’ with ‘particularly’

38 do you need tas/pr in the intro, I would move this information to the methods [We want to keep it here where all indicators are listed in the main text the first time](#)

Data

125 ‘between’ the three ensembles

130 ‘with’ different

131 ‘each initialisation using atmospheric perturbations’

134 replace inter-compared with ‘compared’

136 ‘because of its availability for Europe and because it has similar spatial resolution’

Section 3.1

142 Your first words are ‘The indicators’ - the reader does not remember what this refers to [We added the term indicators in the \(new\) lines 39/40 and refer to Table 1 at the beginning of section 3.1](#)

Section 3.2

174 replace ‘show’ with precipitation increase in all models – models don’t show a result they do something

175 similar standard deviation and range of what? [Of changes in pr and tas; we added this to the text](#)

### Section 3.3

180 sentence is unclear [we changed it to clarify it](#)

186 should read 'is concentrated inside'

192 – stable conditions is unclear [changed to 'no clear change'](#)

193 'this' approach

205 'and' maximum duration

206 – weird sentence – 'even better' is an odd choice of words [we changed the wording](#)

208 comma before the word but

209 'to represent' IAV

210 'to represent'

211 'are not expected'

217 do you mean 'for' the maximum duration?

221 'which is probably the most difficult'

223 what is a 'outlier member'? [we added a short description](#)

227 'with' not after

### Section 3.4

difference 'in'

350 'unstable'

357 'only possible'

### 5 Discussion

Why is the discussion full of really short paragraphs? This reads as a list and could be made flow better and clearer.

[We discuss several aspects of our work that are not necessarily connected, but need to be addressed to set our results into context. A 'flow' can be nice to have, but we do not think it is necessary. We find the discussion clear and good to read.](#)

361 – weird sentence – I would say the number of SMILEs available [we changed the sentence](#)

369 – by biases do you mean different mean states? [Yes, we added a 'mean'](#)

379 'Better agreement' - remove the A

392 – what do you mean by 'in accordance' ? [the variability increases in accordance to the increases in the mean; a higher variability can be expected when the mean increases, just because of higher numbers of the respective variable. We changed the wording to 'in conjunction with'](#)

404-406 and 416-419 are difficult to understand [We added some information to make it clearer](#)

410 might be? Or is? [Well it is not sure, because we did not specifically test this, but we can also not exclude these uncertainties. We replaced 'might' by 'can'](#)

423 what does this tell us?

236 'challenges'

### 6 Conclusions

430 This sentence is hard to read [We do not agree; the sentence is two lines long and has a clear meaning and structure](#)

434/449 This was true, however there a now quite a few studies using multiple SMILEs [we removed the first sentence \(434\) and changed the wording in the second sentence to make clear that there is an increasing number of publications comparing SMILEs, although most publications about SMILEs indeed just use one SMILE](#)

442 remove (dis)

Figure captions:

F4 'shading' and replace 'via' with 'using'

F5 ' during the period'

F5 – mention that individual lines are individual ensemble members I think it is clear enough. The caption says ‘and each ensemble member’

F7 – tell us that the reference period is 1980-2009

F8 – is this variance calculated as square of the std? Yes, this is the definition of variance and standard deviation

F8 how do you define a significant difference? We added the information that this is tested with a Brown-Forsythe Test

F10 how do you pool IAV? The pooled IAV is just a combination of residuals from all members, e.g. 50x30years for CRCM