

## ***Interactive comment on “An observation-constrained multi-physics RCM ensemble for simulating European mega-heatwaves” by A. I. Stegehuis et al.***

**A. I. Stegehuis et al.**

annemiek.stegehuis@lsce.ipsl.fr

Received and published: 15 April 2015

Answers reviewer 2:

\* General comments:

This work analyses an unprecedented (to my knowledge) multi-physics ensemble consisting of 216 summer seasonal simulations, focusing on heat waves over France (2003) and Russia (2010). It provides a fully systematic approach towards the selection of an optimal sub-ensemble to represent heat waves. The paper presents fairly novel concepts and ideas and a new dataset which will probably feed subsequent work. Therefore I suggest the paper to be published after a minor revision taking care of the

C3645

specific comments below, which mainly refer to further discuss some points and solve some doubts to improve the reproductibility of the results.

\* Specific comments:

1) The authors intend to create an optimized WRF ensemble for heat waves (7873:910). What would be the use of such ensemble? The experimental setup used should be taken into account. The simulations shown were run for a few months nudged towards the observed flow. For climate change simulations, nested into a GCM, nudging the atmospheric flow could be a problem. Also, long-term simulations could build up biases not arising in a few months (e.g. related to soil moisture). For seasonal forecasting, the authors recognize problems (7863:15-19) to reproduce observed events due to the chaotic nature of the atmospheric circulation.

- The ensemble could be used for climate change modeling studies. All 5 members of the reduced ensemble differ in physics, which could serve as uncertainty measure.

2) The 5-member sub-ensemble was only tested for heatwaves. The 2007 "normal" season is not shown in Figure 3 or any of those in the supplementary material (even though it is stated that they are in the Suppl. material in 7871:2). If these members are the best in any physically meaningful sense, they should also perform well in the "France 2007" case study. Is that the case?

- Yes, they also perform well in 2007, although the spread of the whole ensemble (216 members) is smaller in 2007 than during the heatwave cases. We added some figures of 2007 in the supplementary material.

3) There are already examples in the literature of "sub-ensembles" breaking model democracy, in which the sub-ensemble outperforms the full ensemble. For instance, Herrera et al (2010) selected a sub-ensemble using mean precipitation and show that this sub-ensemble is also well fitted for extreme precipitation regimes. This result is close to the results found in this work (sub-ensembles selected for a heatwave work

C3646

well for other regions or regular seasons –if this is the case–), and could be added to the discussion, given that it extended the idea to multiple models.

- We added this study in the discussion section.

4) The potential implications of the study for climate modeling (7873:17-) need to be discussed in a wider framework. The authors constrained the ensemble to a particular season, variables and error metrics. In this way, they were able to find an "optimized" set of configurations. However, It has long been recognized (Fernandez et al, 2007), that in a climate simulation an optimal configuration cannot be chosen. Biases and the best-performing configurations heavily depend on the season (Garcia-Diez et al, 2013), variable and even on the metric used (Jerez et al, 2013).

- Yes this is true. We added this in the discussion section. Primary results from a next study suggest however that the small ensemble also performs relatively well in other seasons.

5) Moreover, observational uncertainty was not considered. It has been shown that the reference observations affect model rankings (Gomez-Navarro et al, 2012). This needs to be discussed at some point in the paper.

- We added this in the methodological section, where we discuss the ranking method.

6) Jerez et al (2013) did not use WRF (7865:15). Other potential references here are Awan et al (2011) and Mooney et al (2012). Also, multi-physics ensembles did not start with WRF. There are a few other works with its predecessor, MM5.

- We changed the reference of Jerez in Awan and Mooney, and added that earlier studies have been done with MM5.

7) Weisheimer et al (2011) did not use WRF (7870:15). Remove or rephrase.

- We removed the reference.

8) The authors find probable (7872:18) that the inclusion of another land surface model

C3647

would increase the ensemble spread. This statement can be accompanied by a cite to Mooney et al (2012), where they show strong differences when changing the LSM (see their Fig. 2b, e.g. Sim 9 vs. 11). It is not clear at all why the LSM sensitivity was left out of the study. There is plenty of literature (even cited by the authors) highlighting the role of soil-atmosphere interactions in the development of heat waves and the authors themselves recognize it (7869:16).

- This paper is designed as a methodological paper. Our intention was to test as many physic schemes as possible, including land-surface schemes. However, at the moment we performed the study 4 different land-surface schemes were available: NOAH, RUC, Pleim-Xiu and the 5-layer thermal diffusion scheme. The last one, non-physical, is designed for test cases and cannot be used in realistic situations. The Pleim-Xiu scheme comes with a dedicated set of other physical schemes and does not allow in most cases to combine different possibilities. We performed an ensemble of simulations using the RUC scheme, but we found that it provided extremely sensitive fluxes, with large latent heat fluxes at the beginning of season and extreme subsequent drying in summer months. Sensible heat fluxes also appear overestimated. Comparisons with several FLUXNET sites are now explicitly shown in Figure 1. So even if temperatures of heat waves would match observations in a few combinations of physics schemes, we would almost be sure that this would be for wrong reasons. We also experienced technical problems while running several of the RUC simulations. This made us decide to only use one land-surface scheme and focus on the atmospheric physics processes. We believe that this can be very useful to the many users of the WRF model to examine this sensitivity and to be aware of best-performing physics combinations using NOAH land surface as it is very widely used. We added some sentences to the methodological section, and added a figure with the comparison RUC vs NOAH (new Figure 1).

9) I don't agree with the sentence (7870:12) "By contrast, heatwave temperatures do not seem very sensitive to the planetary boundary layer and surface layer physics schemes". Figure 2d seems noisier than the rest because there are more PBL options

C3648

tested. However, there is a clear, systematic temperature dependence on the PBL. If you imagine a regression line for each PBL scheme, all of them preserve the relationship (slope) with soil moisture, but the the heat wave average temperature is clearly different.

- We adapted the text a little bit to strengthen the point mentioned by the reviewer. The sentence mentioning this point is now: 'Heat wave temperatures seem to be least sensitive to the planetary boundary layer and surface layer physics schemes'.

10) The caption of Figure 2 says correlation where scatterplot is meant. These particular plots show that there is indeed (negative) correlation, but the plots are scatterplots.

- Correlation is changed into scatter plot.

11) X-axis labels in Figs 1 and 3 read "Time (DOY)". I assume it means Day Of the Year but, please, define. Also, in the panels with this axis, two vertical lines showing the heatwave period considered would help, given that different periods were chosen for each event. Also, if Fig 1bc shared the Y-axis with Fig 1a, they could be directly compared with each other. Currently, the normal year seems as hot as the 2003 heat wave, when in fact it is 5K colder.

- DOY is now defined in figure 1a, and we changed the y-axes of figure 1 to be directly comparable to each other.

12) In Fig 3c-August, the cyan circle is missing (probably hidden behind other member). Using non-overlapping symbols would help. The same happens with the pink circle in Fig. 3B

- Different symbols are now used so there is no longer full overlapping.

13) The resolution is stated to be 50km (approx 0.44deg). Was a Lambert grid projection in Kms used? or the Euro-Cordex standard 0.44 rotated lat-lon grid? Please, clarify.

C3649

- We used the Euro-Cordex standart 0.44 rotated lat-lon. This is now added to the methodological section.

14) Observational data is not fully described. Which E-OBS version was used? Was any interpolation carried out in the analyses?

- This is now clarified in the methodology section.

15) The pre-screening of the simulations considering only those within 1K of the E-OBS temperature might be problematic. RCMs have biases. With the method proposed, a fairly physically-consistent simulation could be disregarded, while a simulation unrealistically compensating temperature biases might get in. The latter can easily happen (Garcia-Diez et al. 2014).

- It is true that the 1K temperature bias is rather arbitrary. But because we are really interested in heat waves, and especially the high temperatures, we chose to have this limit anyway. We agree on the fact that within the 55 simulations within the 1K limit, their might be configurations that compensate temperature biases. However, because the simulations are also tested on their ability to simulate well the precipitation and the radiation, we do not expect that this is the case for our 5 (or even 10-15) best simulations.

16) The ranking metrics are not fully clear to me. Daily temperature differences are used (7867:26). But, which score was built out of them? RMSE? Why was temperature considered at a daily scale and precipitation at monthly scale?

- For temperature we used the bias, not the RMSE. Because modeled daily precipitation is much noisier than daily temperature, we decided to used monthly values for this variable. Furthermore radiation data was only available on a monthly timescale.

17) For radiation data, was the model interpolated to the stations to compute the spatial averages? which interpolation method was used? How many radiation stations were available in each region?

C3650

- Yes, the model data was interpolated to the station and spatial averages were computed. We used 'nearest neighbor' for the interpolation. We did only consider France for the radiation, because over Russia the observation data was too scarce. For France we used 21 stations for 2003 and 20 for 2007. This information is added in the methodology section.

18) Regarding the rejection of the members differing in only one scheme (7868:08): How many of these members were disregarded to get the top-5? What is the interest of "keep[ing] a large range of different realistic physics combinations between the simulations" (7868:11)? I see also an interest in the single-step ensemble members. In these ones, the differences can be traced to the single scheme that changed. For instance, "The two simulations that are largely overestimating latent heat flux" (7871:21) are those not using Tiedtke, but this could be just by chance, given that they also differ in other schemes and the schemes interact in a non-linear way (Awan et al. 2011).

- Finally the 'top 5' members already differed with two schemes from each other, so this rejection was not longer necessary. Because we are looking for a variability of physics, we thought about using this rejection. We removed the sentences from the text.

19) How were the extreme configurations selected (7868:22)?

- The two configurations are simply chosen to show the consistency of 'warm' and 'cold' simulations. They are chosen based on daily mean temperature over France during the 2003 heat wave. They are not separate sets for different regions or years. We added a sentence in the text to explain this better.

20) By "the middle of the simulations" (7869:22), I guess you mean the "median".

- Yes, this has been changed.

21) At some point (7870:18), the effect of convective clouds on radiation is invoked. However, note that in WRF the interaction of radiation with sub-grid clouds has only recently been implemented (Alapaty et al. 2012) and included in WRF3.6 for certain

C3651

combination of radiation and cumulus schemes. It was not included in the version used in this work (WRF 3.3.1).

- We agree with the reviewer.

22) The discussion in 7871:18-25 seems to imply (although it is not explicitly stated) that the good performance of the Tiedtke scheme just during the heatwave is just by chance.

- We did not mean to imply this, as we found that the Tiedtke scheme is performing quite well overall. However, we cannot state that the other convection schemes do not simulate the latent heat flux very well, as this is not the case for other years. But we found that also in some other cases, the two of the five best configurations not using Tiedtke, are performing a little bit less well, for example precipitation over the Iberian Peninsula in 2003 (Suppl. Fig. 5c).

23) "We found a large spread" (7872:11) I would highlight, just at the beginning of this sentence "Even though the simulations were constrained by grid nudging,"

- This is now added to the text. Thank you for the suggestion.

24) The journal recommendations suggest that "The model name and number should be included in [the title of] papers that deal with only one model". Replace RCM by WRF in the title.

- We propose to change the title into : 'An observation-constrained multi-physics WRF ensemble for simulating European mega-heatwaves.

(beware I'm not a native speaker)

7862:17, "together with varied physics scheme." sounds odd to me. Please, rephrase.

This has been rephrased.

7866:04, "temperatures differ by less among one another than 0.5C" sounds odd.

C3652

Rephrased to: 'temperatures differ by less than 0.5C among one another'.

7867:01, "Tawary" should read "Tewary".

Corrected.

7869:26, "better" -> "well" (or "better than [what?]")

We rephrased: 'better than the coldest simulation'.

7872:27, missing period "heatwaves Changes"

Corrected.

\* References: Alapaty et al. (2012) "Introducing subgrid-scale cloud feedbacks to radiation for regional meteorological and climate modeling" *Geophys. Res. Lett.* 39:L24808  
Fernández et al (2007) "Sensitivity of the MM5 mesoscale model to physical parameterizations for regional climate studies: annual cycle" *J. Geophys. Res.* 112:D04101  
Gomez-Navarro et al (2012) "What is the role of the observational dataset in the evaluation and scoring of climate models?" *Geophys. Res. Lett.* 39:L24701  
Herrera et al. (2010) "Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel" *J. Geophys. Res.* 115:D21117  
Mooney et al (2012) "Evaluation of the Sensitivity of the Weather Research and Forecasting Model to Parameterization Schemes for Regional Climates of Europe over the Period 1990–95", *J. Clim.* 26:1002-1017

---

Interactive comment on *Geosci. Model Dev. Discuss.*, 7, 7861, 2014.