

MS No.: gmd-2014-175
Review, 22 December 2014

Title: An observation-constrained multi-physics RCM ensemble for simulating European mega-heatwaves

Authors: A.I. Stegehuis, R. Vautard, P. Ciais, A.J. Teuling, D.G. Miralles, and M. Wild

Recommendation: [Major Revision]

GENERAL COMMENTS:

This article presents and discusses the performance of a large multi-physics ensemble configured with the WRF regional climate model system in representing two major heat wave episodes observed in recent years across the European continent.

The paper attempts to objectively assess which configurations perform best in reproducing the heat wave characteristics with the purpose to identify a few best performing configurations which might be recommendable for application in studies on the role of heat wave in Europe under climate change conditions in summer.

The paper is reasonably well written, although I came across quite a few sentences which I found difficult to understand. In addition the manuscript also contains a number of assumptions or claims that go without solid argumentation (or argumentation at all) and are not backed up by references. Some of the figures are difficult to view, in particular the Figure 1 and the figures displaying scatter plots. But with some work that can easily be improved.

A major concern, however, I have with the followed methodology is that the authors have chosen to leave out the land surface scheme from their considerations and to restrict their construction of a multi-physics ensemble to what they refer to as the *atmospheric* physical schemes. There are probably quite a few large-scale weather phenomena that are rather insensitive to the details of a land surface scheme, but for sure such approach does not hold for heat wave conditions across Europe. There is an overwhelmingly large amount of literature that points to the role of land surface processes and their effect on land-atmosphere exchanges in the weeks or months ahead of the onset of a heat wave episode across (a subregion in) Europe, so I really can't understand why the authors have chosen to pursue this approach. The more so as the authors themselves write in their concluding remarks." ... a limitation of this study is the use of only one land-surface scheme; the five selected WRF configurations may actually all compensate for systematic errors of the NOAH land surface scheme ..." etcetera.

In contrast to the use of a single land surface scheme there is a multitude of atmospheric physics schemes examined in this study that can be exchanged for one another, in particular there are six different boundary-layer/surface-layer schemes, but it is not at all made clear to the reader in what aspect they differ or how different they are. To a lesser extent similar considerations apply to the other physics schemes as well. Too many,

unclear what their differences are, and in a way accidentally selected because these schemes happen to be implemented in the WRF system, which doesn't help in making this a "clean" ensemble, meaning that there is no way in which the various members of the ensemble can be neatly discerned from each other in some model physics phase space. In that respect, this ensemble is not so very different from the multi-physics ensemble approaches mentioned in the paper.

To conclude, in my opinion ignoring variations in the type of land surface scheme in building up a multi-physics ensemble makes the approach that is followed quite out of balance, in particular when such ensemble is meant to draw conclusions for the model ability to represent heat wave episodes. I would argue that the authors should first carry out the future study they announce in their "Concluding remarks" in which they intend to investigate the performance of a joint permutation of different atmospheric physics schemes and land surface schemes. On the basis of the results from that study they could then have solidly zoomed in on what they want to present in this study.

MAJOR COMMENTS:

- 1.) The title of the paper should already contain a reference to WRF to directly inform the reader what the paper is about. WRF is not synonymous nor equivalent to RCM, thus conclusions drawn for WRF can probably not immediately be generalized to an arbitrary RCM.
- 2.) Following my general comment on the omission of having varied the land surface scheme which makes the role of the NOAH LSM that is being used even more relevant I want to bring up some issues on the initialization of soil moisture. The authors write that soil moisture is directly taken from ERA-Interim. The authors also state that varying the initial date (1st of May against before 1st of May) results in variations in temperature outcomes of less than 0.5 C, at least for the August 2003 case. I would argue that this result does not at all imply that soil moisture is adequately initialized, it only indicates that there is little sensitivity regarding the date of initialization. My concerns are the following.
 - ERA-Interim employs TESSEL as the land surface scheme which is rather different from the NOAH LSM. How was the actual mapping of TESSEL soil moisture onto NOAH soil moisture be carried out. Through interpolation of soil moisture (in relative volumetric units) or through interpolation of a soil moisture index taking account of the soil moisture field capacity and wilting point parameters in the respective schemes? The 2nd approach is obviously preferable. Please mention in the text what approach is followed in this respect.
 - But even if they did, there are issues of soil water buffering capacity. Soil depths in both schemes may be difficult so, while properly mapping soil moisture content from TESSEL to NOAH, the resulting soil moisture columns might still be quite different. The authors should mention this point.
 - Also, soil moisture must still be predominantly regarded as a model (or module) specific quantity. It is poorly constrained by observations and in the context of data-assimilation often treated as a free parameter that can be used (or abused) to

reproduce observed near-surface parameters like temperature and relative humidity.

So, on the basis of these grounds I would argue that there is considerable amount of uncertainty that can be attributed to the initial soil moisture profile used as the starting point for the set of WRF-simulations. I think the authors could benefit from this situation. Instead of taking a second and, even, third, land surface scheme, (still the most recommendable option, though) or altering the formulation of the land surface scheme, they might bring in the potential role of soil moisture on the evolution of the seasonal slice simulations by perturbing the initial soil moisture profile. For example, an option is to use plausible dry and wet perturbations of soil moisture initial profiles to examine the sustainability of the 5 or 55 best performing configurations that came out of their current exercise. (I don't think it is necessary to redo all 216 configurations that have been done so far).

- 3.) The Russia-region is quite close to the eastern boundary of the EURO-CORDEX domain. This is potentially problematic in simulations of anomalous circulations like those that give rise to heat wave episodes, where there might be a conflict between the circulation sustained by the regional model and the forcing imposed at the near-by eastern boundary. Please, make clear in the text how you paid attention to this aspect. Did you choose a model domain that is actually much larger than the EURO-CORDEX domain? Did you broaden the boundary relaxation zone? Anything else?

MINOR COMMENTS: (in indicating the line number I refer to the line number value in the pdf-document of the discussion paper as outlined by GMD).

- 1.) It seems to me that the proper spelling of "heatwave" must be "heat wave"; similarly "mega-heatwave" should be spelled as "mega heat wave" (see Wiktionary)
- 2.) Page 7862, line 2. "Climate models are often not evaluated ..." This is a too strong statement. Take e.g. Vautard et al. and Kotlarski et al. who evaluated models contributing to EURO-CORDEX.
- 3.) Page 7862, line 6: "sensitive" Sensitive to what?
- 4.) Page 7862, lines 7-9: these 55 combinations that can reproduce" can so because they satisfy an pre-imposed criterion. Please mention that the criterion is less than 1 degree bias during the heat wave episodes.
- 5.) Page 7862, line 11: The statement "...and short wave radiation is slightly underestimated" seems to contradict the results discussed in the paper which clearly show that the model simulations in the mean overestimate the observed radiation, and that this hold for the majority of the five best-performing configurations. Please, restate.
- 6.) Page 7863, lines 13-14: according to Weisheimer et al., 2011, the enhanced sophistication *combined* in land surface hydrology, convection and radiation *proved key* (their words) for a successful reforecast of the 2003 summer in Europe.
- 7.) Page 7863, line 15: What is meant with "easily"? Please explain.

- 8.) Page 7863, lines 15-19: Sentence starting with “However ...” This a a very long and difficult sentence. Please, cut into pieces and clarify. E.g. what is meant with “ the effect of the representation of physical processes.”
- 9.) Page 7863, line 23: “hindcast simulations” → “evaluation experiments” in CORDEX-compliant terminology.
- 10.) Page 7863, line 26-27: Regarding the line “For some models ...for the 21st century”, I am wondering where it comes from. It is not a conclusion taken from Vautard et al. 2013, because that paper was only about ERA-Interim forced RCM-simulations. I am also wondering what the authors intend to say with this line. Because I do not immediately see how the size in bias can be connected with the projected temperature change in the coming century, be, please, more explicit in what you want to say.
- 11.) Page 7863, line 27- page7864, line 3. “Individual ... internal variability”. Again this is very long sentence, and I do not quite understand what you mean to say with the second part starting with “ because ...” Please, clarify.
- 12.) Page 7864, line 28: please change “using the same model” into the “using the same model system”, because these groups are not using the precise same model.
- 13.) Page 7864, line 29: What is meant with “democracy-driven”. Please clarify.
- 14.) Page 7868, lines 8-12: This a again a very long sentence. Please, break up in parts.
- 15.) Page 7868, line 22: It is not entirely clear to me what criterion was used by the authors to determine the extreme configuration. Is it only based on “daily mean temperature”? Throughout the paper there is only one set of two extreme configurations (am I right?), which is used in Fig 1a-h and Fig S2. Or are there separate sets for France and Russia? I think it would be very helpful if you explicitly state how these extreme combinations are configured. (I might have missed it, but I couldn't find it spelled out).
- 16.) Page 7868, lines 23-24. The “large bias” mentioned in these lines is certainly not always large, specifically not for the extreme configuration on the warm side. Please mention.
- 17.) Page 7869, lines 9-12: “The two selected extreme ... misrepresentation of the land water supply” What is meant with “land water supply”? Soil moisture content or evaporation/evapotranspiration from the land surface to the atmosphere? I find the argument presented in this line indeed quite suggestive.
- 18.) Page 7869, line 12: What is meant with “soil moisture”, and also in Figs 2 and S3? Soil moisture of the top soil layer (how thick) or averaged over the whole soil column?
- 19.) Page 7869, lines 17-19: this sentence “This indicates ...summer precipitation” precisely underscores why there should have been at least two different land surface schemes included in this study.
- 20.) Page 7869, lines 20-21: “For solar radiation ... approximately 100W/m2 ...” Difference in solar radiation over France and Russia, or differences over one region within an physics-ensemble. Also, solar radiation over Russia is not shown in Fig 1g or 1h.
- 21.) Page 7869, lines 21-28: Apparently there is a discernible overestimation of solar radiation in the warmest extreme configuration which is not translated in an

overestimation of near-surface temperature. So accordingly the authors suspect there is a cooling mechanism without mentioning what that mechanism would be. This is the interesting part. Is it compensated by an overly large reflected solar radiation (unlikely) or is it participated differently over sensible and latent heat flux than in nature, such that latent heat flux is overestimated. Yet, this is not giving rise to more precipitation (no large precipitation bias, see above), nor to more clouds (positive solar radiation bias), nor is it drying out the soil (because the excess latent heat flux continues, otherwise the partitioning of excess solar radiation would go into sensible heat flux giving rise to higher near-surface temperature.) Please, try to identify what this cooling mechanism could be.

- 22.) Page 7870, line 3: Please rephrase “In order to identify the most sensitive schemes for the development of heatwaves ...” as “In order to identify the parameterizations (or parametric schemes or physics schemes) to which the development of heat waves is most sensitive ...”. Schemes themselves are not sensitive! Check the remainder of your manuscript wrt the use of “sensitive”.
- 23.) Page 7871, lines 23-24: “The overestimation ...for other regions and years ...” I tend to disagree, I find the latent heat flux figure for Russia 2010 (Fig S5e) not very different from the result shown for France 2003 (Fig. 3d). I am wondering how the France-2007 time series for latent heat flux looks like in this respect. Is that comparison available?
- 24.) Page 7871, line 26: “cross-validation” → “cross-comparison” (also page 7873, line 7, and in first line of the caption of Table 3)
- 25.) Page 7872, line 7-11: The first sentence of the section “Concluding Remarks” is again a very long sentence. It is also not a a very adequate line. Why using the word “small” in front to set, you considered all available combinations in this context. Also the phrase “with a given accuracy thresholds for temperature, precipitation and shortwave radiation” is not clear to me. What kind of thresholds have been used for precipitation and short wave radiation?
- 26.) Page 7872, line 23-27: the conclusion might be that the performance of a configuration is related to its ability to adequately represent cloud parameters (cloud amount, liquid water, etc.) or cloud-radiative interactions. In that respect I am wondering which parameterization within WRF is representing the stratiform – or layer-clouds? Can you comment on that.
- 27.) Page 7873, line 4-6: replace “schemes” by “configurations” or “combinations” or “members”; “scheme” refers to a single parameterization, that is not what is meant here.
- 28.) Page 7873, lines 11-16: That points to the heart of the matter as I already mentioned under general comments.
- 29.) Page 7873, lines 17-26: Please mention explicitly that the conclusions from your investigation are only valid for heat wave conditions. There is no guarantee that the constrained ensemble is also better performing for e.g. wet summer conditions or winter conditions.
- 30.) There are two schemes in Table 1 assigned with number (6), namely WRF-SM6 and Tiedtke. Is that correct?

- 31.) I would strongly recommend to split Fig 1 into three Figures, because it is very difficult to read. Fig 1i becomes Fig 1, Figs 1a-c become Figs 2a-c, use column-format like Figs 2a,c,e. Figs 1d-h become Fig 3, also column-format is preferred.
- 32.) Figs 1d,e,f: Preferably use same y-axis range and start at 0.
- 33.) Figs 1g,h: Preferably use same y-axis range\
- 34.) Fig 2, but also Figs S3 and S4. It is quite hard to distinguish the points by their different colours. It would help to choose different plotting symbols as well.

OTHER POINTS:

- 1) Page 7862, line 7-8: “55 Out of ...” → “55 out of ...”
- 2) Page 7862, line 13: “4” → “four”
- 3) Page 7863, line 1: Use “evaluated” instead of “validated”
- 4) Page 7863, line 26: leave out “Celsius”
- 5) Page 7864, line 22: “with different set” → “with different sets”
- 6) Page 7865, line 11: “the number ... were limited” → “the number ... was limited”
- 7) Page 7867, line 1: “Tawari” → “Tewari”
- 8) Page 7868, line 11: “to keep” → “in order to favour”
- 9) Page 7869, line 1: “maximal” → “maximum”
- 10) Page 7869, line 5: “during heatwaves years” → “during heat wave years”
- 11) Page 7869, line 6: “in a lesser extent” → “to a lesser extent”
- 12) Page 7869, line 7: “findings found” → “findings reported”
- 13) Page 7869, line 22: “under” → “below”; “the middle of the simulations” → “the mean value (the median value?) derived from the simulations”
- 14) Page 7870, line 2: “how temperature clusters” → “how resulting temperatures are clustered”
- 15) Page 7870, lines 18-19: “affect radiation before heatwaves” → “affect radiation prior to the onset of heat waves”
- 16) Page 7870, line 23: “of Sect. 2” → “introduced in Sect.2”
- 17) Page 7870, line 24: “model-data” → “model-observation”
- 18) Page 7871, lines 1-2: “The same is found ..” → “The same is not only found ...”.
Please also indicate for each statement the season and region. “for the years 2007 and 2010 in Russia” probably should be interpreted as “for the years 2007 in France, and 2010 in Russia”.
- 19) Page 7871, line 21: “... are largely overestimating ...” → “... are found to considerably overestimate ...”
- 20) Page 7871, line 23: “Tiedke” → “Tiedtke”
- 21) Page 7871, line 25: “fairly simulated” → “fairly well simulated”
- 22) Page 7872, line 27: “before” → “prior to”
- 23) Page 7873, line 1: “feedback” → “feed back”
- 24) Page 7873, line 4: “atmospheric schemes” → “atmospheric physics schemes”
- 25) Caption of Table 1: “Physic schemes” → “Physics schemes”
- 26) Caption of Figure 1: “Daily time series of temperature” → “Time series of daily mean temperature”
- 27) Caption Supplementary Figs 2: “2a-d” → “2a-f”