Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-622-AC1, 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "A global scale evaluation of extreme events in the *eartH2Observe* project" *by* Toby R. Marthews et al.

Toby R. Marthews et al.

tobmar@ceh.ac.uk

Received and published: 13 March 2019

Response to Anonymous Referee #1 (Received and published: 25 February 2019):

General comments The manuscript presents an analysis of a unique dataset that was produced from the eartH2Observe project. This dataset involves the simulation of several hydrologic variables from a number of state-of-art land surface/hydrologic models and using as forcing several satellite and reanalysis dataset. The scale of analysis is global and the focus is on the tails (i.e. low/high extremes) of evapotranspiration and surface runoff. Overall the work is very interesting and the dataset analyzed is very unique. Additionally, the fact that the analysis is performed at global scale provides important information on the regional variability of findings. The manuscript is generally

well written but there are certain parts (especially in the description of methodology) that require additional clarification and discussion. I provide some specific comments below that hopefully will help the authors to improve their manuscript.

– RESPONSE – Thank you very much for the review and we hope that the responses below are sufficient reply to these useful comments.

Specific comments 1. I believe that the title should be revised to better reflect the context of the paper. One of the main elements of you analysis is "uncertainty in identification of extreme events" but this is not reflected in the current title. – RESPONSE – We agree and have added "uncertainty" to the title.

2. Abstract L17-19: I agree but given the focus of your analysis (i.e. identification of extremes) you should be more specific on what your results will allow to comment. For example, models can be quite robust in representing the main body of the distribution of hydrologic variables, which is actually very important for water resources applications. So I suggest to specifically refer again to representation of extremes. – RESPONSE – We agree: we have now specified in the abstract that "Our results are important for highlighting the relative robustness of satellite products in the context of land surface simulations of extreme events" as requested.

3. P3L3: For a multiregional evaluation of satellite precipitation over complex terrain, you may want to consider also Derin, Yagmur, et al. "Multiregional satellite precipitation products evaluation over complex terrain." Journal of Hydrometeorology 17.6 (2016): 1817-1836. – RESPONSE – Thank you for this reference: it was indeed relevant and we have used it at two points in the revised text.

4. Information in Section 2.1 needs to further clarified. What do you mean by "base distribution"? Is this the reference for your EE/yr at each cell? Why you average the five runs and you don't consider each model independently? Do you repeat the same procedure for each product and then compare? Please clarify. – RESPONSE – Thank you for pointing this out: we had not explained why a base distribution (or baseline

distribution) is necessary. We have added to section 2.1 “Extremes for any particular variable may only be assessed in relation to an estimate of ‘normal’ conditions, and for this we took a baseline distribution of values ...”. At this point we have not considered each model independently because we do that at a later point when comparing model uncertainty and data product uncertainty. Finally, yes we did exactly repeat the same procedure for each product (and each model) and compare: this is the basis of the definitions of alpha, epsilon and beta later on in the method.

5. It would be very useful to provide a graphical example to demonstrate the different uncertainty components that you describe in equations 1-3. – RESPONSE – We have indeed included a graphical example of the different uncertainty components: this is Fig. 7. We have considered moving that figure to an earlier point in the paper, but it is difficult to do so because we need to explain more details of how it is calculated before presenting that figure.

6. L24-28 are confusing. First, it is not clear why you consider εx,j >1 as an indicator of model amplification of uncertainty? Do you mean αx,j instead? Also if you want to identify the relative contribution of the different sources of uncertainty, why don't you take the ratio of α/β ? – RESPONSE – We apologise but we have to disagree with the reviewer on this point and stand by the text as we have written it: it is indeed epsilon rather than alpha that measures model amplification. Take this example: Let's say at a particular dryland point on the globe precipitation varies over the range 400-1000 mm/yr. According to our definitions, which are phrased in terms of extreme event occurrence, data input uncertainty DIU will perhaps be 2.5 EE/yr (say we are concentrating on high extremes and those precip numbers equate to extreme high precip occurring between 2 and 7 times per year). Say model 1 is quite ‘flashy’ in the sense that runoff is very sensitive to precip but ET is calculated from an empirical relationship quite insensitive to precip extremes, DOU might be high for runoff highs (e.g. 5.0 EE/yr, say) but very low for ET highs (e.g. 1.5 EE/yr). For model 2 you might easily have the reverse with runoff DOU being low and ET DOU being high. Across all models, let's say

runoff is always flashy; whereas ET is sometimes flashy and sometimes not, depending on whether a conservative empirical representation has been used in that model or something more realistic. With these example values, we would expect to see alpha always high for the ‘flashy’ runoff and quite high for ET too (the flashy runs averaging out the nonflashy ones). We would also expect beta to be low for runoff (because it is consistent across models) but high for ET. Epsilon=alpha+beta is a measure of the total propagation of uncertainty and this is the most appropriate measure to look at when talking about uncertainty propagation, because to look at only alpha is equivalent to ignoring the model uncertainty, and this is one of the important points we are trying to make with this paper. To look at the ratio of alpha to beta is also not appropriate because for that DIU drops out of the calculation and we cannot take account of the variation in the driving data (see eqn1 and eqn2). We have added some text to section 2.2 (“In summary ... product only”) that we hope clarifies this point and the relationship between alpha, epsilon and beta.

7. P6L6: “global average”, why do you consider global average? It is not advisable since the average masks regional variability. Also “ET highs (58.1% vs 41.9%)”, it is not clear what these numbers correspond to. – RESPONSE – This is an introductory point at the start of section 3.1 which we then expand upon in more detail later. It is not irrelevant to point out that the alpha values are universally quite small and do seem to decline with increasing precipitation. the regional variability is displayed graphically in Figs. 4-6.

8. P6L10 “αx,j<-1”, I believe you mean log(αx,j). – RESPONSE – This statement is correct as stated (a log cannot be negative and we felt it was clearer to quote the bound in terms of straight alpha at this point).

9. P6L19-23. Interesting findings, some additional comments are welcome here. For example, why “the magnitude of the increase reduced in wetter environments”? – RESPONSE – We feel that it would be too speculative to include here any of the various theories that could explain why the magnitude of the increase is reduced in wetter

environments. For example, there could be a saturation effect in the environment (but in the absence of soils or land use data we cannot be sure of this) or fast drainage could occur more often under more episodic rainfall (but we have no data on drainage patterns) or the occurrence of convective cells might be very regionally specific (but these are not even visible on most remote sensing products). We have tried to be careful to stick to discussing points that are directly relevant to the results and data that we have presented and we hope in a later study to look at these trends in more detail, but we have omitted any discussion of this here.

10. P6L25: "The global mean value. . ..is a measure of variability". How can a mean value tell you anything about variability? Please clarify/revise. – RESPONSE – If the quantity in question (alpha) is itself a measure of variability, then the mean of alpha will still be a measure of variability even though we agree it will not contain any information about the variability of alpha itself. We have revised the wording here to avoid the apparent contradiction.

11. P6L25-30: In general, this part of the text is quite difficult to "digest". Please improve clarity. – RESPONSE – We agree and thanks: we have removed the middle sentence, which we hope has improved the clarity of the paragraph.

12. P6L31: What do you mean by "internal model uncertainty"? – RESPONSE – We have added in the explanation that this is "a measure of the diversity of the calculation methods used to derive X between models".

13. P7L3-4: ". . .are more sensitive to precipitation extremes in wet environments". Be careful here, you should state ". . .more sensitive to precipitation uncertainty". – RESPONSE – Corrected with thanks.

14. P7L15-16: I believe that there is a confusion here between model uncertainty and uncertainty propagation. This is a very important aspect and the authors should clarify it in their discussion. For example, even with zero model uncertainty, transformation of precipitation uncertainty to runoff uncertainty could potentially amplify as a result of

the nonlinear transformation of rainfall-to-runoff. – RESPONSE – We have specifically defined separate quantities for model uncertainty (beta) and uncertainty propagation (epsilon) and we believe that we have not confused the two issues in this paper: in fact, drawing attention to the difference is one of the overall points of the paper. If runoff is generally 1000-1500 mm/yr with 7 peaks/yr when precipitation inputs are 500-1000 mm/yr with 3 peaks/year, then output uncertainty differs not only in terms of absolute value (which can be a linear effect) but also in terms of distribution (a nonlinear effect). By focusing our study on extreme event occurrence, linear effects should be cancelled out (as long as the extremes are calculated in terms of an appropriate baseline for each quantity, which we have done), however of course there will be nonlinear effects that can give nontrivial values to epsilon (and alpha) even in the case of beta=0 because the number of peaks may still change. At no point in the paper have we assumed that this will not happen: in fact, we have accounted for this in all analyses. At P7L15-16 we have simply stated that model uncertainty is usually greater than data uncertainty. We believe that the reviewer here does not like the implication that when model uncertainty is small then data uncertainty must be even smaller, and it was certainly not our intention to imply that. We have modified the text to say "when a set of models is under consideration, model uncertainty is usually greater than data uncertainty". To avoid the same implication we have added "in a simulation ensemble" to the start of section 4.2 as well.

15. The same point as in 14(above) should be considered in the discussion of section 4.2 (e.g. L26-27). – RESPONSE – Please see last point #14.

16. P9L10: ". . .to improve prediction of water cycle quantities". Ok I agree but the analysis presented has not done anything on the quantitative aspect. Perhaps revise to "improve prediction of water cycle extremes"? – RESPONSE – Thank you for the suggestion: changed to "extremes"

17. Section 4.3. (L15-22). The text here is relevant to work that is evaluating uncertainty and compares against observations. However, this is not the scope of your work.

You isolate (correctly) the forcing and model uncertainty by considering as reference a model/forcing combination. – RESPONSE – These comments are made in a section entitled “4.3 Sources of unquantified uncertainty” and we state clearly in the preceding sentence that we could not analyse these kind of situations in our particular study given the data available. However, we find these issues the be entirely within scope of this study and, in fact, we would have been remiss not to have mentioned them. We hope very much in a follow-up study to find some way to tackle these sorts of issues and we believe it is entirely appropriate to have a brief mention of them here.

18. Fig2: What is (a) and what is (b). Also, some of the explanation on the calculation of results could be added to text in manuscript as well. – RESPONSE – The legend stated “a. Uncertainty in precipitation extreme highs and b. Uncertainty in precipitation extreme lows”, which perhaps was not clear because we did not use parentheses on the (a) and (b), so parentheses have been added in. We agree that the explanatory text in the legend was perhaps too long and was mostly superfluous because the calculation is already described in the main text (section 2.1) so we have now omitted it.

19. Fig3. Similar comment on the explanation. – RESPONSE – Unlike for Fig. 2, the description of the calculation for Fig. 3 is not repeated in the main text, but after reconsidering the legend we would like to argue that the amount of detail here is appropriate: the explanation here simply describes what the rows and columns of this multi-panel plot display and we do not see any way to appreviate this without forcing the reader to hunt through the text for this description. Therefore we have left this text as it is and we hope the reviewer will either reonsider this comment or specify more precisely what change he/she would like us to make, please?.

20. Figure 4. I find this map very useful. It would be nice to provide for the other cases analyzed. – RESPONSE – The other 3*3=9 maps from Figs. 4, 5 and 6 were excluded before simply from space considerations. They have now been added.

21. Figure 7: “erros bars show SE”. Do you mean standard error? And how the error is defined. Perhaps you refer to standard deviation instead? – RESPONSE – We have left this text as it is: what was calculated here was standard error, which differs from standard deviation because you divide by the square root of sample size (the abbreviation SE is standard). The “averaged over 50°S to 50°N” earlier in the legend makes it clear that this is calculated across gridcells rather than time (i.e. sample size is the number of gridcells in this case).

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-622, 2019.