Online discussion

1 RC2**: ['Comment on hess-2021-109'](), Anonymous Referee #2, 27 May 2021**

General comments:

This is a potentially interesting study comparing JULES-CaMa-Flood simulation output with a global dataset of inundation extent for different selected wetland regions.

Specific comments:

- There are 10 case study wetland regions, but the study spends insufficient time on most, if not all of them.

We thank the reviewer very much for their interest in the topic of our paper. We were very much hoping that submitting to *HESS* would elicit wider interest in our work on wetlands and we are very glad that this has turned out to be the case.

It is difficult to respond adequately to this request: all our case study wetlands are individually so well known that entire books have been written on many of them. Given the space limitations of a journal article - and our global perspective - we restricted our general site information for each case study area to a few brief lines in Table 1, where appropriate references have been given. We agree that we could have included a little more information on each wetland, but even devoting a paragraph of introduction about each wetland would have comfortably put us over our word limit for the paper, at the same time as opening us up to criticism that the information was superfluous because already published elsewhere. After much thought, we restricted ourselves to the minimum information on each wetland that was required for the reader to understand and interpret our results, which is the location and size of each (as given in Table 1), and in Table 1 we have collected together good references for each region through which the reader may access much more information about each very unique environment.

- The results section needs improving. The results within and between the regions studied should be compared in quantitative terms. The main/most important findings should be identified and highlighted.

Having considered the text in Results and Discussion, we believe that the problem identified (correctly) by the Reviewer here is that certain statements in our Discussion should have been placed in Results. We have revised the text that refers to Figs. 6 and 7 (formerly Fig. 5) and we hope very much that the new balance between Results and Discussion is much more readable than before.

- There are insufficient insights presented in the discussion. Can the errors and biases (based on the NGE, KGE, alphas and beta) at different locations be better attributed to the quality of JULES versus JULES-CaMa-Flood simulation? Are they predominantly a result of model structure, parameter or forcing errors? To what degree is uncertainty in the remote sensing data responsible? How do climate, season, and hydrotopograhy factor in? The authors must relate their results/findings with existing knowledge of earth system/hydrological processes of the different wetland regions studied.

The reviewer is disappointed that we have not provided a complete explanation of the biases we have detected in terms of physical variables (e.g. hydrotopography) and this is completely fair (and we share the disappointment!). However, in this paper we must restrict ourselves to conclusions that are justified by the data we have collected and the analysis that we have carried out. Ultimately, even though we have carried out a fairly sophisticated analysis, we are comparing only one variable (inundation extent) and this imposes hard limits on how much we can conclude. We could have included much more in the Discussion about what we believe to be the reasons for, e.g. the spatial displacement of inundation we detected in the Niger

Inland Delta or the apparent overestimation of water output from the Okavango swamps. These sections would perhaps have made our Discussion more 'insightful', but we believe that we would have been speculating because we do not have additional observational data to reinforce any such discussion. It may be of interest that we are currently working on a follow-up paper where we attempt to identify mechanisms of inundation in a much more detailled way, but this requires more model runs (with physical processes turned off/on) that we could not carry out for this study.

We have discussed several topics relevant to the hydrodynamics of wetlands in sections 4.2 and 4.3, linking back to previous work in the Amazon. However, there are surprisingly few papers that have compared observed and predicted inundation extents with robust protocols similar to our study, so we do not have many previous papers to compare to. Additionally, most global studies are carried out at spatial scales of 1º or 2º or even coarser scale and report in terms of zonal means or meridional means, which are measures that average over enormous areas of land and we find to be only peripherally useful at our spatial scale of analysis. The studies we have referred to in the Discussion are the most relevant and comparable studies to ours, assessed after a much wider literature search on our part. If this reviewer has particular studies that he/she would like to suggest that we have perhaps missed in our search then we would very much like to hear about them, but given our current data and analysis we believe this Discussion goes as far as our data permit us to go.

- The authors propose the alpha and beta parameters as indicators of model bias in the simulation of evapotranspiration and infiltration. However, these cannot be expected to be constant over time. Additionally, the NSE and KGE were calculated over the full temporal domain. Can the authors be confident that the high performing parameters remain valid at a different time? Additional analysis is warranted to investigate this.

We understand the drift of this comment to be a concern that by averaging over time (as required by the definitions of NSE or KGE), we have lost the information on any variability over time. This is certainly true, but is a consequence of the definitions of these statistics. The reviewer is absolutely correct that we cannot expect these indicators to be constant in time and we did state this in our Methods ("our match statistics were dominated by the regular (seasonal) and irregular cycles occurring at points where inundation was not constant").

During the analysis of this paper we attempted three different analyses relevant to this point, each with their own problems and drawbacks, which may be explained by imagining a series of inundation fraction numberscovering a certain time period P:

1. We looked at the spatial variability across pixels, but with each pixel's data averaged over the whole period P
2. We looked at the seasonal cycle for each wetland, which involved averaging across the spatial extent of the wetland and then compiling each month to acquire monthly averages (all the Aprils, all the Mays, etc.).
3. Looking at a seasonal cycle for each individual pixel

Note that studies based on river gauge data can only attempt approach #2 because they do not have spatial data (and even the gauge data itself cannot be considered a mean of the wetland's dynamics because it is biased towards the dynamics immediately around the gauge itself): Approach #1 especially is a high point of novelty of our study: the vast majority of wetland studies are still based on gauge data only and we are not aware of any other study that has been able to take this spatially-explicit approach.

Having attempted all of these, we quickly discarded #3 because of wetland fluctuations: if the inundation fraction goes to zero (i.e. a particular point finds itself temporarily outside the wetland) or 100% (i.e. the central lake that many wetlands have extends outwards) then our NSE and KGE statistics register –Inf, zero or one values. These values then outweigh all other values in a straight averaging scheme. The usual solution is to mask out all points that are peripheral to the wetland area in this way, but with many wetlands this gave us no remaining points at all. Partly, this is because of the recognised deficiencies of these standard statistics (which we reviewed in Table 2), but effectively it meant that we had to discard this option.

Our results presented in Fig. 3 (formerly Fig. 1) and in Fig. 5 (formerly Fig. 4) are examples of approach #1, the results in Fig. 4 (formerly Fig. 3) are from approach #2 and there is much to be learnt from both analyses about the spatial and temporal variability of these wetlands. Please note that because NSE and KGE require comparing time series, they can only be calculated from approach #1.

We did consider stratifying the data into two periods of 11 years each and comparing 'before' (1992-2003) and 'after' (2003-2014) NSE and KGE values - and we did in fact carry out this analysis – but we found the conclusions we might have drawn from this were statistically unsafe. For example, what would it signify if we found a 10% increase in KGE for the Congo between the before and after periods? It could perhaps suggest climate change, but we could not exclude the possible effects of other social change in the local area (e.g. for the Congo, our 'before' period ended approximately at the end of the Second Congo War). Despite having found some interesting trends between 'before' and 'after', we decided not to include these results in the paper for essentially statistical reasons.

A similar argument applies to the bias-correction parameters *alpha* and *beta*. We agree that we do not expect these to be constant over time, however the nature of the statistics used required a time period for calculation and the period 1992-2014 was the longest period for which we had the required data. However, we do also agree that we have not mentioned this important point in the paper, therefore we have now added the following text to the end of the Results: "Finally, we note the specificity of our results to the time period 1992-2014. Carrying out this analysis for an earlier or a later period would certainly yield different estimates of NSE, KGE, *alpha_min*, *alpha_max* and *beta*. However, we suggest that without significant climate change, or perhaps significant anthropogenic modification of the wetland area concerned, the values of these statistics should remain similar to the values calculated here.".

Technical comments:

- Abstract: the final sentence of the abstract claims "This study provides timely data". However, it is unclear from the manuscript what part of the results/findings this "data" is referring to. The alpha and beta parameters are possibly only useful for JULES-CaMa-Flood-GIEMS users.

This sentence has been rephrased to state clearly that it is information on the biases in data that are useful and timely.

Even though we have been careful to restrct our conclusions to GIEMS and JULES-CaMa-Flood data only, we do not believe that these are the only communities who would find use in these data: all satellite-based inundation data has biases that may be assumed to be very similar to those inherent in GIEMS data, and all model predictions of inundation have biases and uncertainties presumably similar to those that are in JULES-CaMa-Flood predictions, so we believe that this paper also provides a blueprint for users of other model/observational data on how they may assess and account for some kinds of bias in their own data. We have added this comment to the conclusions in section 4.4 as well now.

- Abstract: in line with the earlier recommendation to investigate all the different regions more thoroughly "(including the Sudd, Pantanal, Congo and Amazon)" should be removed.

Parentheses removed as requested

- Line 35: what is being referenced to in the cited reference Saunois et al 2020 is unclear.

Reference removed

- Line 100: "Most hydrological models are run uncoupled from the atmosphere and are therefore reliant on the availability of good precipitation and other atmospheric driving data." – the first part of this statement is inconsequential. Even if hydrological models were run coupled with atmospheric models, a high level of error from the simulated precipitation is still expected.

Perhaps this might be considered a point of terminology only, but we do see a significant difference here. In an uncoupled framework, the user acquires precipitation data from (usually) a published source (e.g. MSWEP data) and the uncertainty in the precipitation is considered a form of data uncertainty. Conversely, in a coupled model context the precipitation is calculated by one of the component models of the coupled framework (the GCM/RCM) and any uncertainty in this is therefore considered to be internal, i.e. model uncertainty. The point being made in this paragraph (and in my 2020 paper referred to here) was about data uncertainty *vs.* model uncertainty. We would completely agree that precipitation has notoriously high uncertainty whatever the modelling framework, but sometimes this is external and sometimes it is internal and this is what was referred to here. We have changed "availability of good precipitation and other atmospheric driving data" here to "the availability of independently-sourced precipitation and other atmospheric driving data" to "availability of high-quality precipitation and other atmospheric driving data obtained from independent sources".

- Include a study area figure at the global scale to adequately introduce the wetland regions and discuss their differences in major processes/controls. This will remove the need to refer to a few of these regions as "the three tropical zones", which can be confusing for the reader.

Thank you very much for this suggestion: we have now included a new Fig. 1 that shows the locations of all 10 study areas, from which we feel it is now much more clear that 3 are larger than the others and are more correctly referred to as "tropical zones" rather than "wetlands".

- Lines 201, 215: references to figures from the results section within the methods section should be removed.

The reference to Fig. 1 on line 201 has been removed, but the reference to Fig. 2 on line 215 has been retained because this figure was indeed truly a 'Methods' figure. In order to clarify the text, the figures have been renumbered so that the 'Methods' figure comes before the 'Results' figure.

- Line 222: "We therefore calculate spatial matching statistics across all case study areas" - it is unclear what is being meant here by spatial matching statistics.

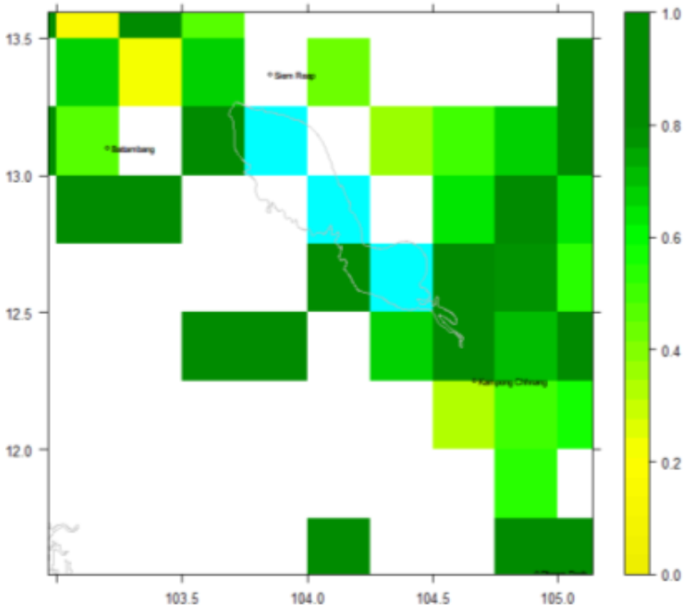
Thank you for pointing this out: we indeed had only used this term once and we agree it is not especially informative. The setnence has been rephrased with "appropriate statistics" instead

- Lines 228-229: The evaluation metrics nRMSE, r, RMSE were not introduced in the methods, nor were their results presented.

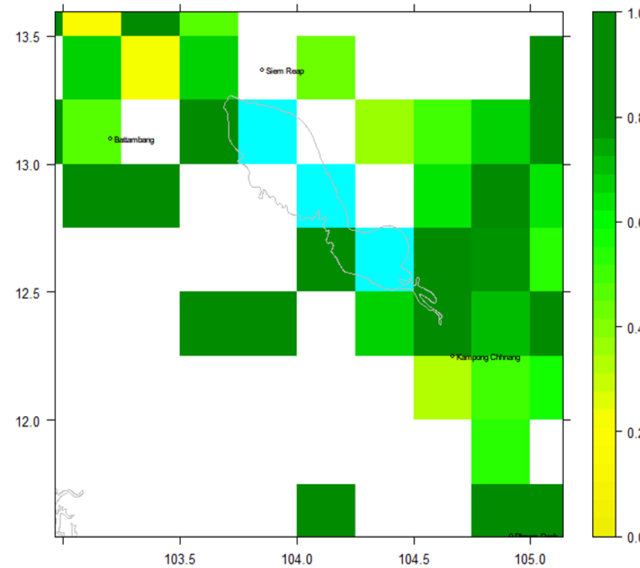
We did state in our Methods section that the results of our RMSE and Pearson's r calculations were not presented. From feedback during the writing of this paper, there is a wide expectation that these statistics should be applied in an analysis such as ours, so we should mention that we have done so, but we found that there were no additional conclusions to be drawn from the RMSE and r plots that could not be drawn from the corresponding NSE and KGE plots (and the trends were always more clear on those plots too). Statistically, this is simply because NSE and KGE include (i.e. may be decomposed into) the definitions for RMSE and Pearson's r (as we describe in Table 2). In summary, we need to mention RMSE and r briefly because of the widespread expectation that they should be applied, but because NSE and KGE include these statistics, there were no additional results to be learned from these plots and we did not wish to mention them more than briefly (especially because presenting these results as well would have meant including an additional Fig. 5 (formerly Fig. 4) and Fig. 6 (formerly Fig. 5a) based on these statistics).

- Line 238: "However, these statistics are not capable of measuring some aspects of the flow regime that are important from the point of view of allowing us to divide out the different sources of inundation in our study wetlands" is unclear.

We have rephrased this sentence with an extra few lines explaining the significance of spatial displacement of inundation.



PDF online version in HESSD



Version submitted (Word Doc)

- Figure 1, 4: the results for regions with a larger spatial domain are difficult to see.
- Figure 5 is blurry

We apologise for the difficulty here: it is difficult to present continental-scale results in an easily-accessible figure in this way because we are limited in terms of available space. However, we do note that we have included high-resolution images in these figures and the user can zoom into the figure to see additional detail. This will also be possible in the online version of these figures as well.

We also note that in the PDF versions of these figures available for review, the quality of these images seems to have been downgraded (see excerpt from Fig. 4 above). This will not occur if the paper moves to HESS for full publication and the images will be much sharper.

- Line 237: "within the borders of the wetland itself" – at/near the wetland boundaries?

Apologies for not having been clear here: we meant "across the wetland" and have rephrased this sentence appropriately.

- The authors' conclusion in lines 284-288 is poorly supported.

The conclusion in question here was our statement that "The visible maxima on these state space plots provide a best estimate of the optimal values of these parameters, with these optima differing markedly between our wetland study areas (Table 1). A notably higher value for NSE or KGE for a particular combination of *alpha_min*, *beta* and *alpha_max* would identify a consistent bias in either the model predictions or the observations (or both)."

With apologies, we believe that this statement is emphatically supported by our data: our search of the parameter space values to obtain these visible maxima was exhaustive and therefore the conclusion that these are real optima is robust. Some of these optima were like the high point on a gently sloping hill rather than a clear mountain summit, but very clear optima on state-space diagrams are in any case rare in analyses such as this. The optima very obviously differ between our case study wetlands (as is visible from Fig. 6, formerly Fig. 5a).

We would request reviewer RC2 to please be more specific as to what aspect of this he/she found unconvincing and we will try our best to redress?

- Line 302: “If overbank flooding is underestimated in our simulation then the water within the river course (the Niger or White Nile, respectively, in these cases) will remain in the river and be taken downstream further than expected, producing a downstream wetland ‘extension’ that exists in the simulation results but not the observed (as we see in our JULES-CaMa-Flood outputs). – this needs to be better linked better to the study results, with examples.

This is an example of the process of spatial displacement of inundation. In response to another reviewer comment above, this has been explained more fully now in the text at lines 245-250.

- The manuscript is informal in tone and unfocussed at some parts, with longwinded sentences that make it hard to read. Additionally, there are:
    - acronyms undefined at first use e.g. CaMA, GIEMS, GLWD, WRR1, WRR2

We apologise for this: all these acronyms have now been defined at first use in the text.

- use of biased/subjective words, e.g. “surprisingly”, “sophisticated”

We have removed the one occurrence of “sophisticated”, but we would like to retain the one occurrence of “surprisingly” in the first sentence of the Abstract. We are stating here that “reliable data on the extent of global inundated areas ... remain surprisingly uncertain” and we do believe that this is objectively surprising.

Without exception, when we have discussed this paper with non-hydrologists in order to receive some feedback on the content of the paper, the first reaction has been surprise that information on inundation extents should not be extremely well-known. It is unclear what this assumption is based on, but we suspect simply because it seems such an easy thing to measure (a depth gauge) and has been measured in some locations for so long (e.g. the Ancient Egyptians measured the inundated extent of the Nile river), so people assume that in the ‘modern’ world we should have universal data on this quantity. Once the difficulties are explained (measurement in ungauged catchments, real time requirements, satellite uncertainties) then it generally becomes clear, but even for hydrologists (and, we assume, the audience of *HESS*) the current state of global data is almost universally a surprise. To remove this word at the start would imply that our results are of far smaller importance than we believe them to be, so we would like please to keep it.

- overstatements, e.g. “widely used” (relative to the few references cited)

We have been trying to avoid over-referencing in this paper. When we stated that CaMa-Flood was a widely-used model, we referred to Hoch & Trigg (2019) and Zhao et al. (2017), both of which are good examples of CaMa-Flood being used but also give references to other studies where the same model has been used for more than a decade now. We believe that either one of these references on their own would be sufficient to support this statement, and we could have added in more examples, but we did not feel that more references were warranted in this case.

- excessive use of brackets and italicized phrases that highly disrupt the flow

We are not aware of any excessive use of parentheses or italicised terms: the topic under discussion is relatively technical, which unfortunately requires use of technical jargon (on the use of which this reviewer has corrected us a few times, for which we are very grateful!) and we have tried as far as possible not to make unsupported statements that may be interpreted as subjective, which requires a fair number of citations, but we do not feel that these have been excessive. We have been through the text completely and have attempted to improve the flow of the text at certain points, and we hope very much that this makes the paper read a little easier.

- generic/blanket statements such as: “We found that our simulated inundation extents (from the CaMa-Flood model, driven by JULES runoff data at 0.25° resolution) sometimes compared very closely to our observed data (from GIEMS satellite-based data), but at many points there were divergences”, and “The spatial displacement of inundation prediction downstream from observed inundation visible especially in our results for the Inner Niger Delta and the Sudd (Fig. 1) is a result of over- or under-estimation of overbank flooding upstream." There are multiple occurrences throughout the manuscript including in the abstract.

The overall readability must be improved.

We have tried to improve the readability of the text throughout. Addressing the two examples highlighted here, these are direct statements of fact, however: our simulated extents did indeed match very precisely at certan points, but not at others (here we were introducing the results and making the general comment that there was variability in how well the modelled extents matched observed) and the second is a statement about spatial displacement of inundation that is indeed usually a result of upstream effects. We fully admit that if our intended audience were hydrologists only then these would be almost superfluous statements, but the readership of *HESS* is wider than just hydrologists so we believe that statements like these are necessary in order to introduce concepts that may not be familiar.

Overall, we would like to thank RC2 as well very much for these results and for his/her time spent on our paper: we understand that it is quite an imposition to do a review and are very grateful for the feedback received.

Toby Marthews et al.

Citation: https://doi.org/10.5194/hess-2021-109-RC2