Author Comment: Evaluation of hydroclimatic biases in the Community Earth System Model (CESM1) within the Mississippi River Basin

Reviewer 1

In this paper, O'Donnell evaluated hydroclimatic biases in the CESM1 within the Mississippi River basin. The evaluation data include USGS gauge data of river discharge, ERA5 reanalysis, GPCC precipitation observations, and LIvneh ET. They also compared the CESM1 simulated runoff with the simulations from several CMIP6 models, including the newer version of CESM - CESM2. They demonstrated that CESM1 has substantial biases in simulating runoff and river discharge and attributed the model discrepancy to the deficiency in the RTM river model. They showed that CESM2 with the more advanced MOSART river model performs better in the river basin.

We thank Reviewer 1 for the thoughtful and useful comments on our manuscript. We do agree with many of these comments and will adopt these changes to better represent the hydroclimatic biases in CESM1 in the Mississippi River Basin and motivations of the work.

While the results are clearly presented, I find that the motivations of this study are not clear and there are likely serious errors in the CEMS1 configuration or simulation or both. As such, the study has limited values and I cannot recommend its publication in this journal. There are two major gaps/issues in the paper. First, the authors have not explained clearly why we need to know the biases of the old CESM1 given that the newer version CESM2 has been used by CMIP6. Does CESM1 have unique features that are not available in CESM2? Is there still a large user base who is using these features for important studies? What are the obstacles that hinder the users to adapt to the new version? Without good reasons, I would question why not to evaluate CESM2 instead.

We agree with Reviewer 1 that the motivations can be further clarified. CESM1 is still of significant value because it is one of the few CMIP5 models that has both a routing model and multiple available modeling projects, including the Last Millennium Ensemble (CESM-LME) (Otto-Bliesner et al., 2016), which includes both full-forcing and single-forcing simulations for the period 850-2005 CE. While we demonstrate here that CESM2, which is a part of the CMIP6 suite and uses MOSART, has significantly improved seasonal timing, neither CESM2 nor other CMIP6 GCMs yet include equivalents of the LME project simulations. Moreover, it is useful to evaluate the degree to which CESM2/MOSART represents an improvement over CESM1/RTM – which we do. A number of studies still use CESM1 to investigate hydroclimate over the last millennium. For example, recently published papers such as *PDO influenced interdecadal summer* *precipitation change over East China in mid-18th century* (Chen et al. 2024. *Nature Climate and Atmospheric Science* 7 (1): 1–11. https://doi.org/10.1038/s41612-024-00666-6.) and *Influence of ENSO and Volcanic Eruptions on Himalayan Jet Latitude* (Thapa, Uday Kunwar, and Samantha Stevenson. 2024. *Geophysical Research Letters* 51 (14): e2023GL107271. https://doi.org/10.1029/2023GL107271.), among others rely on CESM1. Any studies using CESM1 must take into consideration the biases in CESM1, particularly if they focus on the Mississippi River Basin, and if a study focuses on other regions, similar biases should be evaluated before conclusions on hydroclimatic changes are drawn.

Second, the model simulations look suspicious. Table 3 indicates that the modeled surface runoff, subsurface runoff, total runoff and snowmelt are two orders of magnitude smaller than the observations or benchmark data. Given this unbelievably poor performance, I would honestly think the model is useless. It is reasonable to question whether the authors have configured the model or extracted the outputs correctly because CESM1 has been well tested before. Furthermore, there are also several other variables with odd values: 1) the reported precipitation values (Figure 1b and Line 89) are less than 200 mm/year which if true would mean that the Mississippi River basin would be a desert; and 2) the reported runoff values in Figures 3 and the reported snowmelt values in Figure 4 if converted to mm/yr are unrealistically large (1e-6 m/s > 3e4 mm/yr). Also, it is very strange that the model and data are not shown at the same scales in the model-data comparisons (Figures 3, 4, and 7).

We agree with Reviewer 1 that the differences in magnitudes between simulations and observations are large. We have checked this previously, but will confirm any conversions and all values again. We can also add a note in any figure captions where different y-axis scaling is used. While CESM1 has been widely used and some diagnostics are available, the user documentation specifies the need for rigorous validation ("CESM Overview." n.d. CEMS1.1.z User's Guide. Accessed August 9, 2024. https://www2.cesm.ucar.edu/models/cesm1.1/cesm/doc/usersguide/x21.html).

We also thank Reviewer 1 for pointing out this error in reported precipitation value units. The units should be mm/month, and this will be corrected in an updated manuscript.

My other comments are shown below.

L104: Why did you choose the USGS 07289000 which is only available since 2008 for the Lower MS? Why not choose 07295100 Mississippi River at Tarbert Landing, Mississippi which has much longer data records for investigation?

We agree with Reviewer 1 that longer records are useful for comparison. The USGS 07295100 Mississippi River at Tarbert Landing does not have monthly discharge data, but the US Army Corps does for this site. The USGS 07289000 gage data was used for the comparison of simulated to observed data from gages, but the 07295100 Mississippi River at Tarbert Landing data was used in the pre- and post-dam comparison. In an updated manuscript, we will update Tables 1 & 2, and Figure 2 to reflect data incorporated from 07295100 Mississippi River at Tarbert Landing and update citations accordingly.

Table 2: Could you explain why these years can be regarded as the separation of pre-modification vs. post-modification? For example, for the Missouri River, many of the dams were constructed in the 1930s. As a result, I do not think you can see much difference by comparing the model simulation before and after 1967.

We agree with Reviewer 1 that the delineation between pre- and post- modification years needs to be well supported. The years listed in Table 2 are the end of the periods of major river modification when the impacts of the modifications are fully implemented, based on the literature cited in the Table 2 caption. For example, Jacobson & Galat (2008) note that "The six mainstem reservoirs were constructed between 1937 and 1963 and operation as a system began in 1967" on the Missouri. We can further clarify the choice of these years for the subbasins in a revised manuscript.

Section 2: Given the importance of runoff generation and river routing in this study, wouldn't it be necessary to describe CLM and RTM briefly? Particularly, how is water management represented in RTM?

We thank Reviewer 1 for pointing this out, descriptions of CLM, RTM, and representation of water management are useful context and we can add these in a revised manuscript.

L135: QOVER is only a part of surface runoff and does not include surface runoff from standing water (QH2OSFC).

We agree that it is important to include all components of surface runoff. However, in the CESM1 LME project from which we are retrieving data, standing water (QH2OSFC) is not available as a variable and QOVER is representative of surface runoff. While RTM uses CLM4, we also found that QOVER is noted as including QHSOSFC in CLM5 ("Questions on Runoff Components in CLM5 BGC-CROP Mode (Ctsm5.1.Dev118)." 2024. *DiscussCESM Forums*. https://bb.cgd.ucar.edu/cesm/threads/questions-on-runoff-components-in-clm5-bgc-crop-mode-ctsm5-1-dev118.9271/.). We did not include the additionally available surface runoff term, QRGWL (surface runoff at glaciers (liquid only) wetland lakes), since this project does not investigate runoff or the hydrologic cycle components in a glaciated area.

L144: Could you describe briefly these 13 ensemble members? Under what configurations these members were simulated?

We agree with Reviewer 1 that a description of the 13 ensemble members from the CESM LME project, as well as their configurations, are important context and we will add this description in section 2.4 Earth system models and validation approach, where the ensemble members are introduced.

L151: What software do you use to calculate lagged correlation and spectral angle?

We agree that this should be specified and will update the methods section of the manuscript to include that lagged correlation and spectral angle were calculated in Python with the pandas corr and HydroErr sa functions, respectively.

L265: It is probably not true. To my understanding, RTM does not represent two-way land-river coupling. As such, subsurface runoff affects river routing but not vice versa.

We thank Reviewer 1 for pointing out that clarification is needed here. We agree that RTM does not represent two-way land-river coupling. In this paragraph, we intended to highlight that RTM has been shown to have a time lag between runoff and discharge, not suggest that the delayed seasonality of discharge is contributing to any of the timing offsets in runoff. We will edit this paragraph for clarity in an updated manuscript.

Section 3: Please separate results and discussion. The current structure prevents a cohesive storytelling.

We agree with the reviewer that in many cases it is useful to separate the results and discussion. Because of the number of variables being examined, we found it was clearer to group the results and discussion by variables and skill metrics, rather than presenting them as separate sections.

Section 3.3: Why isn't this metric introduced in the methods?

We thank Reviewer 1 for pointing out that clarification is needed. Section 3.3 discusses relative difference, which is introduced in section 2.4 of the Methods and Data section, starting on line 154. We will clarify this section of the Methods in an updated manuscript.