



Interactive comment on “Assessment of the ParFlow-CLM CONUS 1.0 integrated hydrologic model: Evaluation of hyper-resolution water balance components across the contiguous United States” by Mary M. F. O’Neill et al.

Anonymous Referee #2

Received and published: 18 April 2021

The manuscript presents an evaluation of a continental-scale, hyper-resolution model run of a fully coupled hydrologic model. The presented results may serve as a benchmark for model inter-comparisons and guide future ParFlow model improvements. The manuscript is polished and detailed. This reviewer is certainly impressed by its exhaustiveness. Before the manuscript can be recommended for publication in the as-is form, authors are encouraged to address the following concerns:

a) Argument should be provided to explain why only 30 FluXNet sites were chosen?

Printer-friendly version

Discussion paper

What were the criteria for selecting them (e.g., belonging to different PFTs or hydroclimatic regions)?

b) Throughout the manuscript, subjective qualifiers have been added to describe the ability of the model to replicate observations. These include "appropriately", "exceptionally well", "good performance", "acceptable", "poor", "very good ability", "moderate to strong" etc. However, it is not clear for a (future) model user if the results are indeed appropriate (and for what?), exceptionally good (for what?), or acceptable (for what and when), etc. Clear guidance should be provided for decoding these qualifiers, especially as a single qualifier is sometimes used for a range of variables, even when the accuracy is very different. Should these qualifiers be interpreted only for qualitative inter-comparison across sites or are these describing relative performance w.r.t. the performance of other models? To this reader, subjective qualifiers encourage a biased evaluation.

C) While the presented runoff comparisons are quite exhaustive, (future)users of the model and other readers will benefit if exceedance plots for performance metrics are generated. Such a plot will highlight the fraction of observation locations with performance higher than a threshold. These plots may also be generated for different river basins to highlight the relative performance between them. It will help the regional scale modelers. Finally, as it is mentioned that large errors could have been introduced by anthropogenic impacts such as dams etc., separate performance comparisons in Gages-II watersheds could be performed to help address this point. Similar exceedance plots should be generated for other variables as well, as these plots highlight how many grids show performance better than a given threshold, and thus underscore the usefulness of spatially explicit modeling.

D)The uniqueness of this model is its integrated, fully-coupled nature (and its application at hyper-resolution over continental scale). However, none of the comparisons highlight how usage of this kind of model can provide better estimates in at least some variables at certain locations, over other LSMs. Authors are encouraged to consider

showing some relevant comparisons along these lines to significantly increase the impact of the paper.

E) Conclusions focus on highlighting that the model produces good temporal patterns. What is not clear is if one needs such a complex, fully distributed coupled model to obtain comparable correlations. Most temporal correlations in hydrologic responses are largely driven by the temporality of precipitation/melt and Rn/Temperature/VPD. Unless it can be shown the subsurface-surface flow interactions, a "differentiator" characteristic of this model w.r.t. LSMs, has led to improvement in these temporal correlations, demonstration of the real efficacy of this model (considering its data and computational demands) w.r.t. LSMs remains unclear.

Minor comment:

– Southeastern states are cut out from the figures. Interested (future) users of the model in these states will find it difficult to follow through. – Line 248: "Stave IV" should be "Stage IV". – Line 415: "shear" should be "sheer" – Line 454: Fig. 3.3e should be 3e. – Line 567: Fig. 3.6 should be Fig. 6. – Line 597: Figure 6 should be Figure 7 – Fig. 12a: The color palette does not highlight the heterogeneity well. Please consider alternatives.

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2020-364>, 2020.