Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2022-173-RC2, 2022

Comment on gmd-2022-173

Anonymous Referee #2

Referee comment on "Continental-scale evaluation of a fully distributed coupled land surface and groundwater model, ParFlow-CLM (v3.6.0), over Europe" by Bibi S. Naz et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-173-RC2, 2022

This manuscript is an implementation of the ParFlow-CLM at high resolution (3 km) focused upon the European domain. The validation of the model performance is a wide-ranging analysis based upon remotely sensed soil moisture, and ET, as well as ground-based data products of soil moisture, SWE, ET, groundwater depth, and streamflow. It is generally well written, although there is a lack of focus in the key findings. The authors attributed deviations from observed site level behavior (e.g. positive SM and ET bias) primarily to uncertainties with the incoming atmospheric forcing. However, it seems likely that uncalibrated parameters could have just as easily led to these biases.

The authors motivate the analysis by claiming high spatial resolution combined with a representation of lateral groundwater flow is necessary for improved region wide prediction of hydrological variables. However, this reviewer did not find compelling evidence to demonstrate these assertions from this analysis alone, partly because the model skill was not put in context of other simulations. For example, implementing a coarse version of ParFlow CLM, or a version without lateral ground-flow could have better demonstrated these points.

This manuscript is, in fact, complementary to a similar implementation of ParFlow-CLM for the CONUS domain (O'Neill et al). Yet, the author's do not fully address this point until late in the conclusions, and miss an opportunity to provide a more rigorous comparison between the CONUS and European domain performance with ParFlow-CLM.

It is challenging to evaluate this manuscript because in one sense the methods behind the model implementation and evaluation are useful to the LSM or hydrology community. This validation approach (use of statistics based on comparison to RS and site-based observations) could be used as a template for benchmarking other models. Furthermore, this 'evaluation of a previously published model' does fulfill one of the criteria for publication in GMD. On the other hand, the comparison between the model simulation and remotely-sensed and ground based observations lacked a clear focus. Detailed comments

are below.

Line 21: It is a bit confusing what the authors mean by high resolution hydrological modeling, and large-scale hydrologic processes. Better quantification?

Line 30: LSM's are also used commonly for carbon and nitrogen cycling research. Both LSM's and GHMs solve water balance equations.

Lines 40-50: The author seems to be conflating two things: issues of spatial resolution, or issues related to physical processes. It is true a coarse scale model will not capture fine scale hillslope topography which could be important for watershed scale studies, but is this necessary for global scale climate models?

Line 77: You need to spell out remote-sensing (RS) the first time you use it.

Line 90: What is the difference between Parflow-CLM, PF-CLM and PF-CLM-EU3km?

Line 97: Renaming a model to PF-CLM-EU3km usually means you have changed the model equation/structures/parameterizations. I don' think the author's do that here – it is simply the PF-CLM or Parflow-CLM model run at a certain spatial domain (Europe) and at 3 km resolution. A 'new' model hasn't been designed or developed…..

Section 2.0.2 It is completely unclear what is novel about your implementation of ParFlow-CLM other than the domain and resolution. This seems like a model application and not novel development.

Line 134: Not clear what 'inscribing' into the Eur-11 grid means.

Line 144: CLM3.5 is from the Community Land Model, different than the Common Land Model (CLM) described here within ParFlow-CLM.

Section 2.0.4 It seems unlikely that nine years of spinup would be enough to reach equilibrium between prescribing vegetation conditions and subsurface soil moisture state. Did the author's check that the hydrological variables approached an equilibrium. It is also typically not normal to spinup with a single year (1997), you would want to spinup up overall several years (decade if possible) to capture variation in met forcing.

Line 269: "Because of the explicit lateral groundwater and surface flow representation, we show that the PF-CLM270 EU3km model is able to resolve multi-scale spatial variability in hydrological states and fluxes such as simulated river flow, SM, ET and WTD distributions which are strongly correlated with the river network and topography as shown in Fig. 1."

I am not sure I found any evidence of this causal relationship.

Line 339: "The difference is explained by the shallow groundwater system simulated only by PF-CLM-EU3km, which contributes to the saturation of the deeper soil layers leading to higher soil water content, whereas the standalone CLM3.5 model applies a simple approach to simulate groundwater recharge and discharge processes in a single column and neglects explicit lateral groundwater flow."

It appears here that the authors are attempting a comparison against CLM3.5 (the Community Land Model) which was used as the LSM to develop the ESSMRA product, and comparing against the PF-CLM-EU3km. Claiming the differences in SM can be accounted for by differences in the accounting of lateral groundwater flow. This is a complicated comparison for many reasons, one of them being that the ESSMRA product includes observations of the ESA-CCI 'observations'. The PF-CLM-Eu3km does not. It is not a controlled comparison to claim lateral groundwater flow is the cause for the differences…..

It's also extremely confusing that CLM3.5 (Community Land Model) is not the same as the "CLM" (Common Land Model) in PF-CLM-EU3km.

Figure 4: Not clear what we can hope to learn by comparing 3 separate SM products

against each other. Would it not be more helpful to compare the performance of the SM products against in-situ site ISMN observations? I see that this comparison is pushed to the supplement.

Line 387: "Previous studies of PF-CLM-EU3km also indicate……"

Apparently this exact implementation of this configuration of the CLM ParFlow has been done before? Still failing to see the novelty of the study?

Figure 5: It would be more compelling to show mean seasonal cycles for a sampling of sites (model vs. flux tower ET) across a variety of biomes. Seasonal correlations (as shown) should be strong, just based on phenology of vegetation, as well as increase/decreases in SW radiation. You show regional plots in Figure 6, but running at high resolution grid (3 km) should allow you to make direct comparison to flux tower ET data. It is less compelling to show seasonal variation with GLEAM and GLASS given these are data products.

Line 417: "Figure ??" typos show up a few times in this manuscript.

Line 469: "Our comparison of simulated SWE with observed SWE reveals an overprediction of SWE in the Eastern regions which is more likely to be related to the uncertainties in precipitation."

I don’t follow how the authors came to this conclusion. Could not biases in SWE be a result of uncertainties in temperature, or from issues with the snow/energy balance model which simulates accumulation and depletion of snowpack? If some sort of evaluation against in-situ site atmospheric observations was performed that could provide more credibility.

Line 481: “The rigorous evaluation of the PF-CLM-EU3km model over Europe together with the recent study by O’Neill et al. (2021) which evaluated model performance over CONUS paves the way towards a global application of fully distributed physically-based hydrologic models.”

This is the first time, at the end of the manuscript, where the authors mention this serves as a companion paper to the CONUS implementation of the same model. This manuscript would have been much more compelling if comparison in performance were discussed between the CONUS and EU implementations throughout. Or to quantify the benefit of high resolution implementation of this model, with subsurface, later flow against other LSM’s at coarse resolution, or lacking later, subsurface flow.

Line 483: “The protocol of evaluation metrics and methods presented in this study and in O’Neill et al. (2021) can be used as a framework to benchmark future PF-CLM-EU3km model implementations to further improve model simulations in the areas that have been identified or to explore the impacts of groundwater on 485 simulated hydrological states and fluxes by comparing with other existing global land surface model applications.”

Again, it would be more compelling if this manuscript performed a direct comparison of performance against the CONUS implementation or existing global land surface model applications to demonstrate improved utility/skill.