Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2021-38-RC2, 2021

Comment on gmd-2021-38

Anonymous Referee #2

Referee comment on "Coupling the Community Land Model version 5.0 to the parallel data assimilation framework PDAF: description and applications" by Lukas Strebel et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-38-RC2, 2021

The study presents the development of PDAF coupling with CLM version 5.0 for data assimilation. Further, the authors present a 10-year sensitivity simulation, using a single column model over a forested catchment in Germany. The simulation includes data assimilation (DA) with and without parameter updates. Compared to open loop runs, which exhibit a wet bias, the DA could improve the soil water content (especially at 5 cm and 20 cm) compared to deeper layer (50 cm).

This study provides additionally new DA capability to the larger CLM scientific community and also shows its potential for improving the model states by inclusion of joint parameter updates. However, there are several shortcomings in the present manuscript which needs to be addressed before being suitable for publication in GMD.

Major Comments

The motivation for this study is weak. The authors briefly mention about the difference between online and offline DA (Ln 55), but they need to better motivate the coupling CLM5.0 with PDAF. Is it more for the standalone DA with CLM5.0 or for CLM5.0 within the TSMP framework? What new does PDAF bring? How does it reduce the number of core-hours or computation time compared to other offline DA? And, how it scales with increase in domain size and time period of simulation? This needs to be discussed clearly.

Kurtz et al. (2016) already presented the PDAF coupling to TSMP including CLM3.5. So, what is new in this study? I assume that there must be substantial work involved in developing the PDAF interface around CLM5.0 which has different software environment compared to earlier versions of CLM (e.g. CLM3.5). But it is not so clear in the current version of the manuscript.

Ln 85: This comes so suddenly. The authors need to provide better motivation to use single column model.

The literature review is another weak part of the manuscript. The authors make no effort in presenting their results in context of previous findings. Also, does the improvement in soil moisture also improves the surface energy fluxes. For LSMs, improvements need to be explored soil states as well as fluxes. And, a discussion section is missing.

There is no README file or User manual to reproduce the results presented in this study, also please provide a web URL for Zenodo and cite this paper in the References. The upload should also include scripts for processing the figures and observation data for reproducibility.

Minor Comments

- Ln 10: Even tuned second generation LSMs can be "accurate", here maybe the authors want to imply that third generation LSMs better represent the key physical processes. Also, check in the rest of the manuscript.
- Ln 11: more? What type of data?
- Ln 15: Is this further development of PDAF or addition of new interface to connect PDAF with new models?
- Ln 34: common might not be the right word here.
- Ln 48-53: This paragraph needs to be rephrased (framework, external framework, within framework). It has just too many frameworks.
- Ln 70: PDAF with joint state parameter update for CLM was also used in the following study:

Shrestha, P., W. Kurtz, G. Vogel, J.-P. Schulz, M. Sulis, H.-J. Hendricks Franssen, S. Kollet and C. Simmer (2018), Connection Between Root Zone Soil Moisture and Surface Energy Flux Partitioning Using Modeling, Observations, and Data Assimilation for a Temperate Grassland Site in Germany. JGR-Biogeosciences doi: 10.1029/2016JG003753

- Ln 73: "In this study, we present the coupling of .."
- Ln 93: Rephrase. "The paper ends with " is not appropriate.
- Ln 116: 1) variation methods, …2) sequential methods
- Ln 125: Perturbation vector missing in Eq. 1, where y is generally the observation vector. It is discussed much later in Ln 146. What is the measurement error?
- Section 2.3: There is always a discussion about older version, maybe the authors should discuss it before, and present their new formulation, rather than interchanging now and then. Maybe this would also highlight, what new work has been done.
- Ln 181: The "Figure 1" is not helpful, either improve or remove. Also, rephrase and elaborate the discussion.

- Ln 191: What is “CIME”?
- Ln 204: Maybe “clipping” ?
- Ln 218: Rephrase.
- Ln 232: in Wüstebach , and Belgium ?
- Ln 252: Explain the SWC unit.
- Ln 305: “to overestimate SWC” or “wet bias in SWC”
- Ln 333: What is variant here?
- Figures: Add subplot numbers (e.g., a), b))
- Figure 2: “In the diagram NMLST means namelist, SIM means simulation process, HIST means history file output, PID means PDAF identification number.” – this should as legend in Figure.
- Figure 3 caption: red (solid line), light green (dotted line).