Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2021-38-RC1, 2021

Comment on gmd-2021-38

Anonymous Referee #1

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-RC1, 2021

General Comments:

This manuscript coupled the CLM version 5.0 to a data assimilation framework (PDAF) to improve the simulation of soil moisture content at a forested site in Germany. The authors ran a series of assimilation experiments in which they adjust simulated soil water content (SWC) with observations of soil water at depths of 5, 20 and 50 cm. They ran additional simulations where they also adjusted soil sand/clay fraction, and finally organic matter content. They found that all assimilation runs performed better in terms of simulated soil moisture as compared to the (open loop) simulation. The assimilations in which both soil moisture and soil parameters were adjusted performed the best. The influence of water upon land surface models is well established and has an important influence upon surface energy balance, temperature, and vegetation (carbon) behavior. Thus, the use of soil moisture observations to improve land surface simulations in terms of updating both model states and parameters is a worthwhile pursuit.

With that being said, this reviewer found there were significant gaps in the cited literature and motivation for the work (See details below). The authors also did not explore or provide an explanation for how the vegetation behavior influenced the water behavior and contributed to soil moisture behavior and what implications this has for the carbon cycle. The reviewer also felt more justification was needed for why soil characteristics (clay/sand/organic matter) were adjusted in place of true hydraulic parameters within the model.

Scientific/Detailed Comments:

Line 16: Not clear to me how this coupling is ‘novel’. The Data Assimilation Research Testbed (DART) uses similar ensemble capability, and couples an EnKF to CLM5. (Raczka et al., 2021 https://doi.org/10.1029/2020MS002421).

Line 28: Seems like a dated citation (Overgaard et al., 2006)– perhaps reference CMIP5 or CMIP6 manuscripts that compare a range of LSMs performance (e.g. Arora et al., 2020; https://bg.copernicus.org/articles/17/4173/2020/)

Lines 34-36: Need to improve references to water limited regions and the work that has been done to improve water limitation and its connection to the carbon cycle. (e.g. Raczka et al., 2021; Weider et al., 2017; Kennedy et al., 2019) https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016JG003704).

Line 35-40: In general, citations provided here seem rather general, and not focused on particular research topic, which in this case is SWC, hydrology and impact upon latent and sensible heat. You might want to focus more directly on the representation of hydrology within CLM (Swenson et al., 2019; https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019MS001833), and why it suits DA for your application. Including specific advances in hydrology with CLM5.0 (Kennedy et al., 2019; https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018MS001500) might also better motivate this work.

Line 45: Need more background into remote sensing products of soil moisture. There are a growing set of remotely-sensed soil moisture observations that should be referenced here – SMOS, SMAP, ESA-CCI. There are many emerging products. You should better motivate the use of DA precisely because the range of remotely-sensed products is expanding. Also the purpose of DA (especially EnKF) is that unobserved states (subsurface layers) can be adjusted based upon the model state covariance matrix of the modeling system.

Line 47: Awkward sentence, It is common practice….

Line 53: It is unclear what distinction is being made between ‘offline’ vs ‘online’ coupling in data assimilation frameworks. The authors state that ‘offline-coupled’ data assimilation is used for the Data Assimilation Research Testbed (DART) https://dart.ucar.edu/ (Anderson et al., 2009). Furthermore, in offline coupling ‘the framework wraps around the model and does not modify the model’. This is not true. Within DART – the state of the model is modified during the update step of the EnKF. Therefore, the assimilation updates the model state in time so that the trajectory of the model more closely matches the observations being assimilated. Furthermore, DART and CLM are interactive in that DART updates the model state within CLM, and the inflation parameters within DART (Gharamti et al., 2019; doi.org/10.1175/MWR-D-20-0101.1) are also updated in time and influence the ensemble spread of the CLM model state. More recent updates to the CLM code have included model components that call data assimilation components from DART directly. In applications outside of CLM, DART has been used to modify parameters (e.g. Zhang et al., 2021; doi.org/10.5194/tc-15-1277-2021.) within models as well. The authors need to reconsider their assertion that their DA coupling approach is ‘novel’.

Line 57: DART is commonly used with all components of the earth system within CESM including land (CLM), atmosphere (CAM), ocean (POP), and sea/land ice, as well as many other earth system models. See https://dart.ucar.edu/publications/

Line 84: "… [these manuscripts] concluded that the consideration of heterogeneous porosities can increase model performance depending on the model structure. In contrast to these detailed distributed catchment studies, we model the study site from the viewpoint of a larger regional model where the catchment is represented by a single grid cell."

The previous modeling studies suggest that including a description of heterogeneous soil porosities will help model performance in a fine-scale catchment. Presumably a fine spatial scale description is needed to represent a catchment. Therefore it caught this reviewer off-guard that the authors propose to use a coarse, grid cell to represent catchment behavior (see Swenson et al., 2019). Perhaps provide motivation that a DA framework can improve modelling behavior through correcting for known biases in the system – or known errors in parameters.

Method section 2.1

"Furthermore, we investigate whether updating of the soil organic matter parameter via data assimilation can further improve the prediction of soil water with CLM5."

Given your manuscript goals you need to provide some explanation within the CLM methods section of how soil organic matter influences soil water drainage. It is also slightly unclear what benefits either updating to the CLM 5 description or using the PDAF will bring to this analysis, be a bit more specific. Some of this information is included in the appendix (A1-A4), but a bit more explanation within the main text would be helpful.

May also want to mention in the methods of CLM 5 – that it updates the plant hydraulic stress representation (Kennedy et al., 2019) thereby influencing water-carbon coupling, and transpiration. The authors do not really discuss the influence of vegetation (water-carbon coupling) upon their SWC results.

Section 2.2.1

"For example, ensemble members can be generated based on perturbed soil parameters and atmospheric forcings."

Not clear at this point how ensemble is generated for this experiment.

"The state vector ð□□± i contains soil water content (model states), sand and clay fractions (parameters), and organic matter fractions (parameters) depending on the experiment as described in Section 3.3."

It makes sense here to describe how the CLM soil column is constructed (i.e. PFTs, columns, layers etc) within Section 2.1. Are you updating all soil layers of the CLM model for SWC?

Section 2.3:

"Furthermore, for the optional parameter updating it is necessary to provide a function to transform the input parameters, e.g. soil texture, to the model parameters, e.g. the soil hydraulic parameters. CLM5 performs this transformation once during initialization to obtain the hydraulic parameters from the soil texture in the surface file."

It was a bit confusing to this reviewer that the authors were referring to the soil characteristics such as clay/sand/organic matter as 'parameters'. In general, parameters refer to numeric coefficients that influence model equations. This manuscript adjusts the soil characteristics to indirectly adjust the hydraulic parameters (A1-A4). In general, it seems parameter optimization should be limited to parameters which are difficult/impossible to measure. The soil characteristics, on the other hand, could be measured given how well the study site (watershed) seems to be observed already.

3.1 Study Site:

Very unclear how the CLM site-level or gridded simulation was setup. What was the size of the grid cell used in which the soil characteristics / topography were defined? How was this forested site initialized? Was it spun-up from near ground conditions or was a present-day compset used within CLM?

Section : 3.2.1

“The filtered raw data is then spatially and temporally averaged to fit the requirements of the model, i.e., daily averages for the three soil depths.”

I don’t think that’s a limitation or requirement of the model – CLM5.0 can be run on an hourly time basis thus assimilation could be performed hourly. Also there are roughly 25 subsurface potential soil layers in CLM, so it could potentially handle more soil depth observations depending upon the depth of the soil column at this location. I think you performed daily averages of all the soil observation locations to simplify the assimilation process, which is reasonable.

So you averaged all the forested (undisturbed) soil water observation locations into a single value for each depth?

Line 261: Lateral flows are not represented at all in CLM5 – no grid cell to grid cell communication. Surface and subsurface drainage is routed directly to rivers.

Line 287: Be more specific here: Perturbed inputs of *both* atmospheric forcing and soil characteristics of soil/clay and organic matter? What was the purpose of perturbing both? Could you use only atmospheric perturbations if the goal was to only assimilate SWC observations? The additional perturbation of the soil/clay, organic matter was necessary for the parameter updates? Provide a bit more explanation.

Line 287: Do you state anywhere what soil water variable in CLM you are adjusting? I assume it is the prognostic variable H2OSOI_LIQ, but there is also H2OSOI_ICE and the diagnostic variable H2OSOI. Also you are adjusting all vertical layers?

Figure 1: Any physical explanation of why the model would overestimate SWC at shallow depth (5 cm) and at the deepest layer (50 cm), but overestimate SWC at the middle depth (20 cm)? Curious of whether this could be related to the observational uncertainty

of the SWC sensor – and what was used as the observation uncertainty? Also wondering if this behavior was related to the configuration of the root profile within CLM – how much of the root mass was within this layer and therefore what influence this had upon transpiration and removal of water within this soil layer?

This opens up other questions of what the forest state was for your model simulations including things like biomass and leaf area index from the site observations. Were these reasonable? Did you look at the simulated transpiration, evapotranspiration and GPP to determine if these values seemed reasonable? I don't think you had flux tower observations available to check, but perhaps you could infer reasonable values from surrounding sites. The vegetation state will have an important impact of subsurface soil moisture state and to what effect this impacted your simulation is unclear. The vegetation state, including how it was initialized and how it was simulated (other than the PFT setting) was not discussed in this manuscript.

Table 3: It was not completely clear until I viewed this table that the model 'parameters' that were being adjusted within the assimilation were actually the soil characteristics of clay/sand and organic matter. The term 'parameter' is admittedly loosely defined in modeling applications, but in general, this typically refers to 'coefficient' values within the model code that (within a model like CLM) are specific for particular plant functional types. The surface characteristics of the soil, however, are typically prescribed and held constant. The reviewer recognizes that this manuscript is, in part, is a demonstration of the capabilities of the assimilation system, and is apparently following the approach taken in (Naz et al., 2019) but physically, does it make sense to adjust the soil characteristics (generally fixed in time) such that they change with time? Would it not make more sense to adjust the numeric coefficients in equations A1-A4 instead of %sand and %clay? The authors acknowledge this at the very end of the conclusion section, but perhaps more justification could be provided earlier on in the manuscript.

If there were many soil moisture subsurface observations, were any soil characteristic observations available to check the posterior values of the soil characteristics?