

Geosci. Model Dev. Discuss., author comment AC1
<https://doi.org/10.5194/gmd-2021-38-AC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Lukas Strebel et al.

Author 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-AC1>, 2021

Scientific/Detailed Comments: Referee comments in bold and author answer non-bold.

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>).

Our intention of using the term 'novel' in this sentence was to highlight specifically the new coupling of CLM5 with PDAF in contrast to existing coupling of CLM5 to other data assimilation frameworks or previous versions of CLM to PDAF. We will rephrase the section to make this clearer.

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/>)

We will update the dated citation with a more recent citation that highlights the complexity and range of current LSMs.

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>).

We agree and will add suggested references as well as other references more focused on soil water content.

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.

It is true that the references here are rather general. We chose them to highlight the application of CLM to single-point setups for a wide array of studies. We will rewrite this section to make this clearer focusing more on soil moisture and DA as you suggest.

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.

The emerging remotely-sensed soil moisture observations products are important and useful in combination with DA. We did not include many references to remotely-sensed products because our application used only in-situ measurements. However, we agree that remotely-sensed products are a good way to motivate DA applications and we will add this to the section.

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

Yes, we will rephrase the sentence.

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'.

We acknowledge that the distinction between 'offline' vs 'online' is not clear. We use the definition of these terms in this context from PDAF as formulated for example in Kurtz et al. 2016. The distinction is about using the main memory or restart files for the transfer from model states to the DA framework and vice versa. We will correct the 'the framework wraps around the model...' sentence. We did not mean to imply that offline coupling does not affect model states or that DART and CLM are not interactive but rather that in offline coupling it is often not necessary to modify the model source code. We were not aware of the recent updates to CLM that includes calls to DART directly since even recent publications like Raczka et al., 2021 and Zhang et al., 2018 still mention the use of restart files for the exchange between the model and DART. We will correct the comparison to DART in this section to include your corrections. We did not intend to assert that the DA coupling approach is novel, we specifically mention that we mostly re-use and modify existing software infrastructure, we just meant that our specific implementation of coupling CLM5 and PDAF 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/>

We will modify the sentence and mention the use of DART in other CESM components and other earth system models.

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.

Yes, representing a catchment in detail requires a fine spatial scale simulation, however in many applications it is not computationally feasible to represent catchments on such fine scales. Therefore, we think it is reasonable to demonstrate the application of a new DA coupling on the coarse scale of many applications but simplified to a single grid cell to highlight the direct effects of DA in the grid cell where observations are available. We agree that we can better motivate the DA framework and we will rephrase this section.

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.

We will add more details about the CLM5 soil organic matter parameter and its relation to SWC into the main text and include the relations to the plant hydraulic stress representation in CLM5.

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 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?

We will add a reference to section 3.2.2 and 3.2.3 for the specifics of ensemble generation for this specific experiment. We can add details of the CLM soil column structure as you suggest. It is correct that we are updating the SWC of all soil layers of the CLM model and will clarify this in the revised version.

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.

We refer to the soil characteristics as parameters because we treat them as parameters in the joint state and parameter approach. As we already describe in the outlook, we agree that the parameter optimization should be applied to the hydraulic parameters directly. Using the soil characteristics as indirect parameters to be updated has been done in various previous studies and was therefore the baseline implementation. Nevertheless, we will mention in the revised version that the soil hydraulic parameters will be updated directly in future works.

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?

The grid cell size used was roughly 3km by 3km. The model was initialized from a cold start and spun up according to the CLM5 documentation. We will add more details about this in the revised version.

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?

You are right, it is not a requirement or limitation of the model. We wanted to say that these are the requirements we defined for this model setup. We will correct this sentence accordingly. Yes, we averaged the large number of forested soil water observation locations into a single value for each depth (at 5, 20 and 50 cm, the three depths for which observations were available) and will specify this in the revision.

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.

Yes, we just meant that lateral flow in the form of runoff to rivers is represented in CLM5. We will clarify this.

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.

Yes, we used perturbed inputs of both atmospheric forcings and soil characteristics. We perturb both because the perturbation determines the ensemble spread and the ensemble spread represents the model uncertainty. Model uncertainty exists for both atmospheric forcings and the model parameters i.e. the soil hydraulic parameters indirectly determined through the soil characteristics. Even in studies without parameter updates and only SWC observations soil characteristics are often perturbed. We will add these explanations in the revision.

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?

We do not specifically mention the CLM5 variable names. We use the diagnostic variable H2OSOI as a state variable and then adjust both prognostic H2OSOI_LIQ and H2OSOI_ICE variables for all vertical layers. We will make this clearer in the revision.

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?

We currently do not have a physical explanation of this model behavior. For simplicity, the observational uncertainty is assumed be constant and set to a RMS of 2%. We did not explore the effect of the root profile on this behavior. We will discuss this in the revision.

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.

We looked at LAI and evapotranspiration and they are reasonable. For ET even close to flux tower observations. We did not include a figure for ET because the effects of SWC DA

for this specific site were not significant and more analysis is required. We are currently in the process of performing more simulations for different sites and analyzing the effects of SWC DA on ET and other variables in a further study. For this study, we mainly wanted to demonstrate the coupling of PDAF to CLM5 and the direct and clear impact of SWC DA for a simple setup. We will discuss this in the revision.

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?

As mentioned in the previous question, we will be more careful with the terminology "soil characteristics" and "soil parameters". We used the term 'parameter' more generally as a distinction to state variables rather than to differentiate between coefficients and prescribed constants. We fully agree that it makes more sense to adjust the coefficients in A1-A4 directly. The indirect approach using the soil characteristics was an established approach, but in future work we will adjust the numerical coefficients directly. As you suggest, we will mention this earlier in the manuscript and not just in the conclusions. There are soil characteristic observations at different locations in the catchment, but it is not a simple task to 'average' these discrete spatially distributed observations to compare it to the posterior soil characteristic value that represents the whole catchment / grid cell. We will discuss this in the revision.