Geosci. Model Dev. Discuss., 8, C3813–C3825, 2016
www.geosci-model-dev-discuss.net/8/C3813/2016/

Geoscientific Model Development Discussions

Interactive comment on "TerrSysMP-PDAF (version 1.0): a modular high-performance data assimilation framework for an integrated land surface–subsurface model" *by* W. Kurtz et al.

W. Kurtz et al.

w.kurtz@fz-juelich.de

Received and published: 26 January 2016

General Comments
The paper "TerrSysMP-PDAF (version 1.0): A modular high-performance data assimilation framework for an integrated land surface–subsurface model" by W. Kurtz, G. He, S. Kollet, R. Maxwell, H. Vereecken, and H.-J. Hendricks Franssen, presents efforts to couple the TerrSysMP model to the PDAF assimilation library. Overall, the paper is clear and provides an adequate level of description of the new system. This paper will be an important future reference for upcoming assimilation studies performed using this system.

Reply: We wish to thank the reviewer for pointing out the clarity, contribution and novelty of our work.

In the introduction, I'd like to see comparisons with other similar modelling systems (coupled Atmosphere, Land Surface, and Groundwater) capable of assimilation. How does the TerrSysMP-PDAF compare with other assimilation capable systems? Ultimately, to answer the question: what makes the TerrSysMP-PDAF unique?

Reply: One major difference between TerrSysMP-PDAF and other already established data assimilation systems is that TerrSysMP-PDAF includes a physically-based integrated groundwater model (ParFlow) which solves the 3D-Richards equation and coupled surface-subsurface water movement. Most other data assimilation systems (e.g., CaLDAS, GLDAS, WRFDA) usually rely on a simplified representation of the surface water/ groundwater component inherent in land surface models. In these land surface models, 1) water movement is usually only calculated in vertical direction (no lateral flow), 2) surface water and subsurface flow are not coupled and 3) the influence of lateral groundwater flow is not explicitly taken into account.
An exception is the data assimilation system for the physically-based integrated groundwater model MIKE-SHE (*Ridler et al., 2014*) which has similar process descriptions as in TerrSysMP. A difference here is how the data assimilation module is coupled to the flow model. For MIKE-SHE, the OpenMI framework is used to couple the model to the data assimilation utilities of OpenDA. In TerrSysMP-PDAF, the data assimilation algorithms of PDAF and the models of TerrSysMP are coupled in a more integrated fashion, which is more efficient especially for large data assimilation problems.
Another advantage of TerrSysMP-PDAF is its modularity. With TerrSysMP-PDAF it is possible to perform data assimilation with different (combinations of) component models depending on the needs of the actual data assimilation

problem. This allows the user to be more flexible with respect to the process description (i.e., the utilized forward models) and the computational demand of the data assimilation problem.

Furthermore, we would like to stress that TerrSysMP-PDAF is also capable of performing joint state-parameter estimation. Although there is a growing body of literature on the utilization of data assimilation techniques to estimate parameters in land surface models and especially in groundwater models, this functionality is, to our knowledge, still not found in any published data assimilation system. However, we deem parameter estimation important especially for land surface and subsurface models because a large part of model uncertainty is related to the parametrization in these kinds of models.

We will add a paragraph in the introduction that discusses the above mentioned differences to other data assimilation frameworks and that highlights the novelties of our approach.

References:

Ridler, M.E., van Velzen, N., Hummel, S., Sandholt, I., Falk, A.K., Heemink, A., Madsen, H. (2014), Environ. Modell. Softw., 57, 76-89, doi:10.1016/j.envsoft.2014.02.008

The writing style - although understandable can be clumsy to read due to long sentences without commas. I've listed a few cases below. A word processor (like MS Word) will pick up most of these run-on sentences. Overall, I found the paper to be worth publication once the comments above and below are taken into account.

Reply: Thank you for pointing this out. We will correct the cases you mentioned. We will also make every effort to revise the manuscript for clarity throughout.

Specific Comments p.9632 ln.4 –> It's unclear where does the observation to model

index mapping fit into the system? Is the observations handled by the PDAF and assumed to be fixed in location with known model indices? More explanation is required here.

Reply: We agree that the mapping of observations onto the model grid was not very clear from our description. Basically, TerrSysMP-PDAF reads in a separate observation file (netCDF format) at each assimilation cycle. This observation file only contains information on the currently available observations. This information includes the measurement values, the estimated measurement errors (optional) and the location of the observations in the model grid. The latter information is given in the form of model grid cell indices. This means that the user needs to determine the grid cells which match the observation locations beforehand. The grid cell indices provided in the observation files are then handed over to PDAF, which will use these indices to extract the simulation results at observation locations from the state vector. We will clarify this point in the description of the observation mapping in the revised manuscript.

p.9632 ln.13 –> Why is the TerrSysMP restricted to the EnKF? Will it not work with a deterministic (non-localized) filter? Localization is often useful (essential in some cases) in assimilation applications. Is that not implemented?

Reply: In principle, the PDAF library includes various filter variants (e.g., EnKF, ETKF, SEIK, LETKF). For the coupling of PDAF and TerrSysMP (or any other kind of model) it is necessary to provide several user-specified functions in order to define the data assimilation problem for PDAF (e.g., definition of the state vector, definition of the observations, etc.). Some of these user-specified functions are shared among different filter variants (e.g., reading of observations) and others are specific for certain kinds of filters (e.g., specification of grid cell coordinates for local filter variants).

Currently, the main focus of our research is to perform data assimilation with TerrSysMP at the catchment scale using the Ensemble Kalman Filter. As such, TerrSysMP-PDAF is currently configured to work with EnKF, however we are also intending to test deterministic filters for this kind of data assimilation problem in the near future.
We also agree that localization is useful or even mandatory for certain assimilation problems. For example, we are planning to assimilate remote sensing soil moisture data with TerrSysMP-PDAF as one of the next steps. This will make it necessary to use filter variants with localization (e.g., LETKF).
You will surely understand that systems like TerrSysMP-PDAF are always a work in progress and the development is usually driven by the needs of the actual users. Our goal is to provide the full functionality of PDAF with respect to filter variants and filter options in the near future. But for now, the system is still restricted to the Ensemble Kalman Filter.

Figure 8 shows four time series plots of Soil moisture at four locations. Are these four locations where there are observations. It's not very informative to validate using the same measurements that were used in the assimilation. Could you not find 4 points distal (or in between the assimilated measurements) for validation?

Reply: You are right. Fig. 9 shows predicted soil moisture values at four observation locations. In order to make this comparison more objective, we also used four independent points for verification which are in between the observation points (x=2500m, y=1500/2500/3500/4500m). However, the conclusions for these different verification points (see Fig. 1 below) are quite similar to the ones from the previous figure. We will show the results from the independent verification points in the revised manuscript.

p.9643 ln. 20 It is surprising that there is such little change in the energy fluxes after

assimilation. Especially in the grassland area (right) as this regions seems to be where the impact of the assimilation is greatest for both corrections of soil moisture and the hydraulic conductivity. I'm not convinced that the minimal effect is solely due to the high groundwater (as you say) especially for a plant type with short roots.

Reply: We already mentioned in the manuscript that the errors of land surface fluxes were relatively low even for the open loop runs (p.9643 line 15ff). The total AAE values (averaged over all grid cells and time steps) for sensible (latent) heat were 1.003 (1.212) W m^{-2} in the open-loop run. Two possible reasons for these relatively low errors have been nicely pointed out by reviewer 2: (1) Plant physiological parameters were not perturbed (2) Meteorological forcings that affect stomatal conductance were not perturbed. Results from these open-loop simulations show that in our setup the variability of soil moisture content only had a limited effect on land surface fluxes and most of the model domain is not affected by water limitation.
With the assimilation of soil moisture contents, the total AAE values of sensible (latent) heat fluxes were reduced to 0.730 (0.876) W m^{-2}. In relative terms, the assimilation of soil moisture data improved land surface fluxes by about 27% but still the absolute values are relatively low (1 W m^{-2} on average). If one looks at the patterns of land surface flux errors, one can indeed see that most of the improvement by data assimilation can be found in the grassland area.
We will add an additional paragraph to this section where the above mentioned reasons for the low absolute improvement of land surface fluxes are discussed in more detail.

The scaling factor and processor time is measured using two cases (identical ensemble and a more realistic case). It would be interesting to see results (or at least the author's thoughts) on the impact of using a model with a varying topography and thus overland flow calculations on the scaling factor.

Reply: We think that our section on the scaling of TerrSysMP-PDAF is already quite detailed. The comparison of the 'ideal' (identical ensemble members) and the 'heterogeneous' (different ensemble members) test case nicely illustrated the effect of ensemble perturbations on the scaling behaviour of TerrSysMP-PDAF. Even for this setup without overland flow, the results very much depend on the model specification and the ensemble perturbations. So this scaling test is not meant to be a benchmark for the system, but it's rather an illustrative example how the scaling could typically behave under realistic conditions.
Overland flow would, of course, add some additional complexity to the system and then also has the potential to alter the results of this scaling exercise. However, this again is highly case sensitive because the temporal evolution of overland flow (and the corresponding convergence properties) depends on factors like model geometry, model discretization and model parametrization. Our experience is that especially the onset and offset of overland flow at particular grid cells (due to heavy rainfall or recession events) can be computationally very demanding. This is due to the fact that overland flow has much faster timescales than the water movement in the subsurface which can lead to convergence issues at the onset/offset of coupling of the two domains. Such events can have a negative influence on the scaling behaviour of the deterministic forward model because the computationally demanding onset/offset of overland flow only occurs in certain parts of the model domain and the utilization of more CPU's only has a limited effect on speeding-up convergence in these particular areas (see also *Kollet & Maxwell (2006)* and *Osei-Kuffuor et al. (2014)*).
A negative influence of overland flow on the scaling behaviour of the data assimilation system could be expected when only certain realisations experience convergence problems due to the onset/offset of overland flow (e.g., through the chosen perturbations of subsurface parameters or meteorological forcings). In this case, a few realisations might dominate the total runtime of the system which then has a negative impact on the overall scaling of the data assimilation

system. However, such problems are not only restricted to overland flow coupling but could also occur, e.g., when there are strong spatial gradients in model parameters (e.g., hydraulic conductivities) which may pose a problem to the numerical solver.
Generally, the scaling behaviour of ensemble-based data assimilation systems will depend on the numerical robustness of the deterministic forward model towards ensemble perturbations. Therefore, it is important to configure the deterministic forward model well with respect to numerical stability and execution time. This can be achieved, e.g., through the correct choice of solver parameters, an adequate spatio-temporal discretization of the problem and a proper choice of model parameters and ensemble perturbations.
We expect that for a well configured overland flow problem, the scaling results might be quite similar to the 'heterogeneous' example without overland flow. In cases where overland flow related convergence problems occur for all ensemble members simultaneously (e.g., for heavy precipitation events) the scaling properties of the data assimilation system would also not be affected. The only case where overland flow would negatively influence the scalability of TerrSysMP-PDAF is when the ensemble perturbation affects the overland flow convergence for a limited amount of realizations. We will add a paragraph to Sect. 5.2 that discusses the potential effects of overland flow on the scaling results.

References:
Kollet, S.J., Maxwell, R.M. (2006), Adv. Water Resour., 29, 945-958, doi: 10.1016/j.advwatres.2005.08.006
Osei-Kuffuor, D., Maxwell, R.M., Woodward, C.S. (2014), Adv. Water Resour., 74, 185-195, doi: 10.1016/j.advwatres.2014.09.006

Technical Corrections

Assess instead of access (3 times at least) In abstract, p9638 ln 1, and p 9638 ln 21

Reply: We will replace the occurrences of 'access' with 'assess' throughout the manuscript.

p9619 l.18 ... in the model input, which allows to...

Reply: Will be changed.

p9619 l.21 ... model states, which are ...

Reply: Will be changed.

p9619 l.24 . terrestrial system, including

Reply: Will be changed.

p9620 l.9 Typically in these cases, point ...

Reply: Will be changed.

p9622 l.10 integrated view of the terrestrial

Reply: Will be changed.

p9622 l.23 are performed at a high

Reply: Will be changed.

p9622 l.24 realisations are needed ...

Reply: Will be changed.

p9622 l.25 There are a number of

Reply: Will be changed.

p9624 l.9 only in a vertical

Reply: Will be changed.

p9624 l.10 ie. there is no lateral.

Reply: Will be changed.

p9624 l.27 allows it to solve

Reply: Will be changed.

*p9626 l.8 allows *for an integrated view *of the*

Reply: Will be changed.

p9626 l.12 Awkward sentence –> another important point that is noteworthy is that ....

Reply: Will be changed to: *"Another interesting feature of TerrSysMP is its modularity: apart from ..."*

p9626 l.16 has already shown to be highly scalable on the massively...

Reply: Will be changed.

p9629 l.12 remove the ":" and replace with ".", then First.. the Secondly and Thirdly in the paragraph should be Second and Third .

Reply: Will be changed.

p9634 l.2 CLM except for the special ...

Reply: Will be changed.

p9635 l.2 ten layers, pressure, ...

Reply: Will be changed.

p9636 l.8 DWD hasn't been introduced yet. What is it?

Reply: DWD is an acronym for the German Weather Service. We will add this abbreviation to p.9635 line12: *"...are taken from reanalysis data of the German Weather Service (DWD) for the year 2013..."*

p.9638 assessed instead of accessed.

Reply: Will be changed.

Interactive comment on Geosci. Model Dev. Discuss., 8, 9617, 2015.

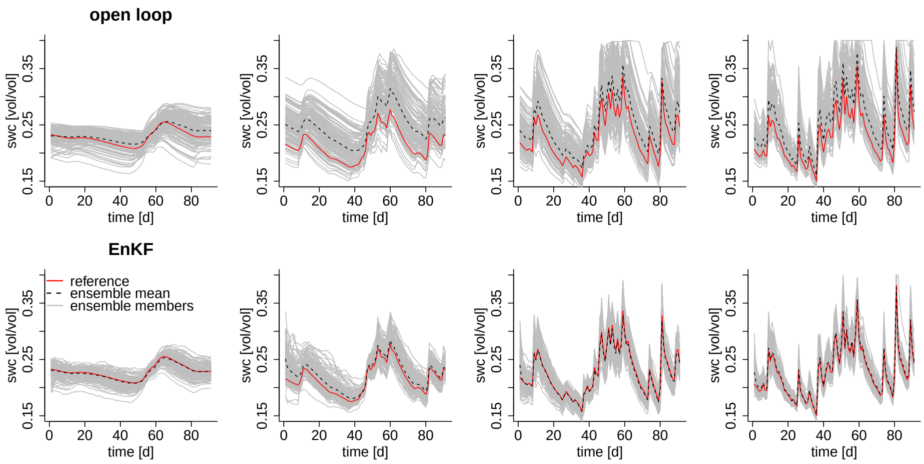

Fig. 1. Simulated soil water content at four verification nodes (x=2500m, y=1500/2500/3500/4500m, from north to south) for April–June 2013. Upper row: open-loop simulations; lower row: assimilation.