Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-296-AC1, 2016

Hydrology and Earth System Sciences Discussions
Open Access
EGU

Interactive comment on "EnKF with closed-eye period – towards a consistent aggregation of information in soil hydrology" *by* H. H. Bauser et al.

H. H. Bauser et al.

hannes.bauser@iup.uni-heidelberg.de

Received and published: 13 September 2016

Summary: *the manuscript presents a study on parameter estimation for an unsaturated zone model. The authors use real soil water content data to estimate parameters and forcings for a 4 layer model using the EnKF. A full iteration of the filter, divided into different steps is introduced to improve the estimation process. During a rain event it is observed that the estimation gets flawed by the models inability to model preferential flow and the observations corresponding to this rain event are excluded from the parameter estimation. This is called a closed-eye period and it is observed that this improves the stability of the estimated parameters.*

General comments: *The manuscript is interesting and is filling a missing space in*

the literatur as being a reasonably controlled test case using real data. However, the manuscript is sometimes difficult to overview and the proposed method feels very ad hoc and leaving a number of open questions. Suggestion is major revision.
(throughout the text references to the manuscript are given as: P9,L78 = page 9 Line 78)

Reply: We thank the reviewer for the constructive comments and suggestions. We have revised our manuscript (see supplement) taking them into account. For consistency with your references, we will still refer to the unrevised manuscript as well.

Major concerns

1. Data assimilation vs calibration

a. The manuscript introduces the concept of a fully iterative filter, which is an approach that I have never seen in the literature before (and obviously the authors neither since there are no citations on this part). Iterative filters are not uncommon, but they are usually local and not global, or restart versions where the model is restarted and ran until next observations. Some relevant references and comparisons to other iterative approaches would be needed as well as a good motivation to the current approach. The authors approach also shows clear similarities to the precalibration exercise by Hubert et al 2010, to the iterative Kalman Ensemble generator by Nowak 2009 (see bottom for references) and to the SODA by Vrugt et al 2005 (already cited, but not compared to the suggested approach).

Reply: Yes, we have not seen the fully iterative approach in the literature as well. We improved the introduction of the approach (page 9, line 12-15) to: "We iterate the whole EnKF scheme and start the next iteration with the final estimation of soil parameters, Miller scaling factors and upper boundary condition of the previous iteration. These iterations differ from the typically applied iterative EnKFs like the Restart EnKF or Confirming EnKF (e.g. Song et al., 2014) and rather resemble the iterations in the Kalman ensemble generator by Nowak (2009), who used a modified EnKF to estimate parameters only. The full iterations applied are required to estimate

constant augmented state components even with a small damping factor and a rather short time period of data. This is especially important for the upper boundary condition, where the time periods of constant values are short."
We have not been aware of the works by Huber et al. (2011) and Nowak (2009). Thank you, for pointing them out to us. We have incorporated Nowak (2009) in the text shown above (page 9, line 12-15) and Huber et al. (2011) in the description of the prior estimation (page 10, line 14): "Highly uncertain properties can exacerbate the performance of the EnKF. For example Huber et al. (2011) showed that the state estimation with an EnKF is superior in case of properly precalibrated model parameters. In our case, there is no prior knowledge about small scale heterogeneity, which can lead to difficulties in the estimation process:"
The SODA approach by Vrugt et al. (2005) is comparable to the Standard EnKF approach in our manuscript, since he as well estimated parameters, while continuously updating the states with an EnKF. But, the estimation with the augmented state instead of the shuffled complex evolution metropolis algorithm is more computational efficient. We added this in the description of the Standard EnKF (page 11, line 6).

b. Further, the suggested approach seems to me a lot more similar to a batch calibration than a filtering, as the same data is used over and over again to calibrate the model towards one final parameter estimate. As it to me is unclear what effect the top boundary estimation has (see below), and as the truly dynamic parts of the data are removed, I have a problem seeing why a filter and not a proper calibration would provide a good solution, given that the filters after all are after all suboptimal for the unsaturated zone?
Reply: Yes, our approach is not a typical filter application for state estimation. Similar to a batch calibration we focus on the parameter estimation, but we argue that our approach goes beyond. Unlike the batch calibration, the EnKF continuously updates the state and could incorporate model errors. Due to the continuous parameter estimation we were able to detect the times when the model assumptions do not

hold (the closed-eye period). Due to the state estimation we can remove these times (which in this case correspond to the dynamic parts of the data). The state estimation enables us to guide the state through this time with a erroneous model and allows us to pick up the estimation afterwards again.
The estimation of the boundary condition is required to prevent the introduction to biases into the parameters. This again is required for a proper detection of the closed-eye period (see below).

2. Top boundary and its update
a. The top boundary leaves a few questions unanswered. It is claimed throughout the manuscript that the top boundary is updated and that this is largely also the reason for the use of a data assimilation method rather than a calibration. However, the authors neither show nor discuss the result of these updates, leaving a reader wondering about the necessity of this inclusion.
b. Further, I do not understand what boundary was used to drive the model. On P8,L10 it is written that the temporal resolution is daily unless a change of mode occurs. On the other hand, all the figures show finer resolved top boundary, and also the scenario descriptions are dealing with temporal resolutions of 10 minutes when describing the rain intensity. If these events are smoothed across a full day (or if the event is shorter than a day, across a full event), it is a considerable smoothing, and it would be much more preferred if the authors also plotted the forcings they are using. Further, wouldn't this risk effecting the parameter estimation during the rain event, if reality has a much stronger peak than the model? Or is the first observation so far down that this has no effect? Please clarify!
Reply: We agree, that we did not discuss the estimation of the upper boundary in sufficient detail in the manuscript. An estimation of the upper boundary condition is not mandatory. A stochastic representation of the uncertainties could be sufficient. But we argue, that a bias in the upper boundary condition could induce a bias in the estimated parameters. Especially if this bias mainly occurs during the rain event, it

could cause a parameter shift during the rain event, similar to the one observed during the closed-eye period. To prevent this, we do estimate the boundary condition. This requires the decreased temporal resolution, which could lead to short-term parameter variations within the rain event. We think that this downside is subordinate to the removal of the bias, especially since it is strongly alleviated by the disspiative soil water movement.

We improved the description of the boundary condition in the manuscript (page 8, line 5-11): "The expansion to an augmented state changes the propagation in time. Each component needs an individual forward propagation. We assume the soil hydraulic parameters and Miller scaling factors to be constant in time. This is not possible for the upper boundary condition where the forward equation is unknown from a soils perspective. However, measurements are available to estimate the evaporation and precipitation. Hence, we assume the forward model constant until a new estimation is available. Then, we switch to the estimated boundary condition. To base the improvement of the upper boundary condition on several measurements, we reduce the temporal resolution of precipitation measurements to change daily and at transitions between precipitation and evaporation. This means, that the upper boundary condition is treated like the parameters, except that the value can change in the forward propagation. The original temporal resolution of the precipitation data (10 minutes) is not required due to the dissipative nature of the Richards equation, which smooths the infiltration front before it reaches the first TDR sensor in a depth of 8.5 cm. The estimation of the averaged boundary condition will ensure that there is no global bias on the parameter estimation during the rain event, but could lead to small short-time parameter drifts within a rain event." and (page 11, line 34): "As we added the boundary condition into our estimation, we can exclude effects due to a bias in the precipitation to cause this parameter shift. The reduced temporal resolution in the boundary condition could only cause short-term parameter changes within the rain event."

We also show the boundary condition with the applied reduced temporal resolution in

Fig. 8 now.

3. Purpose of model?

A nice and stable model is calibrated, that can predict water movement as long as nothing really happens. I find the result very interesting, as it properly questions the use of the standard Richards equation for modelling the unsaturated zone, given that the rainfalls that caused the problems here are quite moderate. The authors should take their space to consider the implications of what they show; can we use the Richards equation in the field? What purpose does a model have that is seemingly data driven (P11,L15) and that cannot predict during rainfall?

Reply: Yes, we agree that the Richards equation cannot describe preferential flow, which limits it's applicability. Quite to the opposite preferential flow is mostly relevant for the top layers only and less for the lower layers. Consequently the Richards equation is valid there and can be applied.

In our manuscript we are interested in the actual processes. We separated the times with and without preferential flow and show that the Richards equation can really describe the times without preferential flow.

Furthermore, if one is not interested in the processes, but an optimal prediction a batch calibration can yield better results, because there effects of preferential flow are incorporated into the parameters, which then in part compensate the non-represented processes.

4. Memory

Since the model used is disconnected from the groundwater, all information within the model has to travel from the top to the bottom in the 1D column. As the lower layers are more dependent on a longer memory to properly assess their behavior (e.g. it takes longer before we see an effect of a rain event in the second layer than in the first), there is a good risk that we smooth this information when continuously updating

the first layer.
a. Wouldn't a continues update of the top layer risk always smoothing the model such that the estimation in the bottom layers becomes difficult?
b. Could this be helped by the closed-eye period (as here, the rain event that is too strong for the top layer, may give valuable insight in the lower ones and if the model is not updated during this period, the front reaches down)?

Reply: For the estimation of the second layer, the state of the first layer at the border to the second layer has to be represented as well as possible. We can picture the first layer as a boundary condition for the second layer. This means, that the state estimation of the first layer will improve the parameter estimation of the second layer. Especially during the rain event this state estimation becomes very important. Without it, the infiltration front would reach the second layer at the wrong time, leading to biased parameters there.

The closed-eye period in the first layer does not improve the parameter estimation in the second layer, because the parameters in the first layer are worse at describing the rain event This requires stronger corrections in the state, which actually might make the parameter estimation in the second layer more difficult.

But, we agree, that in case of a proper representation of the preferential flow a better estimation of the second layer might be possible with a longer memory.

5. Validation

Using data that has been used 10-20 times to calibrate a model with as a validation set is not a particularly strong case. More attention could be given to scenario D, where it is nicely shown that the CE-model has a much better performance than the Standard one. Even better would it be if one of the three observations in the top layer could be taken out of the calibration and used for validating the model; this would be a strong case.

Reply: Yes, we agree, taking one sensor out of the calibration and using it for validation would be a strong case. Unfortunately this is impossible in the given scenario. Due to

the detected small scale heterogeneity, this heterogeneity has to be estimated at each sensor location. Without it, it is impossible to predict the water content at this location. This is a fundamental challenge in a heterogeneous medium and the only way to verify the results is the presented internal consistency.

6. All parameters needed?

The bottom $\frac{3}{4}$ of the model is only briefly discussed in the manuscript. That nothing happens in layer 4 is not so surprising, but how does it look for layers 2 and 3? Do the infiltration fronts reach down here and how is it with the parameter estimation? If nothing happens, do we need to estimate them? How is the effect of the closed-eye period on the second layer?

Reply: We have updated Fig. 4 and included the layers 2 and 3. In both layers there is basically no dynamics as well. We agree, that the estimation of these parameters is operationally not required in this case, but we wanted to address the challenge to estimate parameters there. We could see, that we can already estimate heterogeneity and reduce uncertainty in soil parameters. Additionally, we gained information about the robustness of the procedure. The estimated parameters did not reach values, that would be excluded by expert knowledge. Without the representation of heterogeneity the hydraulic conductivity in those layers dropped rapidly to very small values, which could be excluded.

7. Overview

The authors could consider including a nice block diagram to make their approach clearer (which time period is used for what and where is it iterated and what is feeded to where?). Two example of unclearness: 1) what is used from 2nd iteration onwards as initial condition for period C; the same output from B despite changed parameters? 2) for the top boundary, is it he finally updated value for each model that is also reused in the next iteration?

Reply: We followed your advice and included a block diagram. To the specific questions: 1) the same initial condition generated once is used for all iterations. 2) the updated mean value and standard deviation for each time interval are used to generate the boundary condition ensemble for the next iteration. Both is mentioned explicitly in the text now.

8. How to select a closed-eye period?

One of the key findings of the manuscript it the improvement using a closed eye period. What is not so clear is how this is to be selected. How can we differentiate between a wanted changes of parameters away from a false prior, from the erroneous updates that the authors show in this work? This feels very ad hoc, and hence also leaves the full paper feeling very problem specific. I think the manuscript would give a more rigorous feeling if this issue was discussed in more depth.

Reply: The idea of the closed-eye period is not ad hoc and actually generally applicable. In our case we can detected the closed eye-period due to the iterative approach and thus we can determine, based on the assumption of constant parameters, the closed-eye period due to the change in the direction of the parameter update.

We enhanced the discussion accordingly (page 15, line 14-20): "Due to the iterative approach, we could detect the times when the local-equilibrium assumption is violated: the variations of the parameters are larger than the change from initial to final value during later iterations. The changes in the direction of the parameter update then determine the closed-eye period.

Generally, the closed-eye period can be detected, if the operational limits of the model are known. In our case, we base this on the changing parameters, but e.g. a direct detection in the state or forcing could be possible as well.

The Closed-Eye EnKF omits the incorporation of the model structural errors in the parameters and is a generally applicable concept. In this study, it yields better predictions during periods when the underlying assumptions are fulfilled: the drying period after a rain event when there is local equilibrium, showing the strength of the

Richards equation there."

Minor concerns

9. Longer development plots of parameters

The parameter plots only show the last iteration, however, I would find it highly interesting to know how the development throughout the iterations also looks. Example: mean alpha is initially sampled at 4.8 and has in Figure 6 a value for the standard filter of maybe 4.7 with a jump of 0.05. Hence, it cannot have looked the same during the other 9 iteration, then the value would be different. Similarly, the closed eye filter has a stable value of around 5, but how did the way there from 4.7 look?

Reply: We have added the evolution of the parameters during all iterations. Especially during the first iteration parameters can overshoot to compensate other parameters, leading to the observed behaviour. Additionally, the mean value is not perfectly conserved from one iteration to the next, because the ensemble is not kept, but sampled new from a Gaussian distribution with mean and standard deviation of the final parameters of the previous iteration.

10. Performance of org. model

Please include in the water content figures, also the performance of the mean of the original model, so that the reader has the possibility to assess the positive development of the model during the filtering.

Reply: We included the original model in Fig. 4. It assumes homogeneous soil layers and cannot describe the observed water contents.

11. Why this damping?

As the selection of a damping parameter is anything but obvious, I think it is useful if authors using it gives a one sentence motivation to why it was chosen this way!

Reply: We based our choice on values seen in the literature. We enhanced the text on page 8, line 20-21: "Hendricks Franssen and Kinzelbach (2008) investigated damping factors between 0.1 and 1.0, while Erdal et al. (2014) employed a damping factor of 0.3 for the parameters. The smaller the value, the better can a non-linearity be handled, but the slower will the estimated value approach the optimal value. We choose the values d = [1.0, 0.1, 0.1, 0.1]. This means, that based on the water content measurements, there is no damping for the update of the water content state, but a strong damping for soil hydraulic parameters, Miller scaling factors and upper boundary condition."

12. Resolution of model

The resolution is uniformed 1cm (P5,L20), which for a model with strong boundary fluxes may not be so small. Has the authors checked that the grid size is not also causing issues? Why is "the effect" minimized by 1 cm and not at 1mm?

Reply: We improved our wording to: "We choose a resolution of 1 cm. For the represented soil and boundary conditions this resolution is sufficient to resolve the occurring fronts smoothly, so that we can neglect effects caused by the numerical solver."

13. Why 100 members?

The model has 140 unknowns and is 1D, it cannot be a particular difficult to also run a larger ensemble size and reduce the ad hoc tunings, so why is this setup with a small quite small ensemble chosen?

Reply: We did not investigate the ensemble size, as 100 members proofed to be sufficient. By increasing the ensemble size we only see the localization as a possibility to reduce the applied improvements. We do not see the localization as an ad hoc tuning, though. A limited information range is physically based.

The damping factor is included to attenuate the effect of linearizing non-linear relations, the covariance inflation is added to prevent filter inbreeding, which can be caused

non-Gaussian distributions and the prior estimation of Miller scaling factors is added due to otherwise occurring wrong correlations between hydraulic conductivity and unrepresented heterogeneity.

14. True parameters or heuristic model (was Wollschläger wrong?)

a. I'm a bit confused about this discussion. On P12L24 it is stated that the estimated parameters of the closed eye filter better resembles the believed true parameters and that the standard filter ones are more/only heuristic. However, on P14,L1 it is clearly stated that the estimated parameters are all heuristic and only valid in their estimated range, which hence suggests that there are no such things as true parameters.

b. Also an elaboration on the different result presented here to that of Wollschläger 2009; you get quite different result for a plot quite close by. Are the result so local or where the previously published results not so reliable?

Reply: We agree, there are no true parameters, but there are, what we call 'believed true material properties', which are only valid in the observed range. We have added our definition (page 12, line 24): "This means, if the parameters are constant in time, we believe, that the parameters can represent reality in the observed water content range during times, when the underlying assumptions hold."

Wollschläger et al. (2009) were not wrong. The used data comprised only one TDR sensor in each layer. The obtained results are based on the assumption of homogeneous layers and are optimal in the sense of an inversion when a true model is assumed. As shown in the manuscript the representation of heterogeneity is important, though.

Technical stuff

1. Figure 6: negative alpha values?

Reply: The sign of the α values is connected to the choice how the potential is defined. If bound water content states are chosen to have a negative potential, the α

values have to be negative.

2. Language: should be checked carefully. E.g. what does the sentence "The forcing or embedding in space is the initial condition" mean?

Reply: We corrected the sentence and improved formulations throughout the manuscript.

3. Consider splitting the "Conclusion" section into a "Summary and discussion" (where some of the discussion points taken up in this review would fit) and a short "Conclusion" section which only contains the actual conclusions.

Reply: We kept the existing sections, but improved the Conclusion.

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-296/hess-2016-296-AC1-supplement.pdf

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-296, 2016.