Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-126-RC2, 2016

Hydrology and Earth System Sciences Discussions

Interactive comment on "Multivariate hydrological data assimilation of soil moisture and groundwater head" *by* D. Zhang et al.

Anonymous Referee #2

Received and published: 26 May 2016

In the manuscript titled "Multivariate hydrological data assimilation of soil moisture and groundwater head" authors have assimilated groundwater head (GW) and soil moisture (SM) observations into Mike She hydrological model using ETKF in a synthetic and a real data assimilation scenarios. Later, these assimilation analyses are validated via comparisons of the accuracy statistics for soil moisture, groundwater level, evapotranspiration, and discharge variables. The study builds on earlier studies that have investigated the ensemble size, localization, and model uncertainty impact on performance of a filter assimilating GW in Mike She model in a synthetic setup (Rasmussen et al., 2015; Zhang et al., 2015). Here authors are assimilating real data instead of synthetic data. Overall, the study is interesting and relevant to HESS Journal. However, the manuscript has many missing details and inconsistencies that prevent acceptance of the study in its current form.

Printer-friendly version

Discussion paper

1) Synthetic and real data assimilation experiments are not performed or presented in a consistent way. This prevents results to be meaningfully compared:

1.a) Synthetic studies show "localization degrade analysis for the univariate case when either SM or GW observations are assimilated (Figure 3)" while the multivariate synthetic case and the real data assimilation case shows "improvements in the analysis (Table 3 and Figure 4)". Synthetic studies always reflect the ideal/perfect conditions as they can be fully controlled (particularly the observation errors). I don't see the utility of an application that does not give satisfactory results using synthetic simulations. On the other hand, I am completely puzzled how real data scenario improves the analysis (Table 3) despite synthetic studies fail (Figure 3). Frankly, I would have been more convinced if synthetic results showed improvements while real DA case showed problems (after all real life is not perfect), but not vice versa. This inconsistency should be explained in detail.

1.b) Only the total ET values are given for the synthetic simulations, but not the error statistics of ET (Table 2). They should be given similar to real data assimilation scenario.

1.c) Pick an accuracy statistics (R2; RMSE; or Nash-Sutcliffe efficiency, NSE) and just present it in a consistent way through out the study. Presentation of mixed statistics is very confusing: Table 2 shows R2 and NSE for discharge, Table 3 shows RMSE for SM, Table 4 shows R2 and RMSE for ET while the same table shows R2 and NSE for discharge, Figures 3,4,5,9 shows RMSE. To start with, just give the equations of the error chosen statistics. In general, showing both NSE and RMSE is redundant. NSE (1 - sum(X-Y)ˆ2 / sum(X-mu_x)ˆ2) can be shown to be equal to (1-RMSEˆ2/variance_x), implying there is a directly relationship between NSE and RMSE: if NSE is high we can expect RMSE to be low and vice versa.. This analytical expectation is also supported by the results given in Tables 2 and4: higher R2 experiments have higher NSE and lower RMSE. Briefly, these three statistics are consistent with each other in representing the accuracy of the variable of interest. Just pick one out of these three statistics and just show it consistently for all cases (I recommend RMSE in this case, it is up to authors though).

1.d) Table 2 (synthetic data assimilation experiments) is missing “DA_H” and “DA_SM5” scenarios (no localization scenarios). These scenarios are necessary to see whether or not Localization improves or degrades the analysis accuracy. Given one of the major conclusions the authors are making is “Localization does not only provide better results . . .”, presentation of these results is very important.

1.e) Figure 3 (synthetic data assimilation experiments) is missing “DA_SMBoth” (no localization scenario). Similar to above, these scenarios are necessary to see whether or not Localization improves or degrades the analysis accuracy.

1.f) Tables 3 and 4 (real data assimilation experiments) are missing “DA_H”, “DA_SM”, and “DA_HSM” scenarios (no localization scenarios). Similar to above, these scenarios are necessary to see whether or not Localization improves or degrades the analysis accuracy.

2) I see SM simulations/assimilation experiments are not realistic.

2.a) SM observations often have inconsistencies with models (see Reichle and Koster, 2004). For this reason, SM observations are matched to model before they are assimilated. Have authors done something similar? I see authors have matched the means of GW before observations are matched (page 14, line 9), but not the standard deviations or any other statistical property. To start with, do model and observations have similar statistical property? Did authors check the innovations of ETKF? Is it white?

2.b) Many satellite missions (e.g., SMAP) have the goal of retrieving SM observations with 4% error and consistently Mike She model using real data has similar error magnitude. However, the synthetic Mike She model runs do not seem to be realistic with SM errors of 2%.

2.c) Root-zone SM varies much slower than surface SM. As a result, actual root-zone

SM errors have smaller magnitude than surface SM errors (i.e., root-zone RMSE is expected to be smaller, while RMSE values normalized with actual variability of the variable could be different for two layers). In the current study, root-zone SM errors are higher than surface SM errors (Fig. 4). Deeper layer (22.5cm) SM seems more noisy than surface (2.5cm) as shown in Figure 9. Explanation is needed.

2.d) Above points can be clarified very easily using a table showing: mean and standard deviations of observations and model (no assimilation) for both SM (at 5, 22.5, 50cm depths) and GW. Such a table seems necessary to clarify all of above points.

3) Why select ETKF but not EnKF? In atmospheric sciences ETKF could be a viable option because it does not perturb observations (atmospheric models are chaotic, hence such perturbations could be problematic; but hydrological models are not chaotic). Avoiding the perturbation of observations does not seem to be a sufficient justification to use ETKF. Many studies use EnKF, hence it is more relevant to general audience. Briefly, it is better to say “ETKF selection is arbitrary, EnKF could have been selected as well” rather than justifying the ETKF selection via “it avoids the additional perturbation step”. In case authors would like to support the selection of ETKF with “it avoids the additional perturbation step” argument, then they should support this claim with a reference (i.e., a study shows this additional perturbation could be problematic in hydrological sciences).

4) Many details are not given in the paper and currently the experiments cannot be replicated by another researcher. Besides many experiment set up decisions looks very arbitrary:

4.a) which datasets are available for the warm up period over Ahlergaarde (i.e., for forcing/parameter)?

4.b) forcing and parameter perturbation statistics for real data assimilation case are completely missing (which forcing variables/parameters are perturbed?, additional/multiplicative noise? mean?, standard deviation?, daily/weekly perturbation?, etc). How did authors decide about these statistics, justification?

4.c) for the synthetic experiments observations are perturbed using noise with 0.15m and 5% error standard deviation. Why these numbers? Do authors know Decagon 5TE SM sensors have errors of 5% at daily time step? (i.e., SM observations are assimilated at daily time steps even though they are observed every 30min; implying the errors at daily time steps could be much lower than 30 min time step).

4.d) For the real data assimilation scenario, the soil moisture error standard deviation is assumed constant 5%. Why 5%? Is it something the company who produces these Decagon 5TE sensors suggests? Besides, these half hourly observations are averaged into daily values, implying the observation noise if further reduced. Justification for the observation error standard deviation is needed.

4.e) perturbation details in generation of ensembles is missing as well (the study of Zhang et al., 2015 is referred for these perturbations but the experiment details should be given specifically).

4.f) what is the temporal resolution of the model? Daily? It is forced by daily precipitation and reference evaporation (page 4 line 26), bi-hourly GW observations are obtained for real DA case (page 14, line 2), and GW observations interpolated to weekly time steps are assimilated (page 14, line 28) but the temporal resolution of the model (i.e., at what time step the model is run) has never been explicitly mentioned.

4.g) if the model is run at daily time step, then it is not clear how weekly GW observations are assimilated.

4.h) How ET observations are obtained at the stations? What kind of observations are these (eddy covariance? lysimeter? pan evaporation? Reference ET?)? How frequently obtained? No details are given.

4.i) Which data are used as validation? Do authors use the same stations where observations are collected? Perhaps some stations should be reserved and observations obtained over these stations should not be used in any ways (e.g., assimilation) and should be left purely for validation. After all, if the station has soil moisture probe, then why bother with estimating the SM over that location (this comment is only relevant with direct validation of SM, should not be thought in the framework of SM assimilation to estimate other parameters such as runoff, ET, etc).

4.j) Have you considered assimilating remote sensing-based SM data? Why not?

4.h) "The assimilation performance is evaluated by comparing model output with the actual observations using average RMSE." Now I am puzzled, which observations are assimilated and which observations are used as validation? Are they the same?

5) The study could present more literature review.

5.a) For example Franssen & Kinzelbach (2008) did assimilate GW observations into a hydrological model. Mike She model background/physics (page 4, lines 14-20) could be supported with references as well. ETKF is first introduced by Bishop et al., (2001) not by Sakov et al., (2010); the latter study used EnKF they did not even use ETKF (Page 3, line 16).

5.b) SM and GW time scales are often very different; the impact of precipitation/SM variations could translate to GW variations only after days/weeks/months, depending on the catchment/GW levels/conductivity/etc. On the other hand, the Mike She model simulations have daily time steps (my understanding daily, I could be wrong → authors should clarify this info). If the time scale is very long, it is not immediately clear how SM anomaly for today will give meaningful information about GW anomaly of today. Perhaps the filter (ETKF) should be changed to accommodate past observations? So, sufficient motivation about the utility of SM observations to improve GW simulations using ETKF should be given.

6) "Although the improvements are relatively small, we nevertheless see the benefits in other model process results when improving groundwater head and soil moisture."

It is not clear what is meant with this sentence; do authors imply "we did not find improvement but we think it is still likely in other applications, hence GW and SM should be assimilated together"? After improvements were not found in this study, I think the only comment can be said is "these results should be verified using other models". The interpretation of "seeing the benefits in other model process results" is not can not be made using the results of this study alone.

7) "Localization does not only provide better results . . ." (Page 16, line 25), synthetic results do not support this comment (Figure 3).

8) Abstract requires slight modification: the assimilation method used here "ETKF" has been around for almost 15 years, so this is not the first time it is introduced. Assimilation is not a new method either, has been around for a long time too. So, this current study is just a case study (GW and SM observations are assimilated in a hydrological model). Abstract should be changed to something similar to below:

"Observed groundwater head and soil moisture profiles are assimilated into an integrated hydrological model. The study . . .. model code. Experiments were firstly performed using synthetic data in a catchment of less complexity (the Karup catchment in Denmark), and later performed using real data in a larger and more complex catchment (the Ahlergaarde catchment in Denmark)." âĂČ

MINOR:

Page 3, line 15, ". . . performance of assimilating soil moisture" → ". . .performance of a filter assimilating soil moisture".

Page 4, line 4, "Karup catchment is well-studied catchment", in terms of what parameters/variables?

Page 10, line 17, "assimilating soil moisture at 5 cm depth . . .. and 1 m depth". 5 cm refers to the depth of the assimilated observations, how about 1m? Soil moisture states up to 1m depth are updated? Is it what is meant?

Page 10, line 18-19, similar to above (Page 10, line 17).

Page 13, line 21, “spilt” → “split”

Page 16, line 14, “EnKF” → “ETKF”? I believe authors imply “ETKF is a flavor of EnKF” here, but still use of EnKF is confusing since EnKF is not used in this study.

Craig H. Bishop, Brian J. Etherton, and Sharanya J. Majumdar (2001). Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. Monthly Weather Review, 129:3, 420-436.

Reichle R. H., Koster R. D. (2004). Bias reduction in short records of satellite soil moisture. Geophysical Research Letters, 31, L19501, pp 1-4.

Hendricks Franssen, H. J., and W. Kinzelbach (2008). Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, Water Resour. Res., 44, W09408, doi:10.1029/2007WR006505.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-126, 2016.