Hydrol. Earth Syst. Sci. Discuss., 9, C6702–C6709, 2013
www.hydrol-earth-syst-sci-discuss.net/9/C6702/2013/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Kalman filters for assimilating near-surface observations in the Richards equation – Part 2: A dual filter approach for simultaneous retrieval of states and parameters" *by* H. Medina et al.

Anonymous Referee #1

Received and published: 12 February 2013

General comments

This paper presents a dual Kalman filter strategy for the update of states and parameters in the Richards equation, based on the assimilation of near-surface pressure head or soil moisture. The system state, accounted for in terms of either pressure head or soil moisture, is updated by the classic Kalman filter (KF), while the Unscented Kalman filter (UKF) is used to update the soil properties (hydraulic conductivity and soil retention curves parameters). Although the topic is certainly of interest for the hydrological

community, I have a number of serious concerns that in my opinion must be addressed before the paper can be published.

Overall, it seems to me that the conclusions of the study actually represent a step back in the topic of data assimilation for nonlinear systems, as the strategy is largely based on the classic Kalman filter applied to a linearized numerical scheme for the solution of the Richards equation. Therefore, the Authors are trying to challenge the nonlinearity of the problem by adapting the system to the assimilation algorithm and not vice versa, which is the common strategy. This is of course legit and the choice cannot be questioned "a priori". The problem, in my opinion, is that the results, contrary to what claimed by the Authors, are neither convincing nor encouraging. I will provide more details in the following, under the section "specific comments".

Another major issue of the paper is that the methods are not described comprehensively. I refer in particular to the UKF, whose description is rather confused and difficult to follow. As far as I understand, UKF theory is not straightforward and its application to the Richards equation represents one of the major novelties of the paper: for these reasons I would expect a more precise description, including the details of the several parameters that need to be tuned in order to obtain a satisfactory performance of the assimilation framework. These parameters are hastily discussed in section 5.5, but this is not sufficient: a reader who wanted to reproduce the same results would not be able to do so only with the information provided in the paper. Again, more details follow below.

Specific comments

Page 13331, lines 6-13: given that the paper deals with a 1-D model of unsaturated flow, this discussion is not much relevant and can be shortened. This would allow adding more details on the subsequent paragraph, as reported in the following point.

Page 13331, lines 15-18: this paragraph is much more relevant than the previous and thus should be expanded with a proper number of citations of studies focused i) on

retrieving states only and ii) on retrieving states and parameters.

Page 13332, lines 17-20: this statement cannot be generalized. Other KF-based algorithms, such as the ensemble Kalman filter (EnKF), have been proved to be efficient when implemented with a nonlinear numerical solver of the Richards equation. For instance, Camporese et al. (WRR, 2009) reproduced the same experiments of Entekhabi et al. (1994) and Walker et al. (2001) with a good retrieval of the system state and without any problem of numerical stability.

Page 13333, line 13: "weights"? Perhaps do you mean "parameters"?

Page 13333, eq. (1): I believe that u and v should be at the current time step k. Also, the meaning of the symbol "ˆ" should be stated right after the equations, while it is (probably, see related technical correction) currently placed much later, at line 7 of page 13334.

Page 13334, eq. (5) and elsewhere in the manuscript: you cannot use the symbols "ˆ" and "-" together, as the former means posterior and the latter prior. Please remove "ˆ" from all the prior variables.

Page 13336, eq. (16): this equation should be either demonstrated or given a proper reference.

Page 13336, line (15): the expression S_i = {mi_i, W_i, i = 1. . .L} is erroneous and misleading. It should be i = 0. . .2L and it should be stated that the actual sigma points are the variables W_i, while mi_i are the weights used for computing mean and covariance.

From page 13337, line 7 to page 13339, line 6: all this part is obscure and rather incomprehensible and must be rewritten. Also, when writing down the equations, please remember to define every variable (e.g., missing definitions currently include L_w and Yˆz_k|k-1 in eq. (24), P_wk_yk in eq. (25), R_ek in eq. (26), etc).

Page 13344, lines 11-12: "zero gradient" boundary conditions mean no-flow and not "free drainage". Please clarify what boundary conditions were used in the simulations.

Page 13345, line 1: please clarify what do you mean by "heterogeneous discretisation". Does that mean that the layer thickness is not constant? If so, please give more details.

Page 13345, lines 23-24: stating that a variance of 10ˆ3 cmˆ2 is ten times the initial state does not make any sense, as you are comparing square cm with cm. You should refer to the standard deviation ($\sim$32 cm) or coefficient of variation, which in this case is $\sim$1/3.

Page 13345, line 28: "is updated as a diagonal matrix", please rephrase this sentence.

Page 13346, eq. (43): I believe that using the Mean Absolute Error would be more desirable. A perfect score of ME = 0 in fact does not exclude very large errors of opposite signs which cancel each other out.

Page 13346, eq. (44): this equation is not correct. For a right normalization you should take sigma_SD out of the root square, otherwise the computed RMSE would not be dimensionless and could not be compared between the h-form and theta-form scenarios. This error casts doubt on the results presented in Table 3 and discussed later on the in the manuscript.

Page 13347, lines 10-11: in my opinion this is a critical issue. Could it be related to the fact that the Crank–Nicolson approximate solution can contain spurious oscillations in certain conditions? The backward Euler method is both stable and immune to oscillations but of course must be coupled with a nonlinear KF-extension such as, e.g., the ensemble Kalman filter.

Page 13347, lines 16-23: have you tried a simpler log-transformation of the parameters? It would probably suit better the hydraulic conductivity and the range of search would not be limited between two values, a characteristic that could help when no prior information are available.

Page 13347, line 24: the covariance matrices should be chosen on the basis of physical considerations and not be tuned to ensure the convergence of the method.

Page 13348, lines 7-8: on the other hand, it should be mentioned that, contrary to the h-based form, the theta-based equation does not allow solving problems involving also the saturated zone.

Page 13349, lines 14-15: I would rather say that Ks is not identifiable, as it diverges even when the initial guess is close to the true value.

Page 13349, lines 25-29: perhaps it depends on the quality of the figure (it is very hard to make out the different lines), but I cannot see a clear difference between prior n > 2 and < 2. In particular, scenarios with prior n = 2.6 seem to converge no differently from the other ones. Slight differences can only be discerned in the theta-theta retrieving modes.

Page 13350, lines 7-14: in view of the previous point, this discussion can be removed altogether, as well as Figure 3 and lines 1-8 at page 13351.

Page 13350, lines 26-29: again, this is very subtle and, to tell the truth, it seems to me that the contrary occurs (see previous point).

Page 13351, lines 9-14: I do not agree that these results are encouraging. Neither the results about Ks, which is of fundamental importance and cannot be retrieved, nor those about n and alpha are conclusive. It seems that in all cases the solution converges to values that are not far away from the initial guess, so a few questions remain. What happens if you start from prior guesses that are further away from the truth? Is the solution found by the method better than the prior? In other words, are the errors associated to a posteriori simulations run with the retrieved parameters smaller than the errors computed by simulations run with the prior parameters (open loop)?

Page 13351, lines 21-23: this is very contradictory. It may be that the error on RMSE (see previous point at Page 13346, eq. (44)) affects these results and thus leads to ambiguous conclusions. Please double-check the RMSE.

Page 13352, lines 13-15: this is somewhat surprising. Can you explain why a higher

assimilation frequency would yield worse results?

Page 13352, lines 20-22: again, this statement seems to contradict the consideration at page 13351, line 21.

Page 13354, lines 1-6: much more details are needed here. Stating “the unscented algorithm is also employed for the statistical linearization of the now nonlinear operator H, similarly to what is done for retrieving the parameters” is not sufficient. How has this been done? What parameter values have been used? Rewriting section 2.1.2. could help to shed more light on this subject, too.

Page 13354, lines 10-14: this is one of the reasons for which the ensemble Kalman filter is more popular than EKF for this kind of applications.

Page 13355, lines 3-15: much more details are needed here. The reader might want to reproduce the results and, for doing so, need to know exactly what values have been used in each scenario. Suggest adding a table that includes all the parameter values used for each scenario. Besides, if the algorithm is so dependent on the choice of the matrices, it does not seem to be really robust. Could you please comment on that?

Page 13356: In view of all the points above, I would recommend the conclusions be rewritten. The dual Kalman filter approach seems to have more cons than pros and the implementation of the UKF does not seem so simple. After double-checking the results, the Authors could reconsider some of the statements on the sensitivity to the VGM parameters and the comparison between system state retrieval in the h-based and theta-based scenarios. Finally, the final considerations concerning “the convenience of retrieving state variable of the same type of the observation variable” cannot be generalized. As mentioned earlier, the EnKF has been demonstrated to work fine with nonlinear observation models as well as with linear ones.

Technical corrections

Page 13330, line 9: time-series series -> time-series.

Page 13330, lines 25-26: Richards (1931) is not in the reference list.

Page 13331, line 27: Vereecken et al. (2008) is not in the reference list. Please check all citations for consistency with the reference list.

Page 13332, line 5: a method developed in 1995 is not novel any more. Suggest removing the adjective.

Page 13333, lines 9 and 10: density -> densities.

Page 13334, line 7: "" -> "^"?

Page 13335, lines 1, 8, and 14: please itemize or number (e.g., i) Initialization . . ., ii) Prediction . . ., etc.).

Page 13335, line 12: R_v has already been defined at line 6, remove the definition here.

Page 13335, eqs. (14) and (15): is it "H" or "H_xk"? Please decide the notation and be consistent for the remainder of the paper.

Page 13336, eq. (17): mi_0^(c) -> mi_i^(c).

Page 13340, line 14: Eq. (31) -> Eq. (32).

Page 13340, eq. (36): Deltat^j -> Deltat^k. Same typo in eq. (37).

Page 13341, eq. (39): A_k+1 -> A_k.

Page 13342, line 4: Eq. (37) -> Eq. (39).

Page 13343, line 15: the abbreviation RM is not used anywhere else in the paper. Please remove it.

Page 13344, lines 11-12: stochastically generated daily series -> stochastically generated daily series of rainfall.

Page 13346, line 12: root mean square (RMSE) -> root mean square error (RMSE).

Page 13349, line 13: . . .moisture states is required to reliably constrain the soil hydraulic functions.

Page 13350, lines 13-14: "true parameter values as a function of parameter n", please rephrase this sentence.

Page 13352, line 1: Figure 5j -> Figure 6?

Page 13352, line 7: Fig. 6 -> Figure 7?

Page 13353, Title: Influence of the type OF observed . . .

Page 13353, line 18: . . . would BE significantly . . .

Page 13353, line 21: . . . prior TO applying . . .

Page 13354, line 16: their -> its.

Page 13362, Table 1: suggest removing "Synthetic data (SD)", as this table shows information only on the retrieval experiments. Also, add 1/3 days^-1 assimilation frequency.

Page 13363, Table 2: suggest adding a column that reminds the reader the values of the true parameters.

Page 13364, Table 3: Please add to the caption that ME and RMSE refer to time = 150 days. Moreover, double-check consistency between RMSE in retrieving pressure head and soil moisture.

Page 13369, Fig. 5: please clarify whether these results refer to assimilation frequency of 1 days^-1 (as stated in the caption) or 1/5 days^-1 (as stated in the text).

Page 13371, Fig. 7: please state in the caption that the RMSE values are time-averaged.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 13329, 2012.