Rebuttal letter for the Editor and the Referees

We thank the Editor and the Referees for their interest in our work and for the helpful comments that helped us to greatly improve our manuscript. They brought up some very good points and we appreciated the opportunity to clarify our research objectives and results.
We checked all the general and specific comments and made necessary changes to the manuscript according to their indications. In the following, a point-by-point reply on the comments of each Referee is reported. Attached is also a revised manuscript with track changes. Note that each line and page reported in this rebuttal letter refers to the track changes manuscript.

Reply to Anonymous Referee #1 – changes in the manuscript are highlighted in yellow.

First, I wonder if, in the interest of clarity, it would be possible to reduce the number of scenarios presented (as many as 19!) while preserving the main conclusions drawn in the study.

Thank you for drawing the attention to this point. Our main goal was to provide a comprehensive series of simulation scenarios, both in terms of variables to be assimilated and variables to be updated, in order to make our conclusions as general as possible. In our opinion, the number of scenarios presented (19, including two open-loop cases) is appropriate to assess the impacts and trade-offs associated to the different combinations of assimilated and updated variables. We tried to be concise by summarizing the results of all scenarios in Figure 4, which conveys a synthetic but effective overview. Then, for brevity, we only analyzed detailed results in a number of selected and representative scenarios.

Second, since an EnKF algorithm is used, it might be worthwhile to assess - perhaps by applying a restart - EnKF in some key scenarios - the effects of numerical inconsistencies introduced when updated, and thus statistically modified, states and parameters are merged into the flow model at the data assimilation times.

Thank you for raising this important issue. We actually considered applying a restart-EnKF, but then we decided not to use it for two main reasons. First, although the restart-EnKF is very valuable in the case of solute transport, where it is important that the contaminant mass after each update is maintained and consistent with the updated parameters (e.g., Camporese et al., 2011, 2015), in the case of flow only, possible numerical inconsistencies between the re-initialized system state and updated parameters quickly disappear thanks to the dissipative nature of the Richards equation. Second, but not less important, this strategy would be extremely computationally demanding, as we would need to restart many times from the beginning an ensemble of strongly nonlinear simulations. In summary, the trade-off between increased computational costs and expected improvements of the results would probably make a restart-EnKF not worth the effort in this context.\\

Additional minor comments, requests for clarification and proposed changes are provided in the attached document.

Thank you very much for your detailed evaluation of our manuscript. Each minor comment was carefully considered and properly addressed.

Page 1 line 15: a
amended Page 1 line 15
Page 2 line 11: using
Amended Page 2 line 11
Page 2 line 26: numerical
Amended Page 2 line 26
Page 2 line 31: cost
Amended Page 2 line 31
Page 2 line 32 - 33: unclear – rephrase
Amended Page 2 line 32 - 33
Old “in general to assess possible tradeoffs related to multivariate data assimilation.”

New "and to check which possible trade-offs might occur when assimilating different variables in a multivariate data assimilation framework"

Page 3 line 8: **use the present through this section**
Amended in the whole section 2

Page 3 line 4: for minimizing
We would keep "allow us to minimize"
Page 3 line 5: cancel
Amended page 3 line 10
Page 3 line 10: add Figure (1)
Amended page 3 line 17
Page 3 line 12: is and change the paragraph to the present
Amended page 3 line 20 and all section 2
Page 3 line 14: low –
Amended page 3 line 21
Page 3 line 23: cancel
Amended page 3 line 30
Page 3 line 24: were
Amended page 3 line 31
Page 3 line 27: were
Amended page 4 line 4
Page 6 line 7: Please clarify. Is the model input practically a net precipitation, that is, precipitation minus evaporation? Rain and evapotranspiration phases are imposed iteratively?

Old "rainfall and evaporation rates are imposed."
New "time-variable rainfall or evaporation rates are imposed"

Amended page 6 line 19

Page 6 line 9: whereas
Amended page 6 line 21

Page 6 line 10/a: cancel
Amended page 6 line 22

Page 6 line 10/b: being
Amended page 6 line 22

Page 6 line 11: add "**and the latter having little impact on the CATHY model response in this case."**
Amended page 6 line 23

Page 6 line 24: Please clarify. What does "fully unsaturated" mean? dry soil? or soil at residual saturation? Has this condition been observed/validated before the lab experiment?

As also explained in response to point 6 by Reviewer 2, the soil at the beginning of the experiment was dry (i.e., no saturated zones anywhere), although not at residual saturation. The model initial conditions were assigned based on the tensiometer measurements.

Amended page 7 lines from 11 to 17

Page 7 line 12: I guess T is truly an upper diagonal matrix, obtained from the Choleski factorization of the covariance matrix of the three parameters, alpha, n and thetha_r. Correct? (please explain)
OLD: "**T is the factored covariance matrix"**

NEW: "**T is an upper diagonal matrix, obtained from the Choleski factorization of the covariance matrix**

of the three parameters"

Amended page 8 line 5 and 6

Page 7 line 16: this is always the case, isn't it?
Indeed, we agree with the Reviewer

Page 7 line 19: please revise the sentence

did you mean "**… and the residual were used...**"?

OLD: "**the residuals used to calculate the correlation coefficients between all pairs of observations.**"

NEW: **"the residuals were used to calculate the correlation coefficients between all pairs of observations."**
Amended page 8 line 14 and 15

Page 7 line 28: cancel
Amended page 8 line 25

Page 7 line 29/a: cancel
Amended page 8 line 26

Page 7 line 29/b: were varied
Amended page 8 line 26

Page 7 line 31: with regard to or As to or As
Amended page 8 line 28

Page 8 line 23: tested
Amended page 9 line 20

Page 8 line 24, 25, 26 and 27: In principle, I have not been convinced by the use of dampening coefficients. They seem very much like fuzzy, empirical ways with which one gets to accept/reject or smooth down results of parameter updating. Can you comment on it? Are you aware of any sort of theoretical basis that justifies their use? What type of considerations does one need to make to understand how to "calibrate" these coefficients?

Thank you for raising this important issue, which has been also raised by Referee 3: in general, we agree that the application of a dampening factor can be debatable. But so are covariance inflation and update localization, which all are numerical tricks (Evensen, 2009a) designed to overcome the problems that may be caused by the use of a limited number of realizations in the EnKF. Nevertheless, they are widely applied and accepted in the data assimilation literature (Hendricks Franssen and Kinzelbach, 2008, Erdal et al.,2014, Erdal et al.,2015 and Zhang et al., 2016). In our case, we did not apply update localization as we are working at a small spatial scale where every variable is correlated to each other. Dampening factor and covariance inflation are two different tricks that eventually lead to the same result, i.e, preventing the ensemble variance to decrease too much. We decided to use a dampening factor instead of covariance inflation because it was particularly easy to implement in our data assimilation procedure.

Amended page 5 line 26, 27 and 28

Page 9 line 6: the most
Amended page 10 line 6
Page 9 line 11/b: cancel

Amended page 10 line 11
Page 9 line 11/c: in
Amended page 10 line 11
Page 9 line 11/a: the
Amended page 10 line 11
Page 9 line 12/a: A
Amended page 10 line 12
Page 9 line 12/b: indicates that
Amended page 10 line 12
Page 9 line 12/c: re
Amended page 10 line 12
Page 9 line 13: Likely
Amended page 10 line 13
Page 9 line 23/a: the
Amended page 10 line 23
Page 9 line 23/b: ,
Amended page 10 line 23
Page 10 line 5_6: T
Amended page 11 line 7
Page 10 line 13/a: add (left panels in Figure 5)
Amended page 11 line 9
Page 10 line 13/b: add (right panels in Figure 5)
Amended page 11 line 10
Page 10 line 11: one
Amended page 11 line 14
Page 10 line 14: cancel
Amended page 11 line 16
Page 10 line 16: , whereas
Amended page 11 line 18
Page 10 line 18/a: cancel
Amended page 11 line 20
Page 10 line 18/b:, which
Amended page 11 line 20
Page 10 line 20/a: on
Amended page 11 line 23
Page 10 line 20/b: from
Amended page 11 line 23
Page 10 line 25: add "**and -perhaps- also by the low sensitivity of the assimilated variables to the clay permeability**"
Amended page 11 line 28

References

Hendricks Franssen, H. J. and Kinzelbach, W.: Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, Water Resources Research, 44, n/a–n/a, doi:10.1029/2007WR006505, http://dx.doi.org/10.1029/2007WR006505, w09408, 2008.

Evensen, G.: Data Assimilation, Springer, 2009a.

Camporese, M., G. Cassiani, R. Deiana, and P. Salandin (2011), Assessment of local hydraulic properties from electrical resistivity tomography monitoring of a three-dimensional synthetic tracer test experiment, Water Resour. Res., 47, W12508, doi:10.1029/2011WR010528.

Erdal, D., Neuweiler, I., and Wollschläger, U.: Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model, Water Resources Research, 50, 132–147, doi:10.1002/2012WR013443, http://dx.doi.org/10.1002/2012WR013443, 2014.

Camporese, M., G. Cassiani, R. Deiana, P. Salandin, and A. Binley (2015), Coupled and uncoupled hydrogeophysical inversions using ensemble Kalman filter assimilation of ERT-monitored tracer test data, Water Resour. Res., 51, 3277–3291, doi:10.1002/2014WR016017.

Erdal, D., Rahman, M., and Neuweiler, I.: The importance of state transformations when using the ensemble Kalman fil- ter for unsaturated flow modeling: Dealing with strong nonlinearities, Advances in Water Resources, 86, 354 – 365, doi: https://doi.org/10.1016/j.advwatres.2015.09.008, http://www.sciencedirect.com/science/article/pii/S0309170815002092, data assimilation for improved predictions of integrated terrestrial systems, 2015.

Zhang, D., Madsen, H., Ridler, Marc, E., Kidmose, J., and Jensen, Karsten, H.: Multivariate hydrological data assimilation of soil moisture and groundwater head, Hydrology and Earth System Sciences, 2016.

Reply to Anonymous Referee #2 – changes in the manuscript are highlighted in green.

1)In general I think it could be made clearer what the purpose of the data assimilation framework is, so that it is easier to follow how the performance is evaluated. It makes a difference if the purpose is to make predictions based on continuously measured observations or if predictions should be made also without observations. In the first case the question of parameter identification is less important than in the second case. If observations are available all the time, parameter updates might improve state updates and state predictions, but it is not important that parameter updates yield reasonable parameter values. This is different for long prediction periods without observations. I assume that the purpose is here to make predictions also without observations, as a long validation period is chosen.

We thank the Referee for highlighting this aspect. Indeed, the purpose of the model is to make predictions also without observations: as mentioned in the end of the Referee's note, a long period of the whole simulation has been devoted to the validation of the model (no assimilation) as we wanted to verify the model in a situation where no measurements are available. The overall purpose of the data assimilation framework has been made clearer in the Introduction of the revised manuscript.

Amended – Introduction page 3 lines 12, 13 and 14

2)Page 3, lines 3-4: Here and also at other places it is stressed that the flow processes considered in the experiments are dominated by strong non-linearities. It would be good if this could be explained a bit more in one or two sentences or if the statement could be taken out. I do not see where non-linearities are dominant in the experiments. In Figure 2 it seems as if all variables follow well the rainfall signals. I would expect strong influence of non-linearities in the extreme cases when ponding of water occurs because the flow rate exceeds the conductivity of the soil or when pressure heads go to extremely high negative values due to upwards flow.

We thank the Referee for raising this point. The non-linearities we mention several times in the article refer to the inherent nature of the Richards equation, which is solved by the model. These non-linearities are enhanced by the fact that our hillslope, as opposed to many other previous applications of data assimilation with physics-based models, is characterized by dominant unsaturated conditions. This point has been clarified in Section 3.1 of the revised manuscript.

Amended – section 3.1 page 4 lines 27 - 30

3) Section 2: I think it would be good to write already at this place about the evaporation. In Section 2 evaporation is not mentioned, so one assumes that it is neglected. Only in Section 4.1 does one read about the rates. It would also be interesting to learn how they were measured. Or were they estimated? In this case: How were they estimated?

We thank the Referee for spotting this: the evaporation rate has been measured throughout the experiment by an atmometer. The average measured rate is quite small (< 1 mm/day), consistent with weather and period of the year (November). We included this information in Section 2 of the revised manuscript.\\

Amended – section 2 page 3 line 31 and 33, page 4 lines 1 and 2

4) Section 4.1: How were seepage face boundary conditions realized? By imposing a water pressure of zero or by imposing zero gradients across the interface?

In the CATHY model, seepage face boundary conditions are realized by imposing atmospheric pressure (ψ = 0 m) at the nodes below the "exit point" (i.e., the intersection of the water table with the interface) and zero flux at the nodes above. Additional details can be found in Camporese et al. (2010), already cited in the paper.

5) Eq. (1): The left bracket on the right hand side should open after the hydraulic conductivity.

Equation (1) has been amended.\\ page 4 Equation 1

6) Section 4.1: It did not become clear to me how the initial condition was set in the simulations. From Table 1 and eq. (1) and from lines 23-24 on page 6 I would think that a constant pressure head was set in the whole domain. If this is so, I do not understand why this initial condition was chosen instead of a hydrostatic pressure distribution. With a constant pressure head, the system is far away from any equilibrium and I would assume that a long spin-up time is needed before a good assimilation run is possible. Either water has to flow out or it has to flow into the domain to achieve equilibrium. If I understand the model setup right, it could only flow in and out from the river of 50 cm water depth on the bottom of the hillslope. This will be a slow process, or not? Maybe this is a misunderstanding, but in this case the initial condition should be explained more clearly.

We thank the Referee for raising this point, which is important because it represents one of the peculiarities of our test case compared to similar studies. At the beginning of the experiment, the entire hillslope was in a highly unsaturated condition, as derived from the tensiometer data. This means that, even without an initially hydrostatic (i.e., equilibrium) profile, water is basically prevented to flow in or out of the hillslope due to the very small values of relative hydraulic conductivity. That is why we opted for a uniform value of initial pressure head, based on the average provided by the measurements. Preliminary analyses showed that the model is not sensitive to this value, because in any case the initial hydraulic conductivity is so small that the hillslope responds only when the first rainfall event occurs and starts wetting the soil. That is why, in this specific case, no warm up or spin up was necessary. We included this discussion in Section 4 of the revised manuscript.

Amended – section 4.2 page 7 lines 10 to 17

7) Section 4.2, parameter transformations (16) - (18): Maybe I missed it but it did not become clear to me what transformation was used for what parameter. Or were all three used for all cases? In this case it would be interesting to learn if any of them works better than the other ones.

Only one single transformation of the three available was used for each of the van Genuchten parameters, according to Carsel and Parrish (1988), where the nominal values and the proper transformation applied to each parameter depend both on the type of soil and the parameter considered. For clarity, we indicated the specific transformation for each parameter in section 4.2.

Amended – section 4.2 page 8 lines 1 and 2

8) Section 5.3: I am not sure that I agree that parameter estimation capabilities are discussed here or could be discussed with the observations at hand, as the true parameters are not really known. I think the important question addressed in this section is rather if parameter updates are useful for data assimilation. The parameters might be optimal for a given situation, but it might be that with more observations, the optimal parameters would be different. It would be interesting to see the parameter updates over time. To my understanding, it is an indicator for reasonable model parameters if the parameter updates converge to a value and do not change with boundary conditions.

We thank the Referee for this comment, which allows us to further clarify this aspect. We agree that true parameters are unknown, as in all real-world systems such as our experimental hillslope; assimilation of real data is indeed one of the strengths of our study. Therefore, the Referee is right in that the purpose of Section 5.3 is to discuss the benefits of parameter estimation in our data assimilation framework.
However, we had some information that can help us assess whether the estimated parameters were consistent with the soil type or not. This was already discussed in the manuscript, but, to address the raised issue, we added a figure in the revised version showing the evolution in time of the updated parameters (Figure 9).
The new Figure 9 shows the time evolution of the saturated hydraulic conductivity and van Genuchten parameters (mean values, solid line, together with minimum and maximum values, in dashed lines, to indicate the ensemble spread) for the two types of soil, sand and clay, for scenario S17 (one of the scenarios reported

in figure 10). The results show that convergence towards stable values (in other words, identifiability) in the data assimilation phase (the first 5 days) is only achieved for the parameters K_s and n of the sand layer, while the dispersion is generally higher for theta_r (to which the model is typically not very sensitive) and α and especially for all the parameters of the clay layer. This can be explained by the fact that all the sensors are located in the sand layer, while no experimental data from the clay layer are assimilated in the filter.

Amended – figura 9 + description in section 5.4 page 12 lines 3 to 12

9) Figure 6 and end of Section 5.3: I have problems seeing a real improvement by parameter updates in Figure 6 in case of the pressure head. Although the uncertainty is reduced with the parameter updates, considering the large discrepancy between measured and simulated values, this reduction is not necessarily an advantage, as the observations are no longer inside of the uncertainty interval.

We agree that visualizing the improvement is not so easy. However, by looking carefully, it can be seen that, in the validation phase, the magenta line (ensemble mean with data assimilation) departs from the black line (ensemble mean in the open loop) and gets closer to the green line (measurements). Also, the plots in figures 6 and 10 only report the results in one single location (P2). In order to make the analysis more objective and comprehensive, we used metrics such as the NRMSE, reported in Table 4, by which we can appreciate the real improvement, calculated over the whole simulation and all the measurement locations, of the data assimilation compared to the open loop.
Amended – section 5.3 page 11 lines 2 and 3

10) The pressure plots in Figure 6 and also in Figure 9 seem a bit odd to me. From estimating roughly from Figure 1, I would expect that the water table in the hillslope should be about 50 cm above ground. The sensor for P2 should be about 40 cm above the water table (only guessing, this is not so clear from the sketch). Without rainfall, the pressure head should in this case be -40 cm. It is clear that it increases during infiltration, but the hydrostatic condition would be in this range. The observations show a lower value, but the simulations show a value of -20 cm. This should be much too high. In the validation period, it seems that the pressure head in the model is falling after the rainfall has stopped and it seems not to have reached an equilibrium at the end (in contrast to the observations). Could this be an effect of the initial condition and the spin-up is not finished?

This point is related to previous points 2 and 6. As previously mentioned in our replies to these points, due to the highly unsaturated initial conditions and associated non-linearities of the flow processes, it is difficult to make predictions on the response of the hillslope based on assumptions of linearity. Overall, the pressure plots in Figures 6 and 10 are consistent with the pressure measurements and, after all, the goal of data assimilation is to include observation information (including uncertainties) into model predictions.
Finally, please note that the channel at the toe of the hillslope is never full of water, as it drains out from the opening visible in the plan view of Figure 1. As a consequence, the perched water table that forms as a result of infiltration never exceeds a value of about 20 cm above the sand-clay interface.

11) Section 5.4 first paragraph: Is it so surprising that updating the van Genuchten parameters has a strong impact on water content predictions but updating only Ks not? The water content is related to the primary variables of the model (pressure head) via these parameters, so I do not find the result so surprising.

We agree that these results are not surprising. Nevertheless, they are discussed as part of the physical rationale behind the main point of this Section, which is to describe the tradeoffs in the results stemming from different data assimilation and updating strategies.

12) Page 10, line 31-32: Why is this point shown once more? I think it is an important point to make, but I do not think that it has been made before. Or do you mean that it has been made before in other papers? There remains an open question: How would one proceed in this situation in the best way? In reality, it is very unlikely that all soil zones could be probed sufficiently. So how does one deal with heterogeneous structures that are not covered by observations? I do not think that this question should (or could) be answered, but it is an interesting point.

We thank the Referee for giving us the opportunity to better clarify our position on this point, which was already mentioned (that is why we write “once more” in line 1 page 12). Indeed, we agree that it is unlikely, especially when dealing with large heterogeneous structures, to have every soil zone properly probed. Within this context, it is even more important to assess whether (or not, as in this case) multivariate data assimilation approaches are capable to compensate for the lack of distributed observations with alternative sources of information, such as, in this case, subsurface outflow.

Amended – section 5.4 page 12 lines from 1 to 12

13) Last paragraph of Section 5.4 and results and discussion: In general, the tradeoffs are described, but not really discussed much. Can one understand this behavior so that one could draw general conclusions? Otherwise it is not so clear if the results are specific for the case that is here studied. I find it also remarkable that including pressure head observations leads to reasonable pressure predictions, while water content is poor, and vice versa. This behavior should be linked to the van Genuchten Parameters, which must be poorly matched in these cases. If all observations are assimilated, both predictions are reasonable. Can one see that in the van Genuchten parameters? Are they improved if all observations are used? Again, it would be interesting to see parameter updates.

As mentioned in our reply to point 8, the parameter updates are shown in a new figure, from which we can see that the only van Genuchten parameter that can be clearly identified is n for the sand layer. However, we cannot match it to any real value, as this is not a synthetic test case. Therefore, we have to accept a relatively large residual uncertainty on the retention curve parameters.
By comparing the final ensembles of the sand parameters in S17 with those in S15 (Figure 8), it is also interesting to note that the main difference is given by the estimated values of the saturated hydraulic conductivity, which is strongly affected by the assimilation of pressure head measurements.
We expect our conclusions to remain valid on most applications of soil hydrology with similar characteristics.
Amended – section 5.4 page 12 lines from 1 to 12

References

Carsel, R. F. and Parrish, R. S.: Developing joint probability distributions of soil water retention characteristics, Water Resources Research, 24, 755–769, doi:10.1029/WR024i005p00755, http://dx.doi.org/10.1029/WR024i005p00755, 1988.

Reply to Anonymous Referee #3 – changes in the manuscript are highlighted in orange.

1. Page 2, line 32. It would be good to elaborate on the trade-off problem in the introduction and refer to other studies that have investigated this problem, such as the recent studies by Zhang et al. (2016) and Zhang et al. (2018).

We thank the Reviewer for suggesting further interesting references about the topic, we added to the manuscript, together with a brief discussion of the trade-off issue (see subsection 5.4 of the Results section).

Amended – see Introduction (page 2 lines 34 and 35, page 3 lines 1 and 2) and subsection 5.4. (page 12 lines from 22 to 26)

2. Page 3, line 25. Could you include a brief description or include a reference on the coefficient of uniformity used here.

Thank you for highlighting this aspect. The coefficient of uniformity is discussed in Lora et al., 2016.

Amended Page 3 line 33

3. Page 5, line 10. Use of dampening factor in the Kalman filter update is debatable. It is a factor that needs to be introduced to compensate for improper settings in the Kalman filter, including model and measurement uncertainty descriptions and ensemble approximations. Discussion of these issues should be included.

We thank the Referee for raising this good point. We tried to clarify and highlight better this aspect in the revised manuscript, also with relevant references, as reported also in the reply to Referee 1 [Page 8 line 24, 25, 26 and 27].

Amended – see also reply to Referee [page 5 lines 26, 27 and 28)

4. Section 4.2. It is not clear how the different model and measurement uncertainty parameters have been estimated. Are they based on preliminary sensitivity analyses?

We thank the Referee for highlighting this aspect. Measurement uncertainties have been estimated based on the accuracy of the sensors (tensiometers for the pressure head, water content reflectometers for the water content and tipping-bucket flow gauges for the outflow discharge) and plausible positioning errors. Uncertainties on model parameters and boundary conditions have been estimated mainly based on previous modelling experiences (Camporese et al., 2009a, 2009b) and prior characterization of the soils in the hillslope. In particular, for the Van Genuchten parameters some preliminary sensitivity analyses have also been performed. We better clarified these aspects in the revised version of the paper.

Amended page 6 line 27 and page 7 lines 1 and 2 and page 8 lines 18 and 19 and page 9 lines 18 and 19

5. Page 7, line 23-26. Why are the measurement error covariance matrix, anomalies and innovation vector normalized? There should be no need to normalize with the EnKF.

We thank the Reviewer for the question. In the paper, several scenarios have been carried out where different variables, i.e., pressure head, water content and outflow discharge, have been jointly assimilated. In this case, as the assimilated variables can differ from each other by orders of magnitude, the normalization of the measurements ensures that the covariance matrices in the Kalman gain are not ill-conditioned. Please see Evensen (2003) and Camporese et al. (2009b) for more details.

Amended – section 4.2 page 8 lines 22 and 23

6. Page 8, line 13-14. Instead of the normalization of RMSE used, one could normalize the RMSE by the Nash-Sutcliffe coefficient. That measure would be more appropriate for interpretation of the actual performance.

We thank the Reviewer for highlighting this aspect. Please note that our objective here is not to assess how well the data are matched but rather to evaluate the benefits of different data assimilation scenarios in comparison to each other (and the open loops). To do so, we elected to use the RMSE as this index is traditionally applied in groundwater hydrology analyses, while the Nash-Sutcliffe efficiency is more popular in rainfall-runoff models.

7. Section 5.4. Would be good to discuss these results in relation to other observations of trade-offs reported in the literature, such as in Zhang et al. (2016) and Zhang et al. (2018).

We thank the Reviewer for the comment. As previously noted (point 1), we introduced the suggested references into the paper not only in the Introduction but also in the Results section where the different trade-offs are analyzed.

Amended – section 5.4 page 12 lines from 22 to 26

References

Camporese, M., C. Paniconi, M. Putti, and P. Salandin. 2009a. Comparison of Data Assimilation Techniques for a Coupled Model of Surface and Subsurface Flow. Vadose Zone J. 8:837-845. doi:10.2136/vzj2009.0018

Camporese, M., C. Paniconi, M. Putti, and P. Salandin (2009b), Ensemble Kalman filter data assimilation for a process - based catchment scale model of surface and subsurface flow, Water Resour. Res., 45, W10421, doi: 10.1029/2008WR007031.

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean Dyn., 53, 343–367.

Lora, M., Camporese, M., and Salandin, P.: Design and performance of a nozzle-type rainfall simulator for landslide triggering experiments,CATENA, 140, 77 – 89, doi:10.1016/j.catena.2016.01.018, 2016.

Zhang, D., Madsen, H., Ridler, Marc, E., Kidmose, J., and Jensen, Karsten, H.: Multivariate hydrological data assimilation of soil moisture and groundwater head, Hydrology and Earth System Sciences, 2016.

Zhang, H., Kurtz, W., Kollet, S., Vereecken, H., and Franssen, H.-J. H.: Comparison of different assimilation methodologies of groundwater levels to improve predictions of root zone soil moisture with an integrated terrestrial system model, Advances in Water Resources, 111, 224 – 238, doi:https://doi.org/10.1016/j.advwatres.2017.11.003, http://www.sciencedirect.com/science/article/pii/S0309170817304888, 2018.

Multi-source data assimilation for physically-based hydrological modeling of an experimental hillslope

Anna Botto[1], Enrica Belluco[1], and Matteo Camporese[1]

[1]Department of Civil, Environmental and Architectural Engineering, University of Padova, Italy

Correspondence to: Anna Botto (anna.botto@unipd.it)

Abstract.

Data assimilation has been recently the focus of much attention for integrated surface-subsurface hydrological models, whereby joint assimilation of water table, soil moisture, and river discharge measurements with the ensemble Kalman filter (EnKF) have been extensively applied. Although the EnKF has been specifically developed to deal with nonlinear models, integrated hydrological models based on the Richards equation still represent a challenge, due to strong nonlinearities that may significantly affect the filter performance. Thus, more studies are needed to investigate the capabilities of the EnKF to correct the system state and identify parameters in cases where the unsaturated zone dynamics are dominant, as well as to quantify possible tradeoffs associated with assimilation of multi-source data. Here, the model CATHY (CATchment HYdrology) is applied to reproduce the hydrological dynamics observed in an experimental two-layered hillslope, equipped with tensiometers, water content reflectometer probes, and tipping bucket flow gages to monitor the hillslope response to a series of artificial rainfall events. Pressure head, soil moisture, and subsurface outflow are assimilated with the EnKF in a number of scenarios and the challenges and issues arising from the assimilation of multi-source data in this real-world test case are discussed. Our results demonstrate that the EnKF is able to effectively correct states and parameters even in a real application characterized by strong nonlinearities. However, multi-source data assimilation may lead to significant tradeoffs: the assimilation of additional variables can lead to degradation of model predictions for other variables that ~~were~~ are otherwise well reproduced. Furthermore, we show that integrated observations such as outflow discharge cannot compensate for the lack of well-distributed data in heterogeneous hillslopes.

1 Introduction

Data assimilation, i.e., the process in which observations of a system are merged in a consistent manner with numerical model predictions (e.g., Troch et al., 2003; Moradkhani, 2008), has become increasingly popular in hydrological modeling over the last few decades (Montzka et al., 2012). Among the various techniques available, the ensemble Kalman filter (EnKF) (Evensen, 2003, 2009b) is probably the most widespread, thanks to its ease of implementation, capability to handle nonlinear models and potential to be used as a sequential inverse modeling tool when parameters are included in the update step. Applications in hydrology include studies in different disciplines, such as groundwater hydrology (e.g., Chen and Zhang, 2006; Hendricks Franssen and Kinzelbach, 2008; Bailey and Baù, 2010; Li et al., 2012; Zovi et al., 2017), rainfall-runoff modeling

(e.g., Moradkhani et al., 2005; Vrugt et al., 2006; Weerts and El Serafy, 2006; Clark et al., 2008; Xie and Zhang, 2010; Han and Li, 2008), and land surface modeling at multiple scales (e.g., Reichle et al., 2002a, b; Crow and Wood, 2003; Francois et al., 2003; Pan and Wood, 2006; De Lannoy et al., 2007; Cammalleri and Ciraolo, 2012; Flores et al., 2012; Hain et al., 2012).

As a consequence of such popularity, the EnKF is more and more applied also with integrated surface-subsurface hydrological models (IHSSMs), whereby multiple terrestrial compartments (e.g., snow cover, surface water, groundwater) are solved simultaneously, in an attempt to tackle environmental problems in a holistic approach (Maxwell et al., 2014; Kollet et al., 2017). For instance, Camporese et al. (2009a) and Camporese et al. (2009b) combined the CATHY model and the EnKF to assimilate pressure head, soil moisture, and streamflow data, finding that the assimilation of pressure head and soil moisture is beneficial to subsurface states and river discharge, but the assimilation of river discharge alone does not improve the prediction of subsurface states. Similar conclusions were drawn by Pasetto et al. (2012), who compared the EnKF with a modified particle filter to assimilate discharge and pressure head, ~~again with~~ using the CATHY model. More recently, Pasetto et al. (2015) used the EnKF and CATHY to investigate the impact of possible sensor failure on the observability of flow dynamics and estimation of the model parameters characterizing the soil properties of an artificial hillslope. Ridler et al. (2014) assimilated soil moisture remote sensing products in the MIKE SHE model and found that surface soil moisture has correction capabilities limited to the first 25 cm of soil. With the same model, Rasmussen et al. (2015a), Rasmussen et al. (2015b), Zhang et al. (2015), and Zhang et al. (2016) investigated in detail issues related to uncertainty quantification and biased observations, as well the impacts of update localization and ensemble size on the multivariate assimilation of groundwater head and river discharge at the catchment scale. Kurtz et al. (2015) first presented a data assimilation framework for the land surface-subsurface part of the Terrestrial System Modelling Platform (TerrSysMP), followed by Baatz et al. (2017), who assimilated distributed river discharge data into the TerrSysMP to estimate the spatially distributed Manning's roughness coefficient and Zhang et al. (2018), who tested and compared five data assimilation methodologies for assimilating groundwater level data via the EnKF to improve root zone soil moisture estimation. Within yet another modeling framework, Tang et al. (2017) used the EnKF in conjunction with HydroGeoSphere to study the influence of heterogeneous riverbeds on river-aquifer exchange fluxes.

In spite of such a strong interest, several issues related to the use of EnKF for state and parameter estimation in integrated hydrological modeling remain unresolved. The subsurface component of many IHSSMs is based on the solution of the Richards equation in one or three dimensions and, although recent studies with numerical experiments in synthetic test cases (Erdal et al., 2014; Brandhorst et al., 2017) have shown that the EnKF has great potential for the estimation of soil hydraulic parameters in the unsaturated zone, it is still unclear whether the method is able to cope with nonlinearities and parameter estimation in real test cases, where multiple uncertainties on initial and boundary conditions make the problem much more challenging (e.g., Visser et al., 2006; De Lannoy et al., 2007; Monsivais-Huertero et al., 2010; Shi et al., 2015; Bauser et al., 2016). Also, as more sources of data become available at cheaper costs, it is increasingly difficult to assess which data types are the most suitable or effective to assimilate and ~~in general to assess possible tradeoffs related to multivariate data assimilation.~~ and to check which possible tradeoffs might occur when assimilating different variables in a multivariate data assimilation framework. Zhang et al. (2018) , for instance, found that joint assimilation of pressure head and soil moisture is beneficial only when pressure head is assimilated in the saturated zone and soil moisture in the unsaturated zone, while Zhang et al. (2016) showed that joint

assimilation of groundwater head and water content fails to provide reasonable results if proper countermeasures to spurious correlations are not adopted.

Within this context, the main goals of the present study are: i) to assess whether the EnKF in combination with a Richards equation-based hydrological model is able to effectively improve states and parameters in a real-world test case characterized by dominant unsaturated dynamics and ii) to quantify the tradeoffs associated to multi-source data assimilation.

To pursue these goals, the EnKF is used in combination with the CATHY (CATtchment HYdrology) model (Camporese et al., 2010) to assimilate real observations of pressure head, soil moisture and subsurface outflow collected during a controlled experiment carried out in an artificial hillslope. The experiment is characterized by strong nonlinearities, due to the dominant unsaturated dynamics, but the strictly controlled conditions, as opposed to field studies, allow us to minimize the effects of initial and boundary conditions uncertainty ~~on the problem at hand~~. The behavior and performance of the EnKF-based assimilation framework in terms of its ability to retrieve the correct hillslope response are evaluated in a number of data assimilation scenarios, characterized by different combinations of assimilated and updated variables. In each scenario, a significant part of the simulation is devoted to the validation of the model (i.e., with no data assimilation), to assess the impacts of parameter updating on the model predictions also during periods without observations.

2 The hillslope experiment

The artificial hillslope is placed inside a concrete structure of lenght 6 m, width 2 m and height varying linearly from 3.5 m to 0.5 m, corresponding to a slope of 32 degrees (Figure 1). A total of fifty apertures, which can be kept closed with screw cups when needed, allow the positioning of various monitoring sensors in properly chosen positions on each lateral wall of the structure. The hillslope toe is made with a hollow-brick porous wall, in order to allow subsurface water to drain. The soil ~~was~~ is placed inside the structure to mimic a two-layered hillslope. A uniform 60 cm-thick silty fine sand ~~was~~ is deployed on top of a ~~relatively im~~ low-permeable basement made of sandy clay soil. More details on the soil properties can be found in Lora et al. (2016a) and Schenato et al. (2017). In the following, we will refer to the two soil types simply as sand and clay.

Six tensiometers and six water content reflectometer (WCR) probes are used to measure pressure head and water content in the top soil layer. All the sensors are located in an intermediate position of the hillslope, as shown in Figure 1, which reports a plan view and longitudinal cross-section along with the six positions where each tensiometer has been installed in front of the corresponding WCR. Two tipping-bucket flow gages are placed at the toe of the hillslope to measure surface runoff and subsurface outflow, while the rain is generated by a rainfall simulator that can produce relatively uniform rainfall intensities varying from 50 to 150 mm/h (Lora et al., 2016a).

A Campbell Scientific (CR 1000) data logger ~~has been~~ is used to collect and record all the data with a frequency of 0.5 Hz during a 12-day experiment, ~~which has been~~ carried out by generating rainfall events of different duration alternated with recession periods (no rainfall, evaporation only). Before the experiment, two preliminary tests ~~have been~~ were performed to check the intensity and the uniformity of the rainfall rate. The coefficient of uniformity of the preliminary tests ~~resulted~~ are equal to 72%, for a mean rate of 58.8 mm/h (Lora et al., 2016a). The evaporation rate has been measured throughout the

experiment by an atmometer. The average measured rate is quite small (< 1 mm/day), consistent with weather and period of the year (November 2016).

Figure 2 reports all the data collected during the experiment. Note that the tensiometers P1 and P4 were affected by malfunctioning, therefore their data are not reported and ~~will not be~~ were not used in the data assimilation simulations.

3 Numerical methods

3.1 The CATHY model

The CATHY (CATchment HYdrology) model (Camporese et al., 2010) is a physics-based hydrological model capable of simulating integrated subsurface, overland and channel water flow. The model combines a Richards equation solver for the three-dimensional flow in variably saturated porous media with a surface water flow module for the solution of the one-dimensional diffusion wave approximation of the De Saint-Venant equation. In this study, however, there is no surface runoff. Therefore, only subsurface flow is considered, according to the following form of Richards equation:

$$S_w S_s \frac{\partial \psi}{\partial t} + \phi \frac{\partial S_w}{\partial t} = \nabla [K_s K_r (\nabla \psi + \eta_z)] + q_s. \tag{1}$$

In equation (1), $S_w = \theta/\phi$ is water saturation, θ and ϕ being the volumetric soil water content and porosity $[/]$, respectively, S_s is the specific storage coefficient $[L^{-1}]$, ψ is the pressure head $[L]$, t is time $[T]$, ∇ is the gradient operator, K_s is the saturated hydraulic conductivity tensor $[L/T]$, K_r is the relative hydraulic conductivity function $[/]$, $\eta_z = (0,0,1)^T$ is the vertical direction vector, z is the vertical coordinate directed upward $[L]$, and q_s represents distributed source or sink terms $[L^3/L^3T]$.

The unsaturated hydraulic properties are taken into account by means of the van Genuchten functions (e.g., Wösten and van Genuchten (1988)) $S_w(\psi)$ and $K_r(\psi)$:

$$S_w = S_{wr} + \frac{1 - S_{wr}}{[1 + (\alpha|\psi|)^n]^m}, \tag{2}$$

$$K_r = \left(\frac{S_w - S_{wr}}{1 - S_{wr}} \right)^{0.5} \left\{ 1 - \left[1 - \left(\frac{S_w - S_{wr}}{1 - S_{wr}} \right)^{\frac{1}{m}} \right]^m \right\}^2, \tag{3}$$

where $S_{wr} = \theta_r/\phi$ is the residual water saturation, with θ_r the residual water content, α is an empirical constant $[L^{-1}]$ related to the inverse of the air entry suction, while the dimensionless shape parameters n and m are linked by the expression $m = 1 - 1/n$. The model solves Equation (1) by means of Galerkin finite elements with tetrahedral elements and linear basis functions in space and weighted finite differences for integration in time (Camporese et al., 2010).

It is worth noting that the Richards equation is strongly nonlinear, due to the retention curves (Equations 2 and 3). Such nonlinearities are enhanced in this study by the fact that the hillslope is characterized by dominant unsaturated conditions (Figure 2), as opposed to many other previous applications of data assimilation with physics-based models. This makes our hillslope experiment particularly challenging and distinctive from both modeling and data assimilation perspectives.

3.2 The ensemble Kalman filter

The ensemble Kalman filter (Evensen, 2003, 2009a, b) is a sequential data assimilation scheme, in which states (and parameters) are sequentially updated based on a Monte Carlo approximation of the covariance matrices needed in the standard Kalman filter. The process is Markovian of the first order and the implementation of the EnKF does not require the linearization of the model, making it particularly suitable to handle nonlinear problems.

In this paper the EnKF is implemented according to the numerical formulation proposed by Sakov et al. (2010). Let $\mathbf{X}$ be an ensemble matrix of M rows and N columns, where N is the number of realizations and M is the state dimension, i.e., the number of nodes in the finite element grid, augmented by the number of parameters that are subject to update. The main idea behind this type of implementation is that the matrix $\mathbf{X}$ can be defined as the sum, $\mathbf{x} + \mathbf{A}$, of the ensemble average, $\mathbf{x}$,

$$\mathbf{x} = \frac{1}{N}\mathbf{X}\mathbf{1}, \tag{4}$$

and the matrix of ensemble anomalies, $\mathbf{A}$,

$$\mathbf{A} = \mathbf{X}(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T), \tag{5}$$

where $\mathbf{1}$ and $\mathbf{I}$ are a vector with all elements equal to one and the identity matrix, respectively. As usual, superscript T denotes matrix transposition.

Whenever observed data are available, the EnKF can compute the updated matrix $\mathbf{X}^u$ as the sum of the updated ensemble mean, $\mathbf{x}^u$, and the updated anomalies, $\mathbf{A}^u$,

$$\mathbf{X}^u = \mathbf{x}^u + \mathbf{A}^u. \tag{6}$$

In the following, the lack of a superscript u denotes that the matrix or vector is computed at the forecast stage, i.e., at the previous model time step.

The updated mean can be calculated as

$$\mathbf{x}^u = \mathbf{x} + \boldsymbol{\beta}\mathbf{A}\mathbf{G}\mathbf{s}, \tag{7}$$

where $\boldsymbol{\beta}$ is a diagonal matrix of dampening factors (Hendricks Franssen and Kinzelbach, 2008), whose elements vary from 0 to 1, $\mathbf{s}$ is the scaled innovation vector

$$\mathbf{s} = \mathbf{R}^{-1/2}(\mathbf{D} - \mathbf{H}\mathbf{x})/\sqrt{N-1}, \tag{8}$$

which depends on the measurement error covariance matrix, $\mathbf{R}$, and on the difference between the measurements, $\mathbf{D}$, and the ensemble mean of the simulated observations, $\mathbf{H}\mathbf{x}$. Dampening factors help to prevent the occurrence of unstable updates and filter divergence (e.g., Evensen (2009b)) and were chosen as an alternative of covariance inflation due to ease of implementation.

The matrix $\mathbf{G}$ is defined as

$$\mathbf{G} = \mathbf{M}\mathbf{S}^T, \tag{9}$$

where $\mathbf{S}$ is the matrix of scaled ensemble innovation anomalies

$$\mathbf{S} = \mathbf{R}^{-1/2}\mathbf{HA}/\sqrt{N-1}, \tag{10}$$

$\mathbf{HA}$ being the simulated measurement anomalies

$$\mathbf{HA} = \mathbf{HX}(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T), \tag{11}$$

and $\mathbf{M}$ being defined as

$$\mathbf{M} = (\mathbf{I} + \mathbf{S}^T\mathbf{S})^{-1}. \tag{12}$$

The updated anomalies, $\mathbf{A}^u$, are computed as

$$\mathbf{A}^u = \mathbf{A} + \beta\mathbf{A}(\mathbf{M}^{1/2} - \mathbf{I}). \tag{13}$$

When updating the states only, the elements of $\mathbf{X}$ are the pressure heads at each node of the finite element grid, while the state augmentation technique is used when updating also the parameters. In this latter case, the desired parameters (e.g., hydraulic conductivity, parameters of the retention curves), transformed as described in section 4.2, are added to $\mathbf{X}$ and updated based on their correlation with the system states (e.g., Erdal et al., 2015).

4 Model and data assimilation setup

4.1 CATHY setup

The artificial hillslope is discretized with a surface triangular grid resulting from the subdivision of square cells of 10 cm side. The triangular grid is then replicated vertically for a total of 25 layers to generate the three-dimensional tetrahedral mesh (Figure 3). Fifteen layers are used to represent the top sand, while 10 layers discretize the clay. No flow boundary conditions are assumed at each boundary, except for the subsurface outflow section, where seepage face boundary conditions are used, and the surface, where, time-variable rainfall or evaporation rates are imposed. Finally, the soil hydraulic parameters are assigned as reported in Table 1. The values of saturated hydraulic conductivity and van Genuchten retention parameters are perturbed to generate the ensemble of realizations as described in the following section, ~~while~~ whereas soil porosities and specific storages are considered deterministic, ~~as~~ the former ~~were~~ being well characterized with laboratory tests ~~while the CATHY model is usually not very sensitive to the latter.~~ and the latter having little impact on the CATHY model response in this case.

4.2 EnKF setup

In order to generate the ensemble of realizations needed for the application of the EnKF, we perturb the atmospheric forcing (i.e., rainfall and evaporation rates), soil properties and initial conditions. Table 1 reports a summary of the perturbed variables, along with their nominal mean values as well as the nature and statistics of the perturbations. The uncertainties on model

parameters and boundary conditions have been assigned on the bases of previous modeling experiences and preliminary characterization of the soils in the hillslope (Camporese et al., 2009a, b; Lora et al., 2016b).

The ensemble of time-variable atmospheric forcing rates was generated with a sequence of multiplicative perturbations, $\mathbf{q_k}$, correlated in time as in Evensen (2003),

$$\mathbf{q}_k = \gamma \mathbf{q}_{k-1} + \sqrt{1-\gamma^2}\mathbf{w}_{k-1}, \tag{14}$$

where the subscript k is the time index, $\mathbf{w_k}$ is a sequence of white noise drawn from the standard normal distribution, and the coefficient γ is computed as

$$\gamma = 1 - \frac{\Delta t}{\tau}, \tag{15}$$

Δt being the assimilation interval and τ the specified time decorrelation length, here set equal to 108,000 s, i.e., 30 h.

The initial conditions consist of a uniform value of pressure head, ψ_0, whose nominal ensemble mean is -0.67 m, based on the average provided by the tensiometer measurements . We opted for a uniform value of initial pressure head because, at the beginning of the experiment, the entire hillslope was in a highly unsaturated condition, as derived from the tensiometer data. This means that, even without an initially hydrostatic (i.e., equilibrium) profile, water is basically prevented to flow in or out of the hillslope due to the very small values of relative hydraulic conductivity. This is one of the peculiarities of our test case compared to similar studies. Preliminary analyses showed that the model is not very sensitive to ψ_0, because in any case the initial hydraulic conductivity is so small that the hillslope responds only when the first rainfall event occurs and starts wetting the soil. Therefore, in this study, no warm up or spin up was necessary. The ensemble of ψ_0 values is generated by additive perturbations normally distributed with mean equal to 0 and standard deviation equal to 0.2 m.

Perturbed soil parameters, for both sand and clay, include the saturated hydraulic conductivity as well as the parameters of the van Genuchten retention curves. Table 1 reports the nominal mean values of K_s, based on soil samples analyzed in the laboratory, as well as the parameters used for generating their ensembles. The saturated hydraulic conductivities of sand and clay are perturbed independently from each other with multiplicative perturbations sampled from a lognormal distribution.

The parameters of the van Genuchten retention curves α, n, and θ_r are perturbed taking into account their mutual correlation according to Carsel and Parrish (1988), who described their statistics and transformed them into normally distributed variables via the Johnson system (Johnson, 1970; Bertino et al., 2003). Three main transformation functions are available, i.e, the lognormal (LN), log-ratio (SB), and hyperbolic (SU):

$$LN: Y = ln(V) \tag{16}$$

$$SB: Y = ln[(V-A)/(B-V)] = ln(U) \tag{17}$$

$$SU: Y = sinh^{-1}(U), \tag{18}$$

where V denotes the parameter before transformation, bounded within the range $[A\ B]$, and Y denotes the transformed parameter with normal distribution. In this work, the prior statistics of the van Genuchten parameters are taken from the soil types "sandy loam" and "silt loam" in Carsel and Parrish (1988), assumed as valid representations of our sand and clay, respectively.

In particular, the SB transformation has been used for all the parameters, except for n of the sandy loam and α of the silt loam, for which the LN transformation has been applied. The ensemble of transformed parameters is generated by

$$\mathbf{y} = \mathbf{u} + \mathbf{T}^{\mathbf{T}}\mathbf{z}, \tag{19}$$

where $\mathbf{y}$ is the vector containing transformed α, n, and θ_r, $\mathbf{u}$ contains the transformed variable means, $\mathbf{T}$ ~~is the factored covariance matrix~~ is an upper diagonal matrix, obtained from the Choleski factorization of the covariance matrix of the three parameters (see Table 2), and $\mathbf{z}$ is a vector of normal deviates with mean equal to 0 and standard deviation σ_{VG}. The transformed variables can be included in the matrix $\mathbf{X}$ when updating the parameters, and can then be back-transformed in order to obtain the updated values of the soil retention parameters.

The EnKF algorithm implemented here is actually an ensemble transform Kalman filter (Bishop et al., 2001) that does not require the perturbation of observations. ~~However~~ On the other hand, the measurement error covariance matrix, $\mathbf{R}$, must be assumed to be known a priori. In this work, $\mathbf{R}$ was estimated directly from the measurements, taking advantage of the high time resolution of the collected data. Pressure head and water content data were collected every 2 s and averaged every 10 min over a 40 s window to obtain the observations to assimilate. Over the same time window, the data were linearly detrended and ~~the residuals used to calculate the correlation coefficients between all pairs of observations.~~ the residuals were used to calculate the correlation coefficients between all pairs of observations. The final covariance matrices were then assembled multiplying the correlation coefficients by the relevant standard deviations, assumed as 0.05 m and 0.025 for pressure head and water content, respectively. The subsurface outflow measurements are assumed independent from the pressure head and water content data, with a standard deviation equal to 8% of the measured discharge. All measurement error standard deviations have been estimated based on the accuracy of the sensors and plausible positioning errors.

When assimilating multiple variables, proper normalization of the measurement error covariance matrices, anomalies of the simulated data, and innovation vectors were performed, using values of 0.6 m, 0.58, and 4.17×10^{-5} m^3/s for pressure head, water content and subsurface outflow, respectively. The normalization ensures that in multivariate assimilation scenarios the covariance matrices in the Kalman gain are not ill-conditioned (Evensen, 2003; Camporese et al., 2009b).

4.3 Data assimilation scenarios

A total of 17 data assimilation scenarios have been simulated, whereby ~~we varied~~ the assimilation interval, the assimilated variables, the updated variables, and the uncertainty on the van Genuchten ~~retention~~ parameters were varied. Table 3 reports a summary of the main characteristics for each scenario. Assimilated variables may include water content only or with subsurface outflow, pressure head only or with subsurface outflow, and all three variables together. ~~As regards~~ With regard to the updated variables, three cases have been analyzed: update of the state variables only; update of the state variables and saturated hydraulic conductivities for both sand and clay; update of the state variables, hydraulic conductivities and van Genuchten parameters, for both sand and clay. In all the data assimilation scenarios, observations were assimilated only during the first five days of simulation, leaving the final seven days as a validation period, during which the ensemble was let to evolve freely. For comparison, two open loop simulations, i.e., without data assimilation, have also been carried out (Table 3).

The performance of the simulations has been evaluated by means of the root mean square error ($RMSE$), computed for the different variables, i.e., pressure head, water content, and subsurface outflow. The root mean square error is calculated as

$$RMSE(t) = \frac{1}{N_o}\sum_{j=1}^{N_o}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\Big(S_{i,j}(t) - O_j(t)\Big)^2}, \tag{20}$$

where N_o is the number of observations available (six, four, and one for water content, pressure head, and subsurface outflow, respectively), $S_{i,j}$ refers to the simulated results of the ith realization of the ensemble at the location of the jth observation and O_j is the corresponding experimental value. To obtain a meaningful comparison between the errors of different variables, we also compute the time-averaged normalized root mean square error, $NRMSE$,

$$NRMSE = \frac{1}{N_T}\sum_{k=1}^{N_T}\frac{RMSE_k}{NF}, \tag{21}$$

where N_T is the number of time steps, k is the time index, and NF is a normalization factor equal to 0.58, 0.60 m, and 4.17×10^{-5} m^3/s (i.e., 2.5 l/min) for the water content, pressure head, and subsurface discharge, respectively. The $NRMSE$ is computed separately for each variable and for the assimilation ($N_T = 120$) and validation ($N_T = 167$) periods, but also as a global index of performance averaged over all the variables and the two periods.

5 Results and discussion

5.1 Preliminary simulations

A preliminary sensitivity analysis over a number of EnKF parameters has been performed, in order to select a final and satisfactory setup for the subsequent data assimilation scenarios. First, simulations with N equal to 32, 128 and 256 have been performed and it has been found that an ensemble size of 128 ensures a good tradeoff between performance and computational effort. Some preliminary sensitivity analyses on the value of σ_{VG} have also been performed, resulting in values of 0.1 and 0.25 for the scenarios with and without update of van Genuchten parameters, respectively (Table 3).

Then, several dampening factor values ($\boldsymbol{\beta}$ in equations (7) and (13)) have been ~~tried~~ tested, including combinations of different values for the update of system state and parameters. Based on this analysis, a value of 1 has been chosen for the update of the system state, whereas a dampening factor equal to 0.5 has been selected for the update of the soil hydraulic parameters (K_s, α, n, and θ_r, for both sand and clay). This choice of the dampening factors is consistent with previous studies (Brandhorst et al., 2017) and prevents abrupt changes of the retention curve parameters that could lead to difficulties in model convergence and hence loss of realizations.

According to these preliminary analyses, all scenarios reported in Table 3 have been simulated with an ensemble size of 128 and dampening factors of 1 and 0.5 for system state and parameters, respectively.

5.2 Overall EnKF performance

Table 4 and Figure 4 summarize the performance of the EnKF in all the data assimilation scenarios, expressed in terms of $NRMSE$ for the three measured variables (water content, WC, pressure head, PH, and subsurface outflow, Q), and averaged separately over the assimilation and validation windows. A comparison between scenarios 1-5 and 8-12 in Table 4, characterized by the same assimilated and updated variables but different assimilation intervals, show that assimilating more frequently does not always result in significant improvements of model predictions. The variable that benefits ~~more~~ the most from more frequent updates is subsurface outflow, especially in the scenarios where Q is assimilated (e.g., compare scenarios 3-5 with 10-12).

Figure 4 highlights that most of the data assimilation scenarios result in an improvement of model predictions for pressure head and subsurface outflow, compared to the open loop simulations (data pairs below the 45-degree reference line). However, ~~there are~~ in some scenarios, ~~in which~~ the filter performance in predicting the water content is actually worse than in the open loop. ~~From a~~ A close inspection of the values in Table 4 ~~, it is possible to identify~~ indicates that such scenarios ~~as~~ are those where pressure head is assimilated, alone or in conjunction with water content and subsurface outflow. This is ~~probably~~ likely due to a combination of two factors: i) only four out of six tensiometers are available for assimilation, compared to the six WCR probes available for water content, and ii) the pressure head measurements are characterized by relatively poor quality. This can be appreciated from the pressure head data shown in Figure 2, where diurnal disturbances caused by temperature fluctuations (Warrick et al., 1998) are apparent.

5.3 Parameter estimation capabilities

To assess the capabilities and benefits of parameter estimation with the EnKF, it is useful to compare scenarios with the same assimilated variables but different updated variables. Figures 5a, 5c, and 5e show the ratios between $RMSE$ in data assimilation scenarios S6, S8, S13 and the corresponding open loop values for water content, pressure head, and subsurface outflow, respectively. In these three scenarios, water content alone is assimilated, but the updated variables are system state only in S6, system state and saturated hydraulic conductivity in S8, and system state ~~plus~~ , the K_s and van Genuchten parameters α, n, and θ_r in S13. Figure 5a highlights that progressively updating more parameters brings significant improvements in water content prediction over the validation period, with reductions of the $NRMSE$ with respect to the open loop of almost 20% when updating K_s only and 60% when updating also the retention curve parameters. Moreover, updating parameters improves significantly pressure head predictions in validation and subsurface outflow in both assimilation and validation, as shown in Figures 5c and 5e.

The effect of parameter updating on model predictions for scenarios S6, S8, and S13 can be visualized in Figure 6, which shows ensemble means and 90% confidence bands of simulated water content, pressure head, and subsurface outflow in comparison with the experimental data and the corresponding open loop values. In scenario S6, without parameter update, model predictions during the validation tend to converge again to the open loop simulations both in terms of mean and uncertainty (Figures 6a, 6d, 6g). Updating the parameters (scenarios S8 and S13) results in decreased uncertainty during validation (Fig-

ures 6b, 6c, 6e, 6f, 6h, 6i), due to the reduced variability of saturated hydraulic conductivity and van Genuchten parameters. Note also that the update of K_s is particularly beneficial to pressure head (panel a versus b and c, where it can be seen that, in the validation phase, the ensemble mean with data assimilation departs the open loop and gets closer to the measurements) and subsurface outflow (panel g versus h and i), whereas updating α, n, and θ_r improved significantly the water content (panel e versus f).

5.4 Tradeoffs in multi-source data assimilation

We now focus our attention on the scenarios where multi-source data are assimilated. ~~Analogously to t~~ The right panels in Figure 5 show the ratios between $RMSE$ in data assimilation scenarios S7, S10, S15 and the corresponding open loop values for water content, pressure head, and subsurface outflow. Whereas in scenarios S6, S8, and S13 (see left panels Figure 5) water content alone was assimilated, in scenarios S7, S10, and S15 (see right panels Figure 5) water content and subsurface outflow were jointly assimilated. The updated variables are system state only in S7, system state and saturated hydraulic conductivity in S10, and system state plus K_s and van Genuchten parameters α, n, and θ_r in S15. As previously noted for the scenarios with assimilation of water content alone, the effect of parameter updating is significant mainly in the validation phase. However, ~~we~~ one can now observe an increase in validation $RMSE$ of water content when updating also K_s, with a value that exceeds the one of the corresponding open loop. At the same time, there is a decrease in pressure head $RMSE$ of more than 40% with respect to the open loop, indicating that the update of K_s ~~in this case~~ is beneficial to pressure head, but not to water content. Including the update of van Genuchten parameters, in scenario S15, has an opposite effect: model predictions of water content in validation improve dramatically ($RMSE$ of almost 60% less than in the open loop), ~~while~~ whereas pressure head predictions worsen slightly and align with $RMSE$ values of scenario S13. This indicates that the update of the van Genuchten parameters is more important for predictions of water content than pressure head ~~and~~, which represents an interesting example of the ~~kind of the~~ tradeoffs associated with multi-source data assimilation and parameter updating in integrated hydrological models.

Further insights ~~about~~ on the differences between scenarios S10 and S15 can be gained ~~by looking at~~ from Figures 7 and 8, which report the prior and posterior (i.e., at the end of the assimilation period) distributions of soil parameters. Figure 7 reports the results for scenario S10, where K_s only was updated, and shows that the sand K_s is clearly identifiable, whereas the clay K_s is not, with a large residual uncertainty and a mean value that is not consistent with the actual soil type in the hillslope. As no data are available in the clay layer, the large residual variability should be expected, but the bias in the mean value is probably caused by spurious correlations and, perhaps, also by the low sensitivity of the assimilated variables to the clay permeablity. This is likely the reason for the poor model prediction of water content.

Figure 8 shows the prior and posterior distributions of K_s, α, n, and θ_r in S15. Again, for the sand, the saturated hydraulic conductivity can be clearly identified, as well as the exponent n of the van Genuchten retention function, while parameters θ_r and α are more difficult to estimate. As for the clay, posterior uncertainty is large for all the parameters and the mean value of K_s shows again a bias with respect to the prior value, although in this case the final value is more consistent with the actual soil type and this could explain why the water content model predictions improve significantly compared to scenario S10. This

can also be explained by the fact that all the sensors are located in the sand layer, while no experimental data from the clay layer are assimilated.

An additional perspective on parameter estimation is given by Figure 9 , which shows the time evolution of ensemble mean and spread of the saturated hydraulic conductivity and van Genuchten parameters for both sand and clay in scenario S17 (one of the scenarios reported in figure 10) . The convergence toward stable values in the data assimilation phase is only achieved for the parameters K_s and n of the sand layer, while the dispersion remains generally higher for θ_r (to which the model is typically not very sensitive) and α, as well as for all the parameters of the clay layer.

In summary, the results of parameter updating for the clay point out that data would be needed in all the soil layers. However, when dealing with large heterogeneous structures, it is very expensive to have every soil zone properly probed, which is why it is important to assess whether multivariate data assimilation approaches are capable to compensate for the lack of distributed observations with alternative sources of information. Here, an integrated measurement such as the subsurface outflow does not seem to be sufficient to compensate for this lack of representativeness.

Finally, we analyze the tradeoffs in system state predictions associated to multi-source data assimilation for scenarios S15, S16, and S17. Figure 10 shows pressure head in P2, water content in W6, and subsurface outflow as simulated in the data assimilation and open loop scenarios, in terms of ensemble mean and 90% confidence bands, compared to the measurements. In scenario S15 (Figures 10a, 10d, 10g), where water content and subsurface discharge were assimilated, model results are very good for these variables but not so for pressure head (see also $NRMSE$ values in Table 4). On the other hand, in scenario S16, where pressure head and subsurface outflow were assimilated, pressure head and discharge are well simulated, but not water content (Figures 10b, 10e, 10h, and Table 4). Finally, in scenario S17, where all the available data were assimilated, the model predicts well both pressure head and water content, but at the cost of a slightly degraded prediction of subsurface outflow compared to scenarios S15 and S16 (Figures 10c, 10f, 10i, and Table 4).

Similar issues were reported by Zhang et al. (2016) who found that the joint assimilation of soil moisture and groundwater head does not improve model predictions, unless update localization is used. In their study, this was likely caused by unrealistic cross-variable correlations due to limited ensemble sizes, whereas in our case it might be that the relatively poor quality of pressure head measurements, compared to water content observations, and the lack of observations in the clay layer do not allow us to obtain accurate estimates of the van Genuchten parameters. ~~This is another example of tradeoff that can be expected in multi-source data assimilation, especially when, for some reasons, not all the different layers or zones of the hillslope or catchment are monitored.~~

6 Summary and Conclusions

In this study, a Richards equation-based hydrological model, CATHY, has been used with the ensemble Kalman filter to assimilate pressure head, water content, and subsurface outflow data in a real-world test case, represented by an experimental artificial hillslope. A total of 17 data assimilation simulations have been presented and described to provide a comprehensive overview of possible scenarios. Univariate scenarios with the assimilation of water content or pressure head alone were

compared to multivariate cases where water content and pressure head were combined with outflow discharge or where water content, pressure head and outflow discharge were jointly assimilated. Regarding the updating strategies, single (state variable) and joint (state variables plus saturated hydraulic conductivity with and without van Genuchten parameters) updating scenarios were considered.

Overall, the capabilities of the ensemble Kalman filter to jointly correct the system states and soil parameters in physically-based hydrological models were confirmed, even in a real-world test case such as the one presented here, characterized by dominant unsaturated dynamics and hence strong nonlinearities. Updating of the saturated hydraulic conductivity brought significant improvements in the prediction of pressure head and subsurface outflow, while updating the van Genuchten parameters proved to be highly beneficial to the prediction of the water content dynamics. On the other hand, multivariate data assimilation may lead to significant tradeoffs. For instance, the assimilation of soil moisture in addition to pressure head and subsurface outflow improved water content, but slightly degraded the prediction of the outflow discharge. Moreover, our results suggest that high-quality and representative data are essential for a proper and effective use of data assimilation in physically-based hydrological models, as shown by the relatively poor performance of the EnKF in scenarios when pressure head was assimilated, due to temperature disturbances of the data, and by biased estimates of clay parameters, due to the lack of data in this soil layer.

In future studies, more representative data, including observations in the clay, will be assimilated and the possibility to apply bias-aware filters will be considered to compensate for the effect of temperature in the tensiometric data.

Acknowledgements. We gratefully acknowledge the financial support of the University of Padova, through the grant CPDA148790.

References

Baatz, D., Kurtz, W., Franssen, H. H., Vereecken, H., and Kollet, S.: Catchment tomography - An approach for spatial parameter estimation, Advances in Water Resources, 107, 147 – 159, doi:https://doi.org/10.1016/j.advwatres.2017.06.006, http://www.sciencedirect.com/science/article/pii/S0309170816302019, 2017.

Bailey, R. and Baù, D.: Ensemble smoother assimilation of hydraulic head and return flow data to estimate hydraulic conductivity distribution, Water Resources Research, 46, n/a–n/a, doi:10.1029/2010WR009147, http://dx.doi.org/10.1029/2010WR009147, w12543, 2010.

Bauser, H. H., Jaumann, S., Berg, D., and Roth, K.: EnKF with closed-eye period - towards a consistent aggregation of information in soil hydrology, Hydrology and Earth System Sciences, 2016.

Bertino, L., Evensen, G., and Wackernagel, H.: Sequential Data Assimilation Techniques in Oceanography, International Statistical Review, 71, 223–241, doi:10.1111/j.1751-5823.2003.tb00194.x, http://dx.doi.org/10.1111/j.1751-5823.2003.tb00194.x, 2003.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, American Meteorological Society, 2001.

Brandhorst, N., Erdal, D., and Neuweiler, I.: Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters, Advances in Water Resources, 110, 360 – 370, doi:https://doi.org/10.1016/j.advwatres.2017.10.022, http://www.sciencedirect.com/science/article/pii/S0309170817301355, 2017.

Cammalleri, C. and Ciraolo, G.: State and parameter update in a coupled energy/hydrologic balance model using ensemble Kalman filtering, Journal of Hydrology, 416-417, 171 – 181, doi:https://doi.org/10.1016/j.jhydrol.2011.11.049, http://www.sciencedirect.com/science/article/pii/S0022169411008341, 2012.

Camporese, M., Paniconi, C., Putti, M., and Salandin, P.: Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow, Water Resources Research, 45, n/a–n/a, doi:10.1029/2008WR007031, http://dx.doi.org/10.1029/2008WR007031, w10421, 2009a.

Camporese, M., Paniconi, C., Putti, M., and Salandin, P.: Comparison of Data Assimilation Techniques for a Coupled Model of Surface and Subsurface Flow, 8, 837–845, doi:10.2136/vzj2009.0018, http://dx.doi.org/10.2136/vzj2009.0018, 2009b.

Camporese, M., Paniconi, C., Putti, M., and Orlandini, S.: Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data, Water Resources Research, 46, n/a–n/a, doi:10.1029/2008WR007536, http://dx.doi.org/10.1029/2008WR007536, w02512, 2010.

Carsel, R. F. and Parrish, R. S.: Developing joint probability distributions of soil water retention characteristics, Water Resources Research, 24, 755–769, doi:10.1029/WR024i005p00755, http://dx.doi.org/10.1029/WR024i005p00755, 1988.

Chen, Y. and Zhang, D.: Data assimilation for transient flow in geologic formations via ensemble Kalman filter, Advances in Water Resources, 29, 1107 – 1122, doi:https://doi.org/10.1016/j.advwatres.2005.09.007, http://www.sciencedirect.com/science/article/pii/S0309170805002277, 2006.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, Advances in Water Resources, 31, 1309 – 1324, doi:https://doi.org/10.1016/j.advwatres.2008.06.005, http://www.sciencedirect.com/science/article/pii/S0309170808001012, 2008.

Crow, W. T. and Wood, E. F.: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using Ensemble Kalman filtering: a case study based on ESTAR measurements during SGP97, Advances in Water Resources, 26, 137 – 149, doi:https://doi.org/10.1016/S0309-1708(02)00088-X, http://www.sciencedirect.com/science/article/pii/S030917080200088X, 2003.

De Lannoy, G. J., Houser, P. R., Pauwels, V., and Verhoest, N. E.: State and bias estimation for soil moisture profiles by an ensemble Kalman filter: Effect of assimilation depth and frequency, Water Resources Research, 2007.

Erdal, D., Neuweiler, I., and Wollschläger, U.: Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model, Water Resources Research, 50, 132–147, doi:10.1002/2012WR013443, http://dx.doi.org/10.1002/2012WR013443, 2014.

Erdal, D., Rahman, M., and Neuweiler, I.: The importance of state transformations when using the ensemble Kalman filter for unsaturated flow modeling: Dealing with strong nonlinearities, Advances in Water Resources, 86, 354 – 365, doi:https://doi.org/10.1016/j.advwatres.2015.09.008, http://www.sciencedirect.com/science/article/pii/S0309170815002092, data assimilation for improved predictions of integrated terrestrial systems, 2015.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynamics, 53, 2003.

Evensen, G.: Data Assimilation, Springer, 2009a.

Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, vol. 29, IEEE Control Systems, 2009b.

Flores, A. N., Bras, R. L., and Entekhabi, D.: Hydrologic data assimilation with a hillslope-scale-resolving model and L band radar observations: Synthetic experiments with the ensemble Kalman filter, Water Resources Research, 2012.

Francois, C., Quesney, A., and Ottlé, C.: Sequential Assimilation of ERS-1 SAR Data into a Coupled Land Surface–Hydrological Model Using an Extended Kalman Filter, Journal of Hydrometeorology, 4, 2003.

Hain, C. R., Crow, W. T., Anderson, M. C., and Mecikalski, J. R.: An ensemble Kalman filter dual assimilation of thermal infrared and microwave satellite observations of soil moisture into the Noah land surface model, Water Resources Research, 2012.

Han, X. and Li, X.: An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation, Remote Sensing of Environment, 2008.

Hendricks Franssen, H. J. and Kinzelbach, W.: Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, Water Resources Research, 44, n/a–n/a, doi:10.1029/2007WR006505, http://dx.doi.org/10.1029/2007WR006505, w09408, 2008.

Johnson, S. K.: Distributions in Statistics: Continuous Unvariate Distributions-1, Houghton Mifflin, 1970.

Kollet, S., Sulis, M., Maxwell, R. M., Paniconi, C., Putti, M., Bertoldi, G., Coon, E. T., Cordano, E., Endrizzi, S., Kikinzon, E., Mouche, E., Mügler, C., Park, Y.-J., Refsgaard, J. C., Stisen, S., and Sudicky, E.: The integrated hydrologic model intercomparison project, IH-MIP2: A second set of benchmark results to diagnose integrated hydrology and feedbacks, Water Resources Research, 53, 867–890, doi:10.1002/2016WR019191, http://dx.doi.org/10.1002/2016WR019191, 2017.

Kurtz, W., He, G., Kollet, S., Maxwell, R., Vereecken, H., and Franssen, H. H.: TerrSysMP-PDAF (version 1.0): A modular high-performance data assimilation framework for an integrated land surface-subsurface model, Geoscientific Model Development, 2015.

Li, L., Zhou, H., Gómez-Hernández, J. J., and Franssen, H.-J. H.: Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter, Journal of Hydrology, 2012.

Lora, M., Camporese, M., and Salandin, P.: Design and performance of a nozzle-type rainfall simulator for landslide triggering experiments, CATENA, 140, 77 – 89, doi:https://doi.org/10.1016/j.catena.2016.01.018, http://www.sciencedirect.com/science/article/pii/S0341816216300182, 2016a.

Lora, M., Camporese, M., Troch, P. A., and Salandin, P.: Rainfall-triggered shallow landslides: infiltration dynamics in a physical hillslope model, Hydrological Processes, 30, 3239–3251, doi:10.1002/hyp.10829, http://dx.doi.org/10.1002/hyp.10829, hYP-15-0649.R1, 2016b.

Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J.-O., Ferguson, I. M., Ivanov, V., Kim, J., Kolditz, O., Kollet, S. J., Kumar, M., Lopez, S., Niu, J., Paniconi, C., Park, Y.-J., Phanikumar, M. S., Shen, C., Sudicky, E. A., and Sulis, M.: Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks, Water Resources Research, 50, 1531–1549, doi:10.1002/2013WR013725, http://dx.doi.org/10.1002/2013WR013725, 2014.

Monsivais-Huertero, A., Graham, W. D., Judge, J., and Agrawal, D.: Effect of simultaneous state-parameter estimation and forcing uncertainties on root-zone soil moisture for dynamic vegetation using EnKF, Advances in Water Resources, 33, 468 – 484, doi:https://doi.org/10.1016/j.advwatres.2010.01.011, http://www.sciencedirect.com/science/article/pii/S0309170810000230, 2010.

Montzka, C., Pauwels, V. R. N., Franssen, H.-J. H., Han, X., and Vereecken, H.: Multivariate and Multiscale Data Assimilation in Terrestrial Systems: A Review, Sensors, 12, 16 291–16 333, doi:10.3390/s121216291, http://www.mdpi.com/1424-8220/12/12/16291, 2012.

Moradkhani, H.: Hydrologic remote sensing and land surface data assimilation, Sensors, 2008.

Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water resources research, 2005.

Pan, M. and Wood, E. F.: Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter, Journal of Hydrometeorology, 7, 534–547, doi:10.1175/JHM495.1, https://doi.org/10.1175/JHM495.1, 2006.

Pasetto, D., Camporese, M., and Putti, M.: Ensemble Kalman filter versus particle filter for a physically-based coupled surface–subsurface model, Advances in water resources, 2012.

Pasetto, D., Niu, G.-Y., Pangle, L., Paniconi, C., Putti, M., and Troch, P. A.: Impact of sensor failure on the observability of flow dynamics at the Biosphere 2 LEO hillslopes, Advances in Water Resources, 86, 327 – 339, doi:https://doi.org/10.1016/j.advwatres.2015.04.014, http://www.sciencedirect.com/science/article/pii/S0309170815001050, data assimilation for improved predictions of integrated terrestrial systems, 2015.

Rasmussen, J., Madsen, H., Jensen, K. H., and Refsgaard, J. C.: Data assimilation in integrated hydrological modelling in the presence of observation bias, Hydrology Earth System Sciences Discussions, 12, 2015a.

Rasmussen, J., Madsen, H., Jensen, K. H., and Refsgaard, J. C.: Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance, Hydrol. Earth Syst. Sci., 19, 2015b.

Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic data assimilation with the ensemble Kalman filter, Monthly Weather Review, 130, 2002a.

Reichle, R. H., Walker, J. P., Koster, R. D., and Houser, P. R.: Extended versus ensemble Kalman filtering for land data assimilation, Journal of hydrometeorology, 3, 2002b.

Ridler, M.-E., Madsen, H., Stisen, S., Bircher, S., and Fensholt, R.: Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark, Water Resources Research, 50, 8962–8981, doi:10.1002/2014WR015392, http://dx.doi.org/10.1002/2014WR015392, 2014.

Sakov, P., Evensen, G., and Bertino, L.: Asynchronous data assimilation with the EnKF, Tellus A, 62, 24–29, doi:10.1111/j.1600-0870.2009.00417.x, http://dx.doi.org/10.1111/j.1600-0870.2009.00417.x, 2010.

Schenato, L., Camporese, M., Bersan, S., Cola, S., Galtarossa, A., Pasuto, A., Simonini, P., Salandin, P., and Palmieri, L.: High density distributed strain sensing of landslide in large scale physical model, 2017.

Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., and Yu, X.: Parameter estimation of a physically-based land surface hydrologic model using an ensemble Kalman filter: A multivariate real-data experiment, Advances in Water Resources, 83, 421 – 427, doi:https://doi.org/10.1016/j.advwatres.2015.06.009, http://www.sciencedirect.com/science/article/pii/S0309170815001359, 2015.

Tang, Q., Kurtz, W., Schilling, O., Brunner, P., Vereecken, H., and Franssen, H.-J. H.: The influence of riverbed heterogeneity patterns on river-aquifer exchange fluxes under different connection regimes, Journal of Hydrology, 554, 383 – 396, doi:https://doi.org/10.1016/j.jhydrol.2017.09.031, http://www.sciencedirect.com/science/article/pii/S0022169417306352, 2017.

Troch, P. A., Paniconi, C., and McLaughlin, D.: Catchment-scale hydrological modeling and data assimilation, Advances in Water Resources, 26, 131 – 135, doi:https://doi.org/10.1016/S0309-1708(02)00087-8, http://www.sciencedirect.com/science/article/pii/S0309170802000878, 2003.

Visser, A., Stuurman, R., and Bierkens, M. F.: Real-time forecasting of water table depth and soil moisture profiles, Advances in Water Resources, 2006.

Vrugt, J. A., Gupta, H. V., Nualláin, B., and Bouten, W.: Real-Time Data Assimilation for Operational Ensemble Streamflow Forecasting, Journal of Hydrometeorology, 7, 548–565, doi:10.1175/JHM504.1, https://doi.org/10.1175/JHM504.1, 2006.

Warrick, A. W., Wierenga, P. J., Young, M. H., and Musil, S. A.: Diurnal fluctuations of tensiometric readings due to surface temperature changes, Water Resources Research, 34, 2863–2869, doi:10.1029/98WR02095, http://dx.doi.org/10.1029/98WR02095, 1998.

Weerts, A. H. and El Serafy, G. Y.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resources Research, 2006.

Wösten, J. H. M. and van Genuchten, M. T.: Using Texture and Other Soil Properties to Predict the Unsaturated Soil Hydraulic Functions, Soil Science Society of America Journal, 52 52, 1988.

Xie, X. and Zhang, D.: Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter, Advances in Water Resources, 33, 678 – 690, doi:https://doi.org/10.1016/j.advwatres.2010.03.012, http://www.sciencedirect.com/science/article/pii/S0309170810000618, 2010.

Zhang, D., Madsen, H., Ridler, M. E., Refsgaard, J. C., and Jensen, K. H.: Impact of uncertainty description on assimilating hydraulic head in the MIKE SHE distributed hydrological model, Advances in Water Resources, 86, 400–413, doi:https://doi.org/10.1016/j.advwatres.2015.07.018, http://www.sciencedirect.com/science/article/pii/S0309170815001657, 2015.

Zhang, D., Madsen, H., Ridler, Marc, E., Kidmose, J., and Jensen, Karsten, H.: Multivariate hydrological data assimilation of soil moisture and groundwater head, Hydrology and Earth System Sciences, 2016.

Zhang, H., Kurtz, W., Kollet, S., Vereecken, H., and Franssen, H.-J. H.: Comparison of different assimilation methodologies of groundwater levels to improve predictions of root zone soil moisture with an integrated terrestrial system model, Advances in Water Resources, 111, 224 – 238, doi:https://doi.org/10.1016/j.advwatres.2017.11.003, http://www.sciencedirect.com/science/article/pii/S0309170817304888, 2018.

Zovi, F., Camporese, M., Franssen, H.-J. H., Huisman, J. A., and Salandin, P.: Identification of high-permeability subsurface structures with multiple point geostatistics and normal score ensemble Kalman filter, Journal of Hydrology, 548, 208 – 224, doi:https://doi.org/10.1016/j.jhydrol.2017.02.056, http://www.sciencedirect.com/science/article/pii/S0022169417301361, 2017.

Table 1. Perturbation parameters for the generation of the ensemble initial conditions, hydraulic conductivities and atmospheric forcing.

Variable	Units	Nominal value(s)	Perturbation type	Mean	Standard deviation
ψ_0	m	-0.67	additive	0	0.2
Atmospheric forcing[a]	mm/h	58.8; 47.5; -0.0284	multiplicative	1	0.2
		Sand			
K_s	m/s	10^{-4}	multiplicative	1.17	1.53
S_s	m^{-1}	10^{-3}	not perturbed		
ϕ	-	0.58	not perturbed		
θ_r	-	0.065	see Section 4.2		0.1 or 0.25[b]
α	cm^{-1}	0.070	see Section 4.2		0.1 or 0.25[b]
n	-	1.88	see Section 4.2		0.1 or 0.25[b]
		Clay			
K_s	m/s	10^{-7}	multiplicative	1.17	1.53
S_s	m^{-1}	5×10^{-3}	not perturbed		
ϕ	-	0.40	not perturbed		
θ_r	-	0.067	see Section 4.2		0.1 or 0.25[b]
α	cm^{-1}	0.017	see Section 4.2		0.1 or 0.25[b]
n	-	1.40	see Section 4.2		0.1 or 0.25[b]

[a] Rainfall (positive) and evaporation (negative) rates.

[b] σ_{VG} as described in Section 4.2.

Table 2. Factored covariance matrices used for the perturbation of van Genuchten parameters (from Carsel and Parrish, 1988)).

		θ_r	α	n
Sandy loam	θ_r	0.538	0.017	-0.194
	α		0.014	0.019
	n			0.108
Silt loam	θ_r	0.522	0.030	-0.17
	α		0.082	0.234
	n			0.158

Table 3. Overview of the open loop and data assimilation scenarios.

Scenario[1]	Assimilation interval	Assimilated variables[2]	Updated variables[3]	σ_{VG}
OL1	-	-	-	0.1
OL2	-	-	-	0.25
S1	3 h	WC	St. var., K_s	0.1
S2	3 h	PH	St. var., K_s	0.1
S3	3 h	WC, Q	St. var., K_s	0.1
S4	3 h	PH, Q	St. var., K_s	0.1
S5	3 h	WC, PH, Q	St. var., K_s	0.1
S6	1 h	WC	St. var.	0.1
S7	1 h	WC, Q	St. var.	0.1
S8	1 h	WC	St. var., K_s	0.1
S9	1 h	PH	St. var., K_s	0.1
S10	1 h	WC, Q	St. var., K_s	0.1
S11	1 h	PH, Q	St. var., K_s	0.1
S12	1 h	WC, PH, Q	St. var., K_s	0.1
S13	1 h	WC	St. var., K_s, V.G.	0.25
S14	1 h	PH	St. var., K_s, V.G.	0.25
S15	1 h	WC, Q	St. var., K_s, V.G.	0.25
S16	1 h	PH, Q	St. var., K_s, V.G.	0.25
S17	1 h	WC, PH, Q	St. var., K_s, V.G.	0.25

[1] OL1 and OL2 indicate open loop scenarios, i.e., simulations without data assimilation.

[2] WC, PH, and Q denote water content, pressure head, and subsurface outflow, respectively.

[3] St. var., K_s, and V.G. indicate state variables (in terms of pressure head), saturated hydraulic conductivity, and van Genuchten parameters, respectively.

Table 4. Normalized root mean square errors ($NRMSE$) for the 17 data assimilation scenarios under analysis and two open loop (OL) simulations. The table reports the $NRMSE$ for three variables, water content, pressure head and outflow discharge, both for the assimilation and the validation periods. The last three columns report the mean values calculated over the three variables (WC, PH and Q) and the global average between assimilation and validation. Bold values indicate exceedance of the corresponding open loop errors.

	$NRMSE_{WC}$		$NRMSE_{PH}$		$NRMSE_{Q}$		Mean $NRMSE$		
Scenario	Assimilation	Validation	Assimilation	Validation	Assimilation	Validation	Assimilation	Validation	Global
OL1	0.16	0.14	0.45	0.36	0.72	0.64	0.44	0.38	0.41
OL2	0.16	0.13	0.46	0.36	0.71	0.63	0.44	0.38	0.41
S1	0.09	0.11	0.33	0.27	0.58	0.53	0.33	0.30	0.32
S2	**0.21**	**0.24**	0.21	0.31	0.51	0.47	0.31	0.34	0.33
S3	0.10	**0.16**	0.31	0.21	0.63	0.56	0.34	0.31	0.33
S4	**0.24**	**0.26**	0.25	**0.39**	**0.82**	**0.73**	0.44	**0.46**	**0.45**
S5	**0.18**	**0.22**	0.23	0.26	0.70	0.64	0.37	0.37	0.37
S6	0.07	0.12	0.36	0.36	0.68	0.62	0.37	0.37	0.37
S7	0.08	0.12	0.33	0.33	0.55	0.50	0.32	0.31	0.32
S8	0.08	0.11	0.32	0.27	0.55	0.49	0.32	0.29	0.30
S9	**0.22**	**0.24**	0.17	0.29	0.46	0.42	0.29	0.32	0.30
S10	0.09	**0.18**	0.31	0.21	0.54	0.48	0.31	0.29	0.30
S11	**0.24**	**0.33**	0.22	**0.83**	0.59	0.54	0.35	**0.57**	**0.46**
S12	**0.19**	**0.31**	0.20	**0.68**	0.62	0.55	0.34	**0.51**	**0.43**
S13	0.07	0.06	0.32	0.28	0.52	0.49	0.30	0.28	0.29
S14	**0.33**	**0.41**	0.16	0.21	0.50	0.45	0.33	0.36	0.34
S15	0.07	0.05	0.31	0.29	0.53	0.51	0.31	0.28	0.29
S16	**0.42**	**0.48**	0.18	0.19	0.60	0.57	0.40	**0.41**	0.41
S17	0.10	0.05	0.22	0.29	0.69	**0.66**	0.34	0.34	0.34

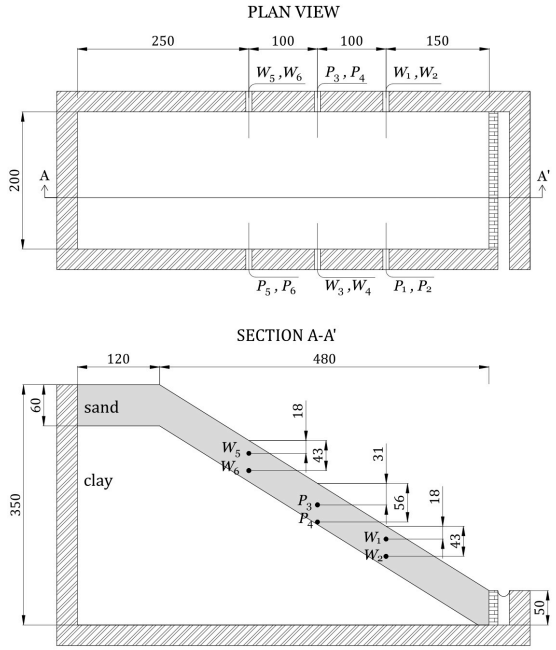

Figure 1. Plan view and longitudinal cross-section of the artificial hillslope, along with the position of the monitoring instruments. Tensiometers are indicated by the letter "P", while WCR probes are denoted by "W". All dimensions are in cm.

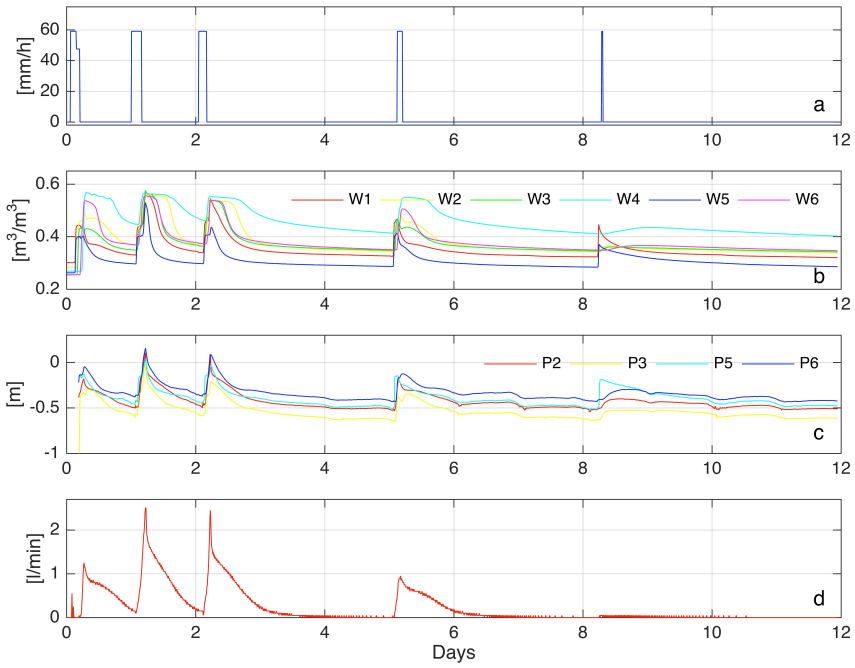

Figure 2. Experimental data collected during the experiment: a) rainfall rate; b) water content as measured by the WCR probes; c) pressure head as measured by the tensiometers; d) subsurface outflow. The black vertical dashed line marks the transition between the data assimilation and the validation phases.

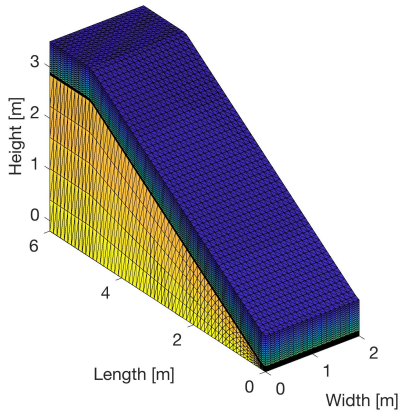

Figure 3. Three-dimensional finite element grid of the hillslope.

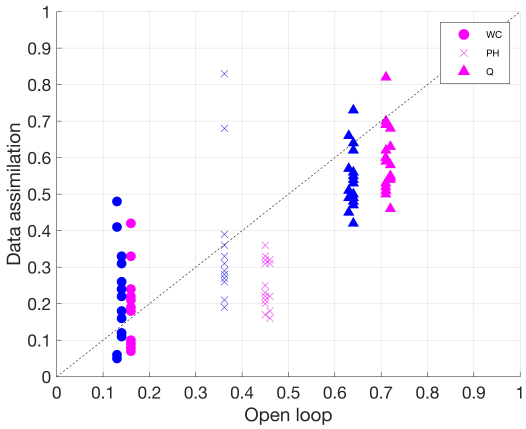

Figure 4. Normalized root mean square errors for water content, pressure head, and subsurface outflow of the data assimilation scenarios versus corresponding values of the open loop. Symbols in magenta represent values calculated over the assimilation period, while symbols in blue represent values calculated over the validation period.

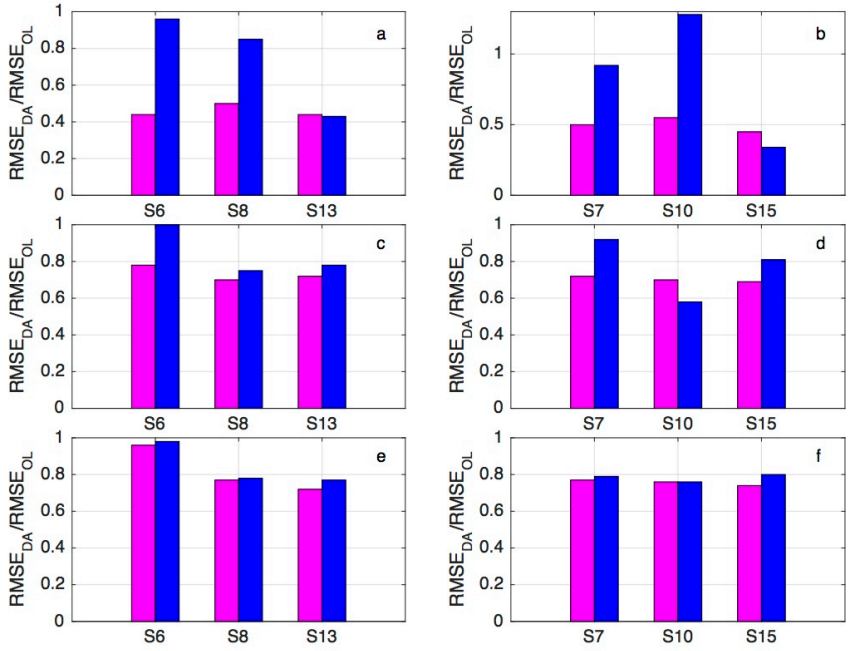

Figure 5. Ratios between $RMSE$s in scenarios S6, S8, S13 (panels a, c, e), and S7, S10, S15 (panels b, d, f) and the corresponding open loop values, for water content, WC (a and b), pressure head, PH (c and d), and subsurface outflow, Q (e and f). The magenta bars refer to the data assimilation period, while the blue ones to the validation period.

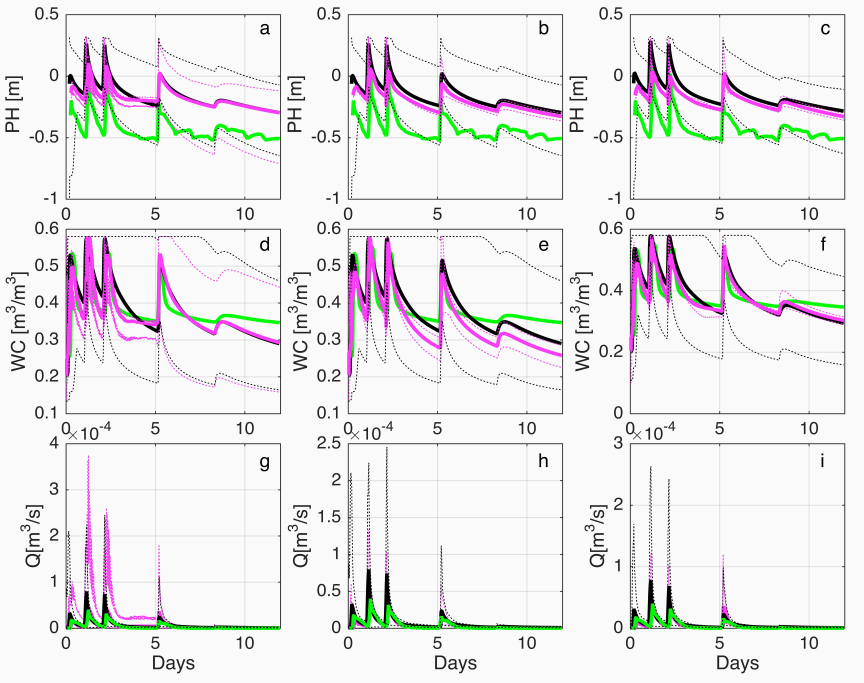

Figure 6. Plots of the pressure head in P2, (a, b and c), water content in W6, (d, e, f) and outflow discharge (g, h, i) for scenarios S6 (a, d, g), S8 (b, e, h), and S13 (c, f, i). Solid green lines represent experimental data while solid black and magenta lines indicate the ensemble mean of the open loop and data assimilation scenarios, respectively. The simulated 90% confidence bands are also reported (dashed black and magenta lines). The locations of the tensiometer P2 and water content probe W6 are shown in Figure 1.

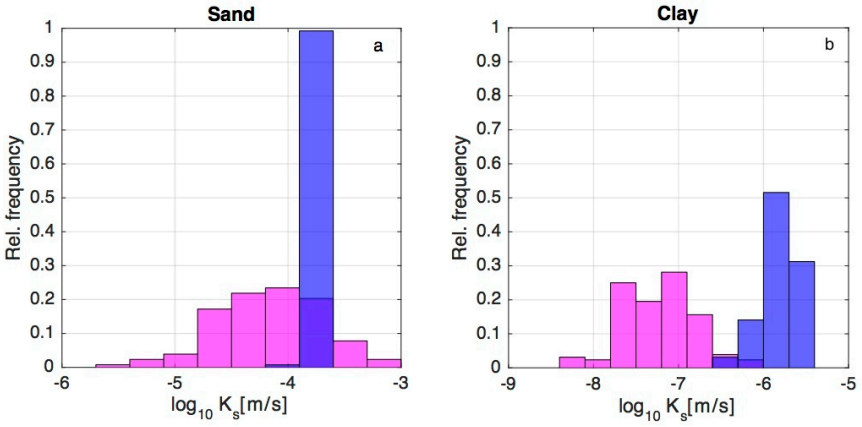

Figure 7. Relative frequency distributions of the saturated hydraulic conductivity, K_s, of sand and clay in scenario S10. Graphs in magenta denote the prior K_s distributions, while the blue ones indicate the posterior distributions, i.e., at the end of the assimilation period.

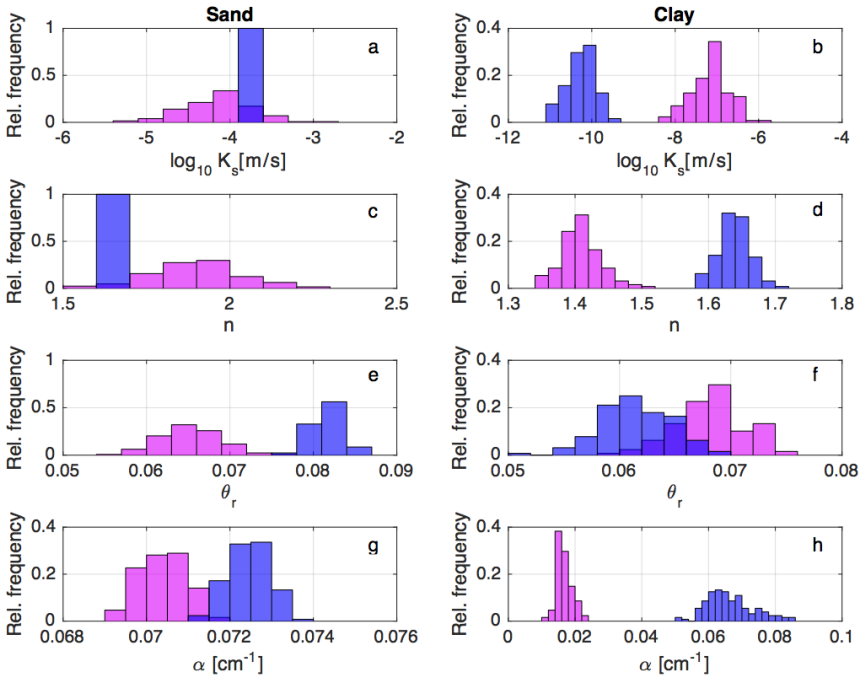

Figure 8. Relative frequency distributions of the saturated hydraulic conductivity, K_s, and van Genucthen parameters of sand and clay in scenario S15. Graphs in magenta denote prior distributions, while the blue ones indicate posterior distributions, i.e., at the end of the assimilation period.

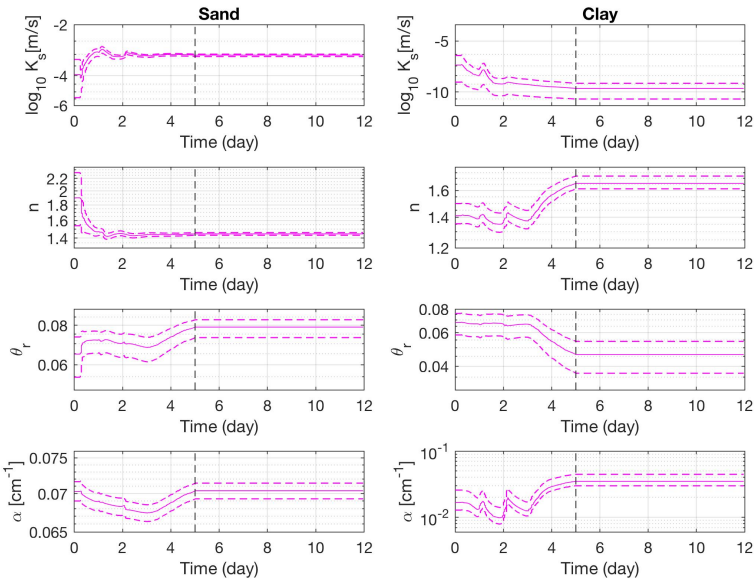

Figure 9. Time evolution of the saturated hydraulic conductivity and van Genuchten parameters (mean values, solid line, together with minimum and maximum values, in dashed lines, to indicate the ensemble spread) for the two types of soil, sand and clay, in scenario S17.

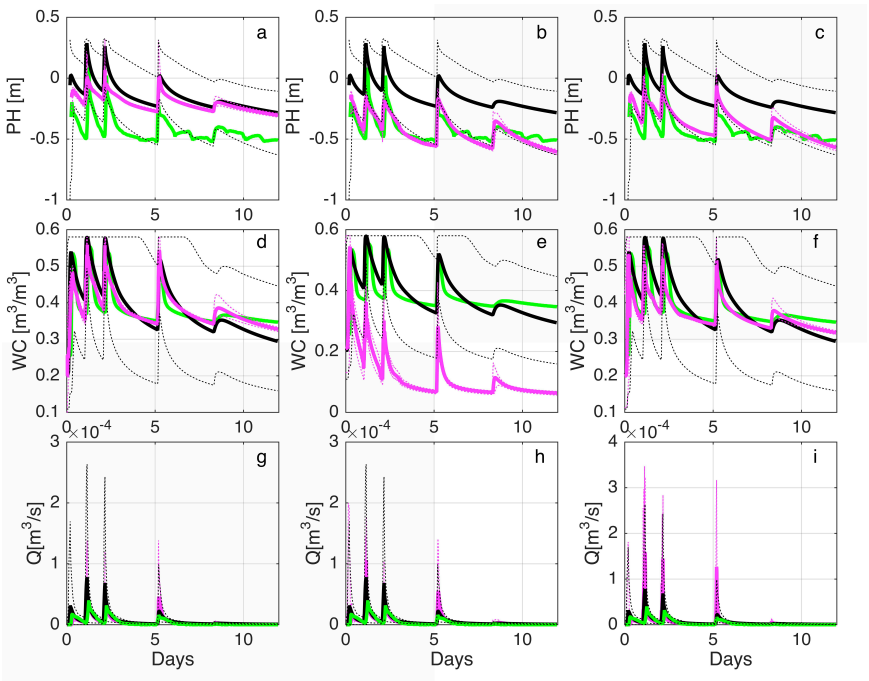

Figure 10. Plots of the pressure head in P2, (a, b and c), water content in W6, (d, e, f) and outflow discharge (g, h, i) for scenarios S15 (a, d, g), S16 (b, e, h), and S17 (c, f, i). Solid green lines represent experimental data while solid black and magenta lines indicate the ensemble mean of the open loop and data assimilation scenarios, respectively. The simulated 90% confidence bands are also reported (dashed black and magenta lines). The locations of the tensiometer P2 and water content probe W6 are shown in Figure 1.