RC2: 'Comment on hess-2024-219', Anonymous Referee #2, 04 Oct 2024 reply

We thank very much the reviewer 2 for his positive comments on our study and took into account most of his comments, improving the paper.

Authors presented a comparison of different DA methods in the context of modular hydrological model for water quality management. The paper is well-written and looks like very comprehensive. I have a couple of comments:

1) In the literature, a few papers about the comparison of DA methods have been published in the field of hydrogeology. Also, those methods are well established, and the disadvantages and advantages are well-known.

This is true about the classical EnKF, less so when looking at iEnKS et ES-MDA which have seldom been used (if ever) for such application and behave very differently from EnKF. Moreover PESHMELBA has some significant peculiarities (see below) that make this study necessary.

Authors highlighted the modular hydrological model in this study instead of many studies using numerical models. if so, authors should clarify why there are differences using different physical models for DA, not only from the results of DA experiments, but from the methodology. Fundamentally, DA methods such as EnKF can be coupled with any transfer functions.

The PESHMELBA is said to be "modular" in the way that it is a coupling of several (independant) modules, each ones representing the processes occurring in a specific element of the landscape and playing a role in pesticide transfers : a vegetative filter strip, a river, a plot of maize, a hedge, etc etc. The meshing is this not a classical one but based on the landscape management leading to hydrological units playing a specific role in the agricultural catchment. Finally, the processes may be physically-based modeled when it is possible (for instance, Richards equation for infiltration in plots), or more conceptually/empirically when there is no equation known to represent it. The aim of this model is not to be a fully physically-based model such as Parflow or Hydrus-3D, but to simulate and compare scenarios of landscape management (e.g., including more or less buffer zones in a catchment), to identify an optimal configuration regarding pesticide transfer mitigation and demonstrate to the stakeholders. For these reasons, the model is a coupling of physical and empirical/conceptual models (also called "semi-conceptual models" in Buytaert et al. 2008) and can depend on thresholds making it highly non-linear. This type of model has not been widely investigated for data assimilation. We can imagine that it will be difficult to find a regular solution for these semi-conceptual models and this is why we think this study is important in data assimilation for hydrological and water quality modeling for decision-making.

This is explained in lines 66-77, and maybe also clarified by the new title : "Comparison of ensemble assimilation methods in a decision support model for landscape management to mitigate pesticide transfer"

Buytaert, W., Reusser, D., Krause, S., and J.-P., R.: Why can't we do better than Topmodel?, Hydrological Processes, 22, 4175–4179, https://doi.org/10.1002/hyp.7125, 2008.

2) in line 273, the true value comes from perturbation from Gaussian noises. does this mean that your ground truth has a Gaussian distribution. How is this close to the real data? Does the real data follow Gaussian distribution? If it has a non-Gaussian distribution, how does those DA methods perform?

This a very good question. Indeed, these methods all assume that the probability densities being manipulated are Gaussian in order to be optimal, which was verified in a previous work (Rouzies et al., 2023, and Rouzies PhD, 2024 (in french)). We added this in the text.

However, such an assumption is not justified in all cases of Rouzies PhD. In particular, it is noted that in some winter scenarios we tested, surface humidity sometimes follows a bimodal distribution. In these cases, a particle filter approach (van Leeuwen and Evensen, 1996) may be an interesting alternative, as it does not rely on any assumption of Gaussianity. This method has rarely been applied in hydrology (Moradkhani et al., 2005 ; Pasetto et al., 2012), although the particle filter remains an attractive method that may be worth exploring in cases where the ensemble Kalman filter fails.

A simpler solution to continue using the EnKF, consists in transforming variables into gaussian ones, using anamorphosis methods (Bertino et al., 2003). See, for instance, applications of anamorphosis in ocean and biogeochemical modeling (Beal et al., 2010) or in meteorological reanalysis (Devers et al., 2020).

Béal, D., Brasseur, P., Brankart, J.-M., Ourmières, Y., and Verron, J.: Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis, Ocean Sci., 6, 247–262, https://doi.org/10.5194/os-6-247-2010, 2010.

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

Rouzies, E., Lauvernet, C., Sudret, B., and Vidard, A.: How to perform global sensitivity analysis of a catchment-scale, distributed pesticide transfer model? Application to the PESHMELBA model, Geoscientific Model Development, 2023, 1–44, https://doi.org/10.5194/gmd-16-3137-2023, 2023.

Rouzies, E. Quantification et réduction de l'incertitude dans un modèle de transfert de pesticides à l'échelle du bassin versant. Mathématiques [math]. Université Grenoble Alpes [2020-..], 2023. Français. ⟨NNT : 2023GRALM025⟩. ⟨tel-04659164v2⟩

3) As we know, those DA methods are impacted by the ensemble size. Have you considered to implement some localizations to constrain the covariance so that the filter inbreeding issue could be reduced? In figure 11, it looks like that, if ensemble size is increased from 50 to 200, the performance of DA gets worse. This does not make sense.

Indeed, localization schemes could be implemented by using the covariance localization or local domain DA . About the covariance, this would be feasible with another implementation of the Kalman Filter, but not with ETKF since the covariance matrix is not built explicitly in this method. About the local domain DA, this would be very interesting and quite relevant, especially considering the structure of soils that are described (see Figure 12 that shows the spatial correlation by soil type).

In the discussion, l. 456, we added this sentence :

l 456. Figure 12 also highlights the absence of spatial correlations between soil units of different soils, advocating for using a scheme with local domain DA in ES-MDA (Asch et al., 2016) to alleviate the computational cost of this method that uses high dimension matrices.

Note that Figure 11 represents a test run. If reproduced multiple times, we would likely observe a trend of decreasing error. However, due to the high cost, we opted not to conduct further testing. The mean trend conforms to what we expected. We have performed this test for sizes 20 and 50, but extending it further is prohibitively expensive.