Reviewer 3

Summary:

This manuscript integrates the Parallel Data Assimilation Framework (PDAF), an open-source data assimilation software, with the HydroGeoSphere (HGS) hydrological model. The integration involves separate and alternating executions of HGS and PDAF, enabling information from one model to inform the other. Similar applications involve combining EnKF with HGS and combining PDAF with ParFlow. PDAF encompasses two fundamental classes of DA methods: Ensemble Kalman Filter (EnKF)-based, offering distributions for estimators, and variational-based, providing point estimators. This study specifically demonstrates the model binding using EnKF-based PDAF, validated through an application in a quasi-hypothetical numerical river-aquifer model. The model's performance on state variables, hydraulic head and soil moisture, is assessed using their ensemble mean. The model parameter, hydraulic conductivity (K), is constrained by the expected prior distribution, aligning with the method's anticipated behavior.

Comments:

Line 21-23: The assertion of operational real-time management may be perceived as over-promising. It heavily depends on the infrastructure of data warehousing and model pipelines.

We couldn't agree more with the reviewer that operational real-time management requires much more than just a model and a DA platform. It also requires sensors, secure and robust data transmission and storage, other infrastructure and pipelines. However, what we assert with our statement is that with the integrated model and this modular DA framework, we have essentially developed the hydrologically and DA wise robust toolbox for developing the basic model for operational management of coupled surface water-groundwater resources. We have adjusted the statement accordingly:

Lines 21-23: "With the integrated model and this modular DA framework, we have essentially developed the hydrologically and DA wise robust toolbox for developing the basic model for operational management of coupled surface water-groundwater resources."

EnKF related:

- Line 169: Clarify the term "state vector with model parameters." Is Xp representing model parameters sampled at a given realization from its latent distribution?

Yes, the term Xp represents the model parameters from a given distribution. To clarify the state vector, we reformulated the equation to describe the state vector:

*Lines 166-175: "In mathematical terms, consider that a state vector **X** can be written as Eq. (1):*

$$\boldsymbol{X}_i = (\boldsymbol{X}_s) \quad (1)$$

where $\boldsymbol{X}_s$ is the state vector with model state variables. When parameters are updated together with the state variables, the augmented state vector can be written as

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_s \\ \boldsymbol{X}_p \end{pmatrix}_i \quad (2)$$

where $\mathbf{X}_p$ is the state vector with model parameters."

- Equation 2: Specify whether the forward transient process is noise-free. While understanding that the noise term may be controlled by parameters in Xp, consider presenting EnKF in the standard state space model format, clearly defining states, parameters, and distributions.

Yes, it's noise free. We have reformulated the equation to describe the state vector. Please see our previous point.

- Equation 3: Define the observation model here to maintain a consistent format with Equation 2, rather than introducing it directly from Equation 4.

We agree that this is not the original version of the observation model which maps the observations to the model state but only to described how the observations are perturbed by the observation errors in EnKF. This has been explained in Burgers et al. (1998). In order to maintain the consistency of such a modified version of EnKF, we leave this formula here.

- Line 200-206: If parameters and states are well-defined, refer to them in this section. Consider adding this information to the suggested flowchart to visually represent the requirements.

We added this information in the manuscript as suggested by the reviewer:

Lines 212-215: "whether the model parameters are included in the state vector for updating along with the state variables. If yes, and if the parameters to be included is the hydraulic conductivity (K),"

As this is also related to the Flowchart of the initialisation of data assimilation, we also updated the corresponding description text:

Line 317-319: "Notice that we may need transferring the original values of the model state or parameters, e.g. for K, the log-transformed K is considered in the state vector rather than the K itself used in the HGS model to ensure that K is always positive during the assimilation process;"

The flowchart itself is not changed as this is part of the definition of the state vector.

- Line 279: I am curious whether the covariance matrix encounters degeneracy problems after many time steps.

In Figure 9 the two realisations of K are different which indicates the covariance matrix is not too small, which in turn means that until the end of the simulation period, there is no covariance matrix degeneracy problem.

Flowchart related:

• Lines 134-152: Clarity in this paragraph could be enhanced with the inclusion of a flowchart, similar to Figure 1.

We have added a flowchart (new Figure 1) to clarify the workflow of HydroGeoSphere.

• Consider improving the flowchart quality in Figures 1 to 4 by incorporating consistent boxes and colors to distinguish observation, model run, configuration, and output steps. Providing a flowchart illustrating the connections between different modules can offer a more comprehensive overview.

In Figure 1-5, the green blocks are the HGS model related parts, the yellow blocks are the model bindings, the blue block is the PDAF software, and the orange blocks are the observation related parts. Figure 3 shows the connection between the different modules/subroutines. The parameter modules are not shown in this figure as they are not the process module and are predefined and used by the initialisation subroutines. This is already described in the manuscript in section 3.3.2.

Figure 8: Strengthen your claim by plotting the standard error for ensemble mean, demonstrating statistical significance in error reduction to bolster your argument.

Currently we only store the ensemble mean for the variables in the state vector and the standard deviation if not saved. Thus, plotting the standard error is currently not possible.