Reviewer 2

In this manuscript, the authors present a coupling framework to integrate a data assimilation toolbox with the HydroGeoSphere (HGS) fully-coupled groundwater – surface water model. This is timely work as there is increasing interest in operationalizing structurally and physically complex models like HGS, and robust data assimilation methodology is required. The manuscript is suited for GMD, well written, and in general, well organized. I only have a few minor comments for the authors to consider.

L101: this bullet point needs clarified.

This is now clarified:

Lines 99-101: "2) a modular tool to handle different types of observation data, which enables to assimilate one or multiple types of observations simultaneously, currently programmed for hydraulic heads, soil moisture and solute concentration measurements."

L123: Replace 'Saint-Venant' with 'diffusion wave'. Could also mention one-dimensional open channel flow.

We replaced this as suggested by the reviewer. In HGS, the surface water flow is represented as two-dimensional depth-averaged areal flow.

L140: Comma not needed behind 'files'.

Corrected as suggested by the reviewer.

L165: multiple realizations of a numerical model.

Corrected as suggested by the reviewer.

L304: values for the nodes (I believe these are nodal properties referred to in this sentence).

Corrected as suggested by the reviewer.

L364: What is the clock speed for these CPUs? Were the individual HGS simulations also parallelized, if so, across how many cores?

The clock speed per computing node is 2.25 GHz. No, the individual HGS simulation is not parallelized, i.e. each HGS model is run on 1 core. This is now clarified in the manuscript:

Lines 374-375: "The clock speed per computing node is 2.25 GHz. The individual HGS simulation is not parallelized, i.e. each HGS model is run on 1 core."

L391: river bank filtration pumping wells,

Corrected as suggested by the reviewer.

L412: (tint)

Removed as suggested by the reviewer.

L414: could remove (i.e. with maximum pumping regime)

Removed as suggested by the reviewer.

L417: brackets around (i.e. K).

Corrected as suggested by the reviewer.

L423: producing a heterogeneous parameter field.

Corrected as suggested by the reviewer.

L430: would saturation at these points not be dependent on head, hence head and saturation at coincident points is redundant?

The saturation and hydraulic head depend on each other in the unsaturated zone. When hydraulic head and soil water saturation are updated, they are both combined in the state vector and updated simultaneously using the covariance matrix. In the example shown in the paper, when these two variables are updated together, the initial condition for the next prediction cycle was only based on hydraulic head. This is now explained in the manuscript:

Lines 460-461: "When hydraulic heads and soil water saturation are updated together, the initial condition for the next prediction cycle is only hydraulic head."

We would like to state that the functional relationship between saturation and hydraulic head suggested by the reviewer is only applicable if unsaturated conditions are present. If the groundwater level rises, the head can still change yet the degree of saturation will be at 100%. As we are jointly simulating saturated/unsaturated conditions it is important to consider both saturation and head. Note also that the functional relationships are often associated with large uncertainties and processes such a hysteresis, which is not considered in our models. The consideration of these two variables is therefore not necessarily redundant. Given that our case is a purely illustrative example to demonstrate the modularity of HGS-PDAF, it is therefore out of scope of the paper to analyse the effects of different DA strategies when assimilating both hydraulic heads and saturation simultaneously.

L433: This perturbation is quite small in relation to variability in a natural system of similar scale, and in particular 1 % SD in moisture content is almost negligible. Could the authors comment on what would be considered reasonable values for a real-world scenario, and how run times might be affected?

We agree that 1% SD is low for a real-test case study where the spatial representation of measured saturations may be influenced by local scale heterogeneities and preferential flow paths. However, the example shown in the paper is based on a synthetic model setup and the observations are also generated synthetically, we determine the observation error values based on the prior knowledge and the tuning experiments, e.g. 5 % and 10 % have also been tested as the saturation error. We use a relatively small observation error to better illustrate how HGS-PDAF works. We have added a few sentences to describe this in the manuscript:

Lines 445-447: “The values of the observation errors are determined by our prior knowledge and tuning experiments. Different percentages such as 5% and 10% were tested and subsequently defined to provide a most illustrative use case.”

In a real-world scenario, such a measurement error may be higher to also account for measurement representation. Specific values for measurement errors are case specific but must always respect proper balance between goodness of fit and over-fitting to preserve the consistency of the updated states.

In terms of runtimes, HGS uses adaptive time steps for numerical iteration. Once DA is implemented, as described in the manuscript, the simulation will be interrupted per assimilation frequency and the model will always need to be restarted and initialised with an initial-small time step. This will certainly increase the overall run times.

General comment:

- Could the authors comment in the manuscript on how perturbations in head and moisture content affected the numerical stability and time-step intervals for subsequent simulations? Is there a sweet spot for the amount of perturbation so that both data assimilation and model run times can be optimized? It is my understanding that if updates to the model state induce shocks or instabilities into the initial condition then simulation run times can appreciably slow down.

For this synthetic model we have tested different saturation error values such as 1%, 5% and 10% to monitor the model stability against the assimilation performance and to define an optimal error to to achieve a most illustrative use case. The total simulation run times are similar for the three cases with different observation errors, while the best results are obtained with the smallest observation error, i.e. 1%. As this example is based on a synthetic model setup, and the observations are also generated synthetically, such a small observation error doesn't induce shocks or instabilities in the initial condition and therefore doesn't significantly increase the simulation runtimes, so the sweet spot depends only on the assimilation performance. However, since the focus of this paper is to show the structure of the developed HGS-PDAF framework, and this synthetic experiment is purely an illustrative exercise to show how DA can be achieved via HGS-PDAF, nothing can be generalised from such a synthetic model. We agree that in a real case, the deviation between the model simulation and the real observation can be large, and updating the model state with a small observation error can affect the numerical stability, thus increasing the time step intervals and the total simulation run times. However, as this is purely illustrative, to carry out an analysis of perturbations and time steps on data assimilation performance is beyond the scope of this paper.

- EnKF has been used now for a number of HGS DA applications. However, as the authors note, the PDAF toolbox supports many other DA approaches. Could the authors add a table to the manuscript that lists the other DA approaches, previous application of these approaches towards hydrologic modeling, and general guidelines for users of the HGS-PDAF framework to select the most suitable approach for their application? Or perhaps list the strengths and weaknesses of the different approaches WRT fully coupled groundwater – surface water modeling?

We have added a table (Appendix 2) as suggested by the reviewer. It shows the DA approaches supported by PDAF, field of application and examples of reference. DA approaches are application dependent, and the classical EnKF should be fine when the number of observations is rather low, as in our illustrative example, and is therefore widely used in hydrological *simulation. If the observation number is high, we can also consider different types of ensemble transform Kalman filters, such as ETKF (Bishop et al., 2001) and ESTKF (Nerger et al., 2012). In particular, if the number of observations is large, localisation (Nerger et al., 2006) should also be considered.*

Lines 201-202: "The available DA approaches and their application fields as well as several example references are listed in Appendix 1."

"Appendix 1: Data assimilation approaches in PDAF and their known application fields

Data assimilation approaches			*Fields of application*	*Examples in hydrogeology (if not applicable, we give references in other fields and marked with *)*
Ensemble based	*Global*	*EnKF*	*Meteorology, oceanography, hydrology, hydrogeology, land surface*	*Tang et al. (2017); Tang et al. (2018)*
		ETKF	*Meteorology, oceanography, hydrology, hydrogeology, land surface*	*Rasmussen et al. (2016); Zhang et al. (2016)*
		SEIK	*Meteorology, oceanography, hydrology, hydrogeology*	*Schumacher (2016)*
		ESTKF	*Meteorology, oceanography, hydrology, hydrogeology*	*Li et al. (2023b)*
		NETF	*Meteorology, oceanography*	*Nerger (2022); Tödter et al. (2016)**
		PF	*Meteorology, oceanography, hydrology, hydrogeology, land surface*	*Abbaszadeh et al. (2018); Berg et al. (2019)*
		SEEK	*Meteorology, oceanography*	*Brasseur and Verron (2006); Butenschön and Zavatarelli (2012)**
	Local	*LEnKF*	*Meteorology, oceanography, hydrology, hydrogeology, land surface*	*Hung et al. (2022); Li et al. (2023a)*
		LETKF	*Meteorology, oceanography, hydrology, hydrogeology, land surface*	*Sawada (2020)*
		LSEIK	*Meteorology, oceanography*	*Liang et al. (2017); Liu and Fu (2018)**
		LESTKF	*Meteorology, oceanography*	*Zheng et al. (2020)**
		LNETF	*Meteorology, oceanography*	*Feng et al. (2020)**
		LKNETF	*Meteorology, oceanography*	*Shao and Nerger (2024)**

Variational		*3DVAR*	*Meteorology, oceanography, hydrology*	*Cummings and Smedstad (2013); Li et al. (2008)**

”