Reviewer 2

This paper focuses on uncertainty and sensitivity analysis techniques applied to the holistic SWAT+ model, which is coupled with a new gwflow module for physically based groundwater flow modeling. The study evaluates these techniques in four different watersheds across the United States, each with distinct hydrologic characteristics. The main parameters of the coupled SWAT+*gwflow* model are estimated using parameter estimation software (PEST), and model performance is assessed based on various hydrological metrics. The results are intriguing.

Response: We appreciate your feedback! Your comments have been greatly valuable to us, and we've thoroughly reviewed them, making every effort to respond appropriately.

1. Please make sure the abbreviation is defined at the very beginning, GLM, NHD, SWAT, etc.

Response: Thank you for your comment. We could include a list of abbreviations at the beginning of the article (after the keywords but before the Introduction). However, we are not sure if this is acceptable to the journal. Therefore, we have elected to define all abbreviations in the first instance they occur in the paper (highlighted in red in the revised version of the article).

2. For table 1, what is the reason for using 1000 m instead of 500 m resolution for Arkansas Headwaters?

Response: Thank you for the comment. The grid size of 1000 m is selected for Arkansas Headwaters due to the large size of the watershed (7,940 km^2), which leads to a significant increase in computation time if a cell size of 500 m is used. In preliminary simulations, a cell size of 500 m vs. a cell size of 1000 m did not have a noticeable difference on streamflow.

3. Figure 1, please add unit of DEM.

Response: Thank you for pointing this out. The DEM unit is added in Figure 1.

4. Figure 7, it looks like the predicted values are smaller than observed value in most cases, what is the reason? It needs more discussion to evaluate the model performance.

Response: Thank you for the comment. Yes, generally the PBIAS is positive for all models, indicating an underestimation in streamflow by the model as compared to measured flow. This is not always the case, as there are months (Nanticoke River, Arkansas River, Cache River) in which the model estimates higher streamflow than measured. Underestimation typically occurs for peak flow conditions, which, since the models are also tested for groundwater head, likely is due to inaccurate spatial and temporal representation of significant precipitation events. We believe this is the case with all hydrologic models.

5. Figure 9, why there are only 2 points in the mean absolute error plot?

Response: We are not sure what the reviewer is referring to in this comment. Groundwater head results have been evaluated at 128 monitoring wells and are distributed as follows: 7 in the Winnebago River, 26 in the Nanticoke River, 92 in the Cache River, and 3 in the Arkansas Headwaters. The box plot and histogram in Figure 9 display the frequency of the mean absolute error (in meters) of groundwater head for the 128 monitoring wells across the four study watersheds. Please, check line 223.

6. Could you add more detail how you calibrate the model.

Response: We appreciate your feedback. The calibration process is explained in Section 2.2.1, which covers Method #1: Parameter ESTimation Tool (PEST) followed by Sensitivity Analysis, and in Section 2.2.2, which discusses Method #2: Iterative Ensemble Smoother (iES) for Parameter Estimation and Uncertainty Analysis (UA).

Technically, the calibration procedure using the Parameter ESTimation (PEST) software for a hydrologic model involves the following steps:

- Model Setup: Configure the hydrologic model to work with PEST. This may include setting up model input and output files in the required formats.
- Parameterization: Recognize the parameters within the hydrologic model that require to be calibrated. These are the parameters that you want PEST to modify to enhance the model's performance.

- Observation Data: Prepare observed data that will be utilized to assess the model's performance during calibration. This data should be in a format that PEST can work with, usually a text file. In our study, we used two watershed responses: streamflow and groundwater head.
- PEST Control File: Create a PEST control file (usually with a .pst extension) that identifies the optimization settings, calibration parameters, observation data, and other relevant information.
- Run PEST: Execute PEST with the control file as input. PEST will interact with your hydrologic model to run simulations with different parameter values.
- Model Simulations: PEST will perform multiple model simulations with several parameter combinations as it gets to minimize the objective function (a measure of the model's fit to observed data of either streamflow or groundwater heads).

The number of iterations in a PEST calibration can vary broadly depending on various aspects, including the complexity of the hydrologic model, the quality and quantity of observed data, the number of parameters being calibrated, and the convergence criteria set by the user. PEST employs an iterative optimization algorithm to adjust model parameters to minimize an objective function (typically a measure of the difference between observed data and model predictions).

Until how many iterations do you consider good calibration?

We have included a text in section 2.2.1. of revised article as follow:

Lines 234-236:

"In our study, we set the maximum number of optimization iterations to 50. However, often PEST converged after 22 iterations (1600 model calls) for Winnebago River, 13 iterations (674 model calls), 36 iterations (2705 model calls) for Arkansas Headwaters, and 13 iterations (843 model calls) for Cache River."

1. More studies with similar hydrology parameters should be tested to see if the method could have consistent performance.

Response: Yes, we agree with the reviewer. We are planning to do many more studies, with other watersheds, in the near future. For this study, we have now included results (calibrated parameters, streamflow statistics, hydrographs, uncertainty plots) for the stand-alone SWAT+ model (i.e., without including the *gwflow* module). These are referenced in the text.