Summary:

This study implements a reservoir scheme into ParFlow, an integrated hydrological model that simulates many other hydrological components but lacks reservoirs and dams. Specifically, the study uses idealized and real-world situations to demonstrate the proposed scheme functions as expected, ensure mass conservation, and is computationally reasonable. The manuscript presents formulations on the new scheme and details how it works in idealized and real-world settings. The East-Taylor watershed is used as a test bed toward the end of the manuscript. Overall, I was excited about the topic and agreed to review, also because the work comes from two leading groups in hydrological modeling. After reading the manuscript, I didn't find my expectations to be met, primarily because the manuscript entirely lacks validation. Please see my detailed comments below.

We thank the reviewer for their careful review of our work. We are sorry to hear that the manuscript did not meet the reviewer's expectations. Based on the comments below we understand that there is a disconnect here between our intended purpose of the paper and the reviewer's impressions. We have provided a detailed response below which we hope will clarify our purpose and the reason we did not provide the type of validation that the reviewer was expecting. We would like to emphasize that we do provide validation and sensitivity analysis of the modeling capability we present here. We do not provide site specific parameter evaluation as that is a test of local model calibration and not the model capability itself.

Major Comments:

A major concern I have on this manuscript is the lack of validation. I waited until the end to see if some validation (real validation with observed flow or reservoir storage variation) is presented, but I found none. Validation is fundamental to any modeling study, which applies to reservoir modeling as well. A number of previous studies I note below have presented validation across the US and worldwide with available data on reservoir storage and release. If this manuscript were to be published in HESS, the authors should demonstrate model applicability across a range of basins in the US (and even beyond) and present a detailed validation of whether the model reproduces observed flows in the downstream of dams and storage variations. Substantial data exist for many reservoirs, at least those in the US (please see numerous studies I note below). Without such explicit validation, I'm not sure what the value of the present study is.

While we acknowledge the reviewer's concern regarding the validation of modeling studies, it's important to clarify that the primary objective of our study was not to develop watershed models, but rather to develop a generic reservoir modeling capability within ParFlow. As such the validation we provide in our paper is to validate the new modeling capability which we are presenting here. The goal of our test cases is to demonstrate that our implementation is sufficiently robust for integration into a broad range of practical applications. We would argue that the suite of test cases we provided do provide explicit validation of the new module's capability to faithfully implement reservoir releases that are specified by the user.

What the reviewer is suggesting here would not be a test of whether the model itself is functioning properly, but rather a test of whether the operating parameters that we input to

the model are realistic. This is of course an important test in any real-world simulation but is not a test of the model itself rather a test of the parameterization and model calibration. As the reviewer points out there have been other studies focused on rule curve development and validation, this is really a separate issue.

We hope that the reviewer will appreciate this distinction and the fact that we did design our test cases and model validation very thoughtfully. Building generic model capabilities such as this is a large undertaking and before we can validate site specific parameters the model itself must be validated and rigorously tested. That is the goal of our paper.

Page 2: Literature review is very limited in scope and doesn't include many of the early studies; reservoir modeling has been there for over two decades and numerous studies have attempted to incorporate reservoir in different kinds of models; I suggest that the authors acknowledge the previous students and expand the introduction; for example: (Dang et al., 2020; Haddeland et al., 2006; Hanasaki et al., 2006; Hanazaki et al., 2022; Shin et al., 2019; Vanderkelen et al., 2022; Wada et al., 2014) and many more. Even Hanasaki et al. (2006), one of the very early studies, has not been mentioned.

We appreciate the suggestion and acknowledge that some early citations were overlooked. We would be happy to expand our literature review in the revised manuscript to include these citations and a more thorough discussion of earlier work.

Introduction: It is generally well-written but it overly emphasizes implementation of dams in ParFlow. I suggest that the authors place more emphasis on: why this is needed (given existing models) and what is the novel contribution of the study?

Thank you for the suggestion we did not intend to overly emphasize ParFlow. We will edit the sections regarding need and novel contribution to be clearer, adding paragraphs as needed.

Page 5, Line 162 and elsewhere: how is dead, active, and flood storage considered in the model? Please describe in the manuscript.

We capture both dead and active storage using the minimum_release_storage attribute. Our model does not distinguish between flood storage and normal storage. We will include this clarification in the revised manuscript.

Equation 4: Are reservoirs considered to be of rectangular shapes?

Reservoirs modify flow within a single grid cell (or two adjacent cells), effectively represented as rectangles. We will add text to this section noting that reservoirs can only be resolved to the grid resolution and that we are not explicitly representing bathymetry here.

Line 257: Where are the "release curves" taken from? It is critical to elaborate this point.

The release curves used here are idealized and not drawn from observations. As noted in our response above o*ur intent with the paper is to demonstrate a generic modeling capability. The purpose of the rule curve in this case is to demonstrate and test the model's capability to handle releases under a range of conditions. We understand that this may not have been made clear in the initial manuscript and will revise the text accordingly here to note the source and purpose of the rule curves used here.*

Page 10, Line 267: How are the volumes of 7 MCM and 5 MCM determined? Please provide the basis for these numbers.

As noted above this is an idealized test case designed to test the capabilities of the reservoir module and not to match observations. We designed a reservoir sized to accommodate approximately a year's worth of rainfall (as mentioned in lines 266-267), with a minimum release storage chosen arbitrarily. The release rates were selected to reflect the expected intake rate of the reservoir during simulations. As previously noted, these decisions were deliberate as our tests were not aimed at replicating a real-world scenario, but rather at validating fundamental properties such as the reservoir's mass balance. Therefore, we believe that opting for rounded numbers within a physically realistic range allows readers to verify our expected outcomes using straightforward calculations. We will add additional clarification to the revised manuscript to make the purpose and design of our test cases clearer.

Line 275: there are certain model parameters noted here, which also seem arbitrary; the authors should present a sensitivity analysis to demonstrate that these parameters are reasonable/robust.

Here too we would like to emphasize the purpose of our testing which is exclusively to validate the performance of the reservoir module. The parameters in question do not directly interface with reservoirs therefore a sensitivity analysis of these parameters would not help our validation efforts. We want to emphasize here that we did conduct a sensitivity analysis of parameters directly linked to reservoirs and provide details of this in the following section.

L331, “maximum possible storage”: Is this calculated? Maximum storage is available in reservoir database such as the GRaND data.

The maximum possible storage is a user-supplied field. If the user opts to utilize GRaND data, it will align accordingly; however, if they choose a different dataset, it will align with that dataset instead.

Figures: Figures including 4 and 6 do not provide important information; these could be placed in a supplemental document.

The purpose of Figure 4 is to present a sensitivity analysis of parameters related to the reservoir, as requested in a previous comment. We believe this analysis is warranted; however, we understand that the purpose of this sensitivity analysis may not have been clear and will provide additional explanation of the purpose of these figures in the revised manuscript.

Figure 6 represents our real-life domain and provides important context for interpreting Figures 7 and 8.

Figure 5, “maximum possible storage”: I was expecting some validation with observed data but seems like this is just a comparison of simulated vs. “calculated” max storage. Please present actual validation.

Please refer to our earlier response regarding the purpose of this test case.

Line 355-359: seems like this is AI-written and not checked by the authors; for example, it reads: “insert citation”.

This line was not generated by AI; it was an oversight by one of the authors while gathering additional citations for the final edit. We will correct this in the revised manuscript.

L359: xx et al.?

We will correct this in the revised manuscript.

Figure 7: please present validation with observations; rich observations are available from USGS, US Bureau of Reclamation, and US Army Corps of Engineers for US reservoirs.

As noted in our previous responses, validating against observations does not meet the purpose of our manuscript. We are not testing reservoir parameterizations (i.e. what would be needed to calibrate to observations) we are presenting a generic modeling capability and as such have provided validation and sensitivity analysis of that capability directly.

Our objective is to show that our implementation is feature-complete and ready for application in real domains. We will emphasize this in the revised manuscript.

Figure 9: I wasn’t able to understand the point of this figure; looks like a supplemental figure.

This figure describes the domain for our performance test case. We will explicitly mention this in the figure caption to provide clarity.

L492: How is it novel? Please clearly describe this in the introduction and here.

Our implementation is the first support for representation of reservoirs in ParFlow’s class of models (fully-integrated physical hydrology models). It will both allow people using this class of models to make more accurate models, and for people interested in reservoir modeling to ask new questions that these types of models are best suited to answer. We will further emphasize this in the revised manuscript.

L498: how is it “user friendly”, especially compared to existing reservoir models.

User friendly is of course a matter of opinion but we appreciate the comment and can be more explicit on this point. We will provide a direct comparison to a typical user interface for adding reservoirs to clarify how our implementation enhances user-friendliness.

Overall, this manuscript can be a valuable contribution only if sufficient validation is presented to demonstrate that the reservoir module can reproduce observed release and storage.

We respectfully disagree with this opinion. Model development is a very important (and time consuming and challenging) part of the scientific process. We feel it is vitally important the model capabilities are rigorously validated and documented before application. It is important that we value rigorous model development as this is what allows us to put faith in their later applications. Absolutely any future application of this tool will require site specific validation and evaluation of reservoir parameters. However, the validation the reviewer is

suggesting is fundamentally different from what we presented here and is a test not of the model itself but of the site-specific model parameters.

Minor issues:

Line 23: there is a time stamp, perhaps put by AI

L243: please fix citations

Line 265: n.d.?

These minor issues were oversights while producing a final copy and will all be addressed. Thank you for your careful read.