This paper presented the SPUX-MITgcm framework, an approach to the calibration of a hydrodynamic model for a highly spatiotemporally heterogeneous observational dataset. The current form of the paper contains a lot of information and model exercises but was not presented in the most suitable way (missing key information or details) to guide the readers to well understand the value of the work. As such there are some difficulties in evaluating the work.

Major issues:

The information for the Particle filter, LSTM, are too general, missing quite a bit of detail. The paper could be improved by focusing on these methods used in the study including

1) a good introduction to PF with details on how it works, and its strength and limitations compared (not just statements, but compare the way how it works with other variational or KF) to educate the readers.

2) Details how the PF is configured and why, sensitivity analysis, convergence rate, etc.

3) The same for LSTM, LSTM framework should be introduced, how does it work, describe training and validation process and evaluation. 3) how does EMCEE Sampler work, how did you evaluate its performance?
4) I am also not quite sure about the BiLSTM's role in the proposed framework.

6) A better highlighting of the novelty and achievement of the work in the context with comparison to a similar or alternative approach. My impression is that it is a novel framework applied to the 3-D model, but not fully sure how effective and efficient it improve the simulation results. I think this can be significantly improved if the authors can re-structure the manuscript to highlight key information about the models used in the study.