General comment: I appreciate the authors' efforts in addressing my previous comments. I found most of my comments were well addressed. I recommend a minor revision with following additional comments for the authors to consider.

Regarding computation times, the calibration with ANN mapping over periods p1 and p2 took 95 hours. This calibration involved 350 iterations, corresponding to 350 calls to the adjoint model, and was performed using 16 threads. For comparison, a single adjoint model run takes approximately 16 minutes, whereas a direct model run takes around 5 minutes using the same number of threads.

Regarding computation times, the calibration with ANN mapping over period p1 took 31 hours. This calibration involved 350 iterations, corresponding to 350 calls to the adjoint model, and was performed using 10 threads. For comparison, a single adjoint model run takes approximately 6 minutes, whereas a direct model run takes around 1 minute and 30 seconds using the same number of threads.

Comment: Why 350 integrations is used for calibration?

You are correct, and we apologize for the confusion. The routing scheme is indeed parallelized, and we will correct this in the manuscript. What we intended to convey is that the routing scheme cannot be fully parallelized over the entire spatial domain, as it must be solved in a sequential manner from upstream to downstream. This distinction will be clarified in the revised version.

Since the time-stepping loop cannot be parallelized at all and the routing scheme cannot be fully parallelized at pixel scale over the entire spatial domain, as it must be solved in a sequential manner from upstream to downstream.

Comment: I cannot agree with the authors. Routing scheme can be fully parallelized over the entire spatial domain. A well paralleled river routing model does not need to be solved in a sequential manner from upstream to downstream.

Reporting the performance of the default forward model is not particularly meaningful in the context of this conceptual hydrological model. Our "default" model refers to a spatially uniform calibration of the parameters, which serves as a baseline for comparison.

Comment: I cannot agree with this response. It is critical to compare the calibrated model with the default model with default (not uniform calibration) parameter values. If there is no improvement from default parameters to calibrated parameters, then why spending efforts on make the model differentiable? I highly suggest the authors to report the performance of default model.

The shallow water model we refer to includes the convective inertia term in the momentum equation, which allows for a simple finite difference solution.

Comment: In this case, it is not non-inertial shallow water model. It is confusing.

Any hydrological or land surface model can be integrated into the smash framework, provided it is compatible with the constraints of automatic differentiation imposed by the Tapenade tool. In particular, the model must be written in a differentiable form and comply with Tapenade's supported syntax and structures. We provide technical guidance on how automatic differentiation is implemented within smash in the online documentation, in the section: Automatic Differentiation – Development Process Details.

Comment: My concern is that it is challenging to rewrite a complex hydrological model in a differential form. I am not aware of any example of differentiable model for complex hydrological models. Then the issue is why we need differentiable model for simplified hydrological model, which is computational cheap and can be easily run tens of thousands of times for calibration. It will be helpful for the authors to add some discussion on this limitation.