The article presents a methodology for local-scale compound flood modeling using global input datasets and a newly developed hydrodynamic model (SFINCS). The framework is applied to a coastal catchment in Mozambique that recently experienced flooding from two tropical cyclones (Idai and Eloise). The proposed methodology for developing local-scale models based on global datasets/inputs is very interesting, and I believe the paper could be an important contribution to the compound modeling literature. However, I believe some more analysis/discussion on the model validation is needed before this work can be published. Therefore, I recommend a moderate revision.

My main concern/issue with the paper is that the results presented in section 4.1 are not compelling. It seems like the local-scale model and the global CaMa model perform similarly well for the two historical cases, which calls into question why someone should go through the trouble of setting up a high-resolution local model if similar accuracy can be achieved with an existing global model. To be clear, I believe there is a lot of value in using a high-resolution local model for flood hazard analysis, I just don’t think the results presented in section 4.1 do a good job of showing the additional benefit. Can any additional validation data, performance metrics, discussion, etc. be added to this section to show more clearly the benefit of using the SFINCS model? The ability to efficiently set up and run local-scale compound flood models for any catchment across the globe is really promising, but we need more confidence that the local-scale model will provide higher accuracy compared to existing global models.
I have some other specific comments below:

3.1.1

I wonder if an ocean model with 2.5 km coastal resolution can adequately capture peak storm tides from TCs, which tend to produce extreme storm surges over relatively small geographic areas (cite).

Lines 142-149: I was confused here as the sources of the storm surge, wind setup, and tide heights were not clear. The 5.0 m max water level (and 3.8 m for Eloise) reported here is based on what? Gauge, high water marks, reports, models? What is the max water level predicted by the author’s global model? I see 4.0 m as the max surge estimate, but what about the total modeled water level? Also, how is the “operational forecast” generated? Is this another global model that the authors compared with? In general I think this paragraph needs to be re-written to be clear about how their model results compare with the results from other models or other sources.

3.2

Figure 3: What does “actual event discharge” mean? The river discharge based on the gauge records or the CaMa discharge? If the latter, I would call it model-based discharge since it is not the “true” discharge.

3.3

In addition to simulated flood extent, can any comparison be made using simulated vs
satellite-based flood depth? In low-lying regions, the flood extent could be similar between
the model and satellite, but the depth could be significantly different. I’m not asserting
that the author’s flood model is inaccurate, but just want to point out that a comparison
based on flood extent alone does not provide a complete picture about whether flood
dynamics are being accurately captured by the modeling framework.

4.1

It seems that although SFINCS simulates a larger extent of flooding than CaMa (due to
incorporation of pluvial runoff), CaMa consistently predicts higher flood depths for both
storms (except at location 5). I wonder if the authors have any ideas why CaMa would
estimate higher flood depth than SFINCS?