Reply on RC2

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.

We would like to thank the reviewer for the thorough review and comments, which we believe have led to an improvement in the manuscript. We are pleased to read that the reviewer believes our manuscript is an important contribution to the compound modeling literature. Based on the suggestions of both reviewers we have made several changes to the manuscript. Our response to the specific comments can be found in the paragraphs below.

Most importantly we have tried to highlight the benefits of the globally-applicable framework better in the abstract and conclusions, see updated text below. We improved the validation with two more flood extent images, one for each event. Furthermore, we have made some small improvements to the model setup by including more small rivers as boundary conditions to the SFINCS model and focus the analysis on the areas connected to the Buzi and Pungwe floodplains.

L18: To test the framework, we simulate two historical compound flood events, Tropical Cyclones Idai and Eloise in the Sofala province of Mozambique, and compare the simulated flood extents to satellite-derived extents at multiple days for both events. Compared to the global CaMa-Flood model, the globally-applicable model generally performs better in terms of the critical success index (-0.01 – 0.09) and hit rate (0.11 – 0.22), but lower in terms of false alarm ratio (0.04 – 0.14). Furthermore, the simulated flood depth maps are more realistic due to better floodplain connectivity and provide a more comprehensive picture as direct coastal and pluvial flooding are simulated.

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).

We agree with the reviewer that the resolution of the hydrodynamic model is important to correctly capture storm surge peaks as local bathymetry could have significant effects on the surge. However, Dullaart et al. (2020) have shown that the Global Tide and Surge Model (GTSM) forced with ERA5 was able to simulate the storm surge peak for five TC events with a bias ranging from -22 to -2 cm. Furthermore, for that case of Beira which is located on a wide continental shelf (~140 km) a 2.5km grid resolution is deemed sufficient to capture the storm. We argue that this level of uncertainty is adequate for a globally-applicable model framework considering the bias in available global elevation datasets.

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.

We have rewritten the paragraph to be more clear about the data sources.

L136: Due to a lack of observations from coastal water level gauges, a quantitative validation of the simulated water levels is not possible, but some general observations about the simulated data can be made. The maximum simulated water levels in GTSM during both events (5.0 m+MSL during Idai; 3.8 m+MSL during Eloise) occurred close to neap tide and are caused by surge (3.2 m during Idai; 2.6 m during Eloise) and wind setup (1.2 m during Idai; 0.7 m during Eloise). The maximum surge and its *timing during Idai (4.0 m) are in line with the operational forecast of 4.4 m based on the HyFlux2 model forced with NOAA Hurricane Weather Research and Forecast atmospheric data (ERCC, 2019; Probst and Annunziato, 2019). As tide and wave effects are not simulated by this model, total water levels are not available for comparison. In comparison with the tidal constituents of International Hydrographic Organization (IHO) station at the Port of Beira as retrieved using the Delft Dashboard (van Ormondt et al., 2020), the highest astronomical tide is expected to be around 3.8 m while our simulations result in 4.5 m, indicating an overestimation. This has however little effect on the maximum water levels which occurred close to neap tide.*

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.

Thanks for noticing this ambiguous wording. We used “actual event discharge” to distinguish it from the climatological discharge but agree that the wording is not clear. We have rephrased the caption to be more specific:

L136: Figure 3: SFINCS boundary conditions during cyclone Idai (left column) and cyclone Eloise (right column) for total sea level from GTSM and ERA5 (top row); discharge from CaMa-Flood (center row); and spatial average runoff from ERA5 (bottom row). The full lines show the total water levels and discharge, as used for the validation, see Section 3.3.1. The dashed lines show the tidal water level component only (top row) and normalized discharge to match the climatological mean (center row), as used in the compound driver analysis, see Section 3.3.3. Only one coastal location (H4) and the two main rivers (Q1 - Buzi and Q4 - Pungwe) are shown to improve the readability of the plots. The location labels in the legends correspond to the locations as shown in Figure 2B.

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.

Unfortunately, hardly any ground observations of water level / depth are available for our case study. We also checked if ICESAT-2 data are available for any of the events but found there is no overpass between 18-21 March 2019 and 24-27 Jan 2021. However, we do agree with the reviewer about the need for flood depth observations and the potential for satellite-derived observations and discuss this in Section 4.1, 4.3 and in the conclusions in Section 5:

L439: Our results also demonstrate that a commonly used metric to evaluate flood models such as the critical success index can mask large differences between model results and should be evaluated together with the false alarm and hit ratios and inspection of the geographical patterns and differences. An additional comparison with flood levels, if available, would allow for an even more comprehensive validation (Stephens and Bates, 2015; Wing et al., 2021).

L512: We recommend investigating whether remote sensing, e.g. satellite laser or radar altimetry data, can be used to validate extreme inland and nearshore water levels (Andreadis et al., 2020; O'Loughlin et al., 2016; Urban et al., 2008).

L545: We also reiterate the importance of observed water levels for a more comprehensive comparison of flood simulations.

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?

In section 4.1 we argue that this is because of limited connectivity between neighboring floodplain cells in CaMa-Flood. CaMa-Flood is in essence a quasi 2D model as the main connection between cells is along the river network defined by a single downstream neighboring cell, while some water is exchanged with other adjacent cells through its bifurcation scheme (Yamazaki et al., 2014). This scheme is however too limited to represent the connectivity in the large low-gradient floodplains of the Buzi and Pungwe rivers. Therefore, more water is retained in the Pungwe in Buzi river cells compared to the SFINCS model. We have made some changes in section 4.1 to clarify this.

L378: The difference between both models in the Buzi floodplains is likely due to the limited connectivity between floodplains of neighboring cells in the CaMa-Flood model through its so-called bifurcation schema. This scheme is however too limited to represent the connectivity in the large low-gradient floodplains of the Buzi and Pungwe rivers. This can be seen in the downscaled CaMa-Flood flood maps, which show unrealistic sudden local drops in flood depth at the interface of unit catchments during cyclone Idai (Figure 4A) and larger simulated water levels in the Buzi in CaMa-Flood compared to SFINCS (Figure 5A/B).

References

Dullaart, J. C. M., Muis, S., Bloemendaal, N., and Aerts, J. C. J. H.: Advancing global storm surge modelling using the new ERA5 climate reanalysis, Clim. Dyn., 54, 1007–1021, https://doi.org/10.1007/s00382-019-05044-0, 2020.

Yamazaki, D., Sato, T., Kanae, S., Hirabayashi, Y., and Bates, P. D.: Regional flood dynamics in a bifurcating mega delta simulated in a global river model, Geophys. Res. Lett., 41, 3127–3135, https://doi.org/10.1002/2014GL059744, 2014.