Dear anonymous reviewer,

We kindly thank you for critically reading the submitted manuscript and providing helpful comments and questions on current shortcomings.

Hereafter, we will react to your major comment first. Subsequently, we will go through the minor comments whereby we will implement the grammar and spelling comments in the revised manuscript and thus are not considered here explicitly. For better readability, we first state your comment in black and subsequently add our reaction in blue.

Major comment

I found the discussion of the model results in Sections 3.1.3 and 3.2.3 to be too descriptive; i.e., they were simply a rehashing of the statistics presented in Tables 1, 2, and 3. Rather than summarizing statistics, I would like to know if there's some greater insight that could be gained from these runs. I would like the authors to put a little more effort into interpreting the results. Even a few sentences would be helpful.

We thank the reviewer for this constructive criticism. While the output statistics are the foundation of a good discussion, we can follow the reviewer's reasoning. In the revised manuscript, we will have elaborated more extensively on the model results and on their wider implications for inundation modelling and hazard assessments.

Minor comments

- Define NSE before using acronym.
  We will provide the necessary definition in the abstract
- Recommendation: cite the work of J. Syvitski in this area. Perhaps "challenges of a) establishing a modular and flexible model coupling framework (e.g., Syvitski et al., 2014) and b) applying...".
  Many thanks for providing this reference and a way to include in the manuscript. We will add the reference to the revised manuscript at a sensible location in the text.
- It's odd to title this section "GLOFRIM 2.0" when the subheadings are of the models that are included within it. A better section title might be "The coupling framework and its component models". Section 2.1 could then be titled "GLOFRIM 2.0".
  We thank the reviewer for this helpful remark. Indeed, the current structure may be misleading as it implies the models were part of GLOFRIM. This is, however, not the case and thus we will update the manuscript structure accordingly.
- Appendix A is not present.
  The appendix is provided in the assets of the manuscript under the header supplements. However, from the file name provided there it does not become clear that it is appendix A despite it was the original name of the uploaded file. We will try to fix this together with the Copernicus publishing team.
- Some of the authors of this paper are also authors of the PCR-GLOBWB model. I have no correction to offer here, but this just feels a little odd. Perhaps the paper might be stronger if an outside model had been used.
  We can guarantee that the fact that one author of this manuscript also worked on the PCR-GLOBWB model (and in fact another one on the CaMa-Flood model) has no influence on the set-up or results of the study presented here.
- D-Flow FM is referenced here and in Figure 1, but not discussed in the paper.
  Many thanks for pointing at this. Indeed, the model is not discussed in the paper as it was not used. A description can be found in the supplement of the paper as well as in the online resources. What is misleading is, that the associated header is 2.2.4 which is part of 2.2 "The models". This is not correct and 2.2.4 should stand individually as 2.3 which will improve clarity.
- What are the time steps of the models?
  We kindly thank the reviewer for mentioning this ambiguity. The time step of the hydrological model is daily and the hydrodynamic models (CaMa-Flood and LISFLOOD-FP) use an adaptive timestep scheme. The data between the models is exchanged at a daily time step. We will update the revised manuscript accordingly.
- Why are KGE and NSE chosen? One or two sentences on why these are the appropriate measures would be useful.

Many thanks for this critical remark, we will append the information to the revised manuscript. The metrics were chosen as they are very common in discharge analysis in general. Also, the KGE and its components provide information about a wide range of analytics. The NSE was chosen as it is particularly sensitive for deviations in peak discharge and thus powerful for flood hazard assessments such as shown in the submitted manuscript. We will add a sentence to the manuscript explaining this.

- Why is KGE > 0.7 significant?
  Thank you for this comment. As with many hydrologic performance metrics, there is no clearly defined threshold when performance is deemed good or to have "skill". Since the maximum value of the KGE is 1.0, KGE values above 0.7 can, in our view, be interpreted as skillful as most uncalibrated discharge simulations typically range in lower spheres.
- What are KGE_α and KGE_β? I can guess what they are, but they should be explained.
  Many thanks for reporting this inconsistency in the manuscript. They are indeed explained in section 3.1.2 but are named α and β there, respectively. In the revised manuscript, we will make sure that the nomenclature is consistent throughout the document.
- I applaud the authors for making their code publically available. In the interest of open science, are the model runs available, as well? We thank the reviewer for those kind words. We were already discussing whether we want to make (parts of) the model input data to make the simulations as open and FAIR is possible.