Reply to reviewer 1

Dear reviewer,

We thank you for your positive assessment of our work and the constructive suggestions. We have addressed all major and minor comments below, with changes implemented in the revised manuscript.

l. 11: Abstract: when you speak about integration in a real-time early warning system, you should add some numbers on computational time of your 1h simulation

We agree and have now added specific runtime examples from our results.

l. 101: t the 1D domain decomposition for the 2D is somehow unclear, explain more

We have expanded the explanation to clarify that the domain is decomposed along one spatial dimension (typically the longer axis of the rectangular domain) into equal-sized subdomains per GPU.

l. 131ff: if you use 2m resolution, 1 cell has 4m², Berlin area of 900km2 then would require~225 mio. cells, why is your number ~double as high, similar for the other resolutions

Thanks for pointing this out. The reported cell numbers (418.9 M at 2 m, 67 M at 5 m, 16.7 M at 10 m) refer to the cells in the full rectangular raster (left panel of Figure 1) which completely encloses the administrative boundary of Berlin plus a small buffer. However, during the simulations, only the cells inside the actual city mask are physically active. We have now added some explanation to clarify this.

l. 183, 205: rain is not a boundary condition but a source term, check in the document

Replaced throughout the manuscript with "source term" or "rainfall input".

sec. 3.1: the performance gain using several GPUs is several times quite poor, give some explanation why, any parallel overheads?

The decrease in scaling efficiency with a higher number of GPUs can mainly be because of increasing inter-GPU communication overhead (ghost-zone exchange) and load imbalances caused by heterogeneous wetting/drying patterns and building masks across sub-domains. We have added some text to explain this in the manuscript.

l. 360: compare Berlin to the similar approaches of other federal states such as North Rhine-Westphalia

We have now added a comparison of this approach to what is implemented in NRW as an example.

l. 366: these deep uncertainties must be mentioned / discussed, otherwise I would delete or rephrase

it is now removed.

l. 383: criterion for affected persons, is this your definition or from the literature

We define "affected persons" as individuals residing in areas impacted by flooding, i.e., those whose dwellings or immediate surroundings are inundated ("get wet feet"). This definition is consistent with established literature in flood risk assessment. For example, Winsemius et al. (2013) define the affected population as those living in cells with positive water depth in a given flood scenario. Similarly, PBL (2018) use this approach to quantify exposed populations. We have now added these references to clarify that our criterion follows common practice.

-Winsemius, H.C., et al. (2013). A framework for global river flood risk assessments. Hydrology and Earth System Sciences, 17, 1871–1892.

-PBL Netherlands Environmental Assessment Agency (2018). The Geography of Future Water Challenges.

Fig 8: how is the number of effected persons computed, is it per cell, how can it be smaller than 1

The number of affected persons is derived directly from the WorldPop 2020 population dataset. The RIM2D simulations max water depths are aggregated to the WorldPop grid resolutions and overlayed with the data. A WorldPop cell is classified as affected if the maximum simulated water depth within it exceeds 0.1 m ("wet feet" threshold). The entire population assigned to that cell by WorldPop is then counted as affected. Values below 1 person arise from the WorldPop data because many cells, especially sparsely populated areas such as parks, industrial zones, or outskirts, contain only a fraction of one person according to the WorldPop disaggregated data.

Sec. 3.4: comment more on uncertainties, friction, infiltration, sewer system? as your model is that fast, you could do parameter variations eg for friction and infiltration

We have now added more discussion on the mentioned uncertainty sources and cited some studies that have done sensitivity analysis with similar models. And yes, due to the short runtimes, sensitivity analysis with hundreds (or thousands) of simulations is feasible (e.g. https://doi.org/10.5194/nhess-25-975-2025), however, a full sensitivity analysis or calibrated risk assessment for Berlin was not added as the primary goal is to demonstrate the technical capability of RIM2D, that state-wide, high-resolution, physically based, real-time pluvial flood forecasting is achievable, rather than to deliver definitive risk maps for Berlin.

l. 426: you argue that such speeds are only achievable through multi-GPU; but such speed are also achievable through HPC cluster with many cores / CPU; add this here and also earlier where you argue similarly

added.

in the context of real time prediction, you should also mention that there are several promising machine / deep learning / artificial neural network approaches

We have now added this with citing some recent works on this topic.

Minor:

- sometimes you speak about the state of Berlin, sometimes about the city, I suggest to unify

Unified.

- l. 97: can you give a reference ?

Added.

- unify all headlines, sometime 1st small, sometimes capital

Unified.

- l. 119: it is larger 3.8 or 3.9 check

Population corrected to 3.89 million (as of end of 2024)

- sec 2.2: add a reference to Fig 1, in principal to each figure

Reference added.

- l. 141: give a reference and / or explain

Reference added.

- l. 144: unit -1/3 should be exponent

It is now fixed.

- sec. 2.4: add references to Tab 1+2, in principal to all tables

Reference added.

- l. 336: sometime you write dx = 2 m, sometimes as here without dx -> unify in document

Unified.

- further typos, minor comments are in an attached pdf, no need to comment on them

All further typos and small remarks from the attached PDF have been corrected.

Reply to reviewer 2

Dear Reviewer,

Thank you for taking the time to review our manuscript and the constructive suggestions. All comments have been addressed below, and the revisions have been implemented in the manuscript.

This is a nice contribution to the literature on Berlin flood forecasting, and merits publication. The benchmarking with the June 2017 flood is particularly useful. It would be helpful if some probing sensitivity analyses could be carried out, e.g. on the sewer capacity. Suppose there had been a major rainfall event a few days earlier, how reliable would the forecasting for the June 29-30 event have been? The authors should comment on the reliability of flood forecasting under alternative extreme conditions.

We thank the reviewer for this suggestion. In the current study, our primary objective was to demonstrate the technical feasibility of real-time, state-wide, high-resolution flood forecasting using the multi-GPU capabilities of RIM2D. Nevertheless, we agree that discussing the sensitivity of predictions to hydrological preconditions (e.g., sewer saturation) can strengthen the manuscript. We have therefore added some explanations to the Uncertainties section (3.4) where we explicitly address these topics.

Sewer capacity sensitivity:

As the reviewer points out, sewer capacity can play an important role in urban pluvial flood modelling. In our work, sewer capacity is calculated following standard German design practice, in which sewer capacity is derived using a 2-year, 15-minute design rainfall (DWA-A 118E), consistent with the approach described by Apel et al. (2024). We have now added a discussion of uncertainty ranges reported for RIM2D and also other similar high-resolution models in response to varying parameters (sewer capacity, roughness, resolution), citing previous sensitivity studies. Moreover, we also added clarification that the present simulations assume an initially empty sewer system, as no information on antecedent pipe filling was available for this event. Additionally, to assess how reduced sewer capacity affects the results, we ran a set of simulations with lowered capacities and compared the resulting flooded areas and water depths. The corresponding results are now summarized in Section 3.4.

Reliability under alternative extremes:

In the revised manuscript, we now also discuss the reliability of RIM2D under alternative extreme conditions. The main uncertainties in such scenarios stem from model inputs such as rainfall, DEM, sewer infiltration capacity, etc. while the hydrodynamic solver itself contributes comparatively little to the overall uncertainty. Because of RIM2Ds short runtimes, ensembles representing alternative antecedent and extreme scenarios can be generated operationally, which increases forecast reliability.

We also note that RIM2D has already been applied successfully to several other extreme flood events, demonstrating stable performance under very different hydrometeorological conditions, such as the 2021 Ahr Valley flash flood (Khosh Bin Ghomash et al., 2024), the 2023 pluvial flood in Braunschweig (Khosh Bin Ghomash et al., 2025), and urban inundation tests in Dresden (Apel et al., 2024). These applications confirm that RIM2D is reliable beyond the June 2017 Berlin event.

-Apel, H., Benisch, J., Helm, B., Vorogushyn, S., and Merz, B.: Fast urban inundation simulation with RIM2D for flood risk assessment and forecasting, Frontiers in Water, 6, 1310 182, 2024.

- Khosh Bin Ghomash, S., Apel, H., and Caviedes-Voullième, D.: Are 2D shallow-water solvers fast enough for early flood warning? A comparative assessment on the 2021 Ahr valley flood event, Nat. Hazards Earth Syst. Sci., 24, 2857–2874

- Khosh Bin Ghomash, S., Apel, H., Schröter, K., and Steinhausen, M.: Rapid high-resolution impact-based flood early warning is possible with RIM2D: a showcase for the 2023 pluvial flood in Braunschweig, Nat. Hazards Earth Syst. Sci., 25, 1737–1749