Dear colleagues - editor and reviewers,

We thank sincerely both reviewers for the thorough and constructive evaluation of our manuscript. We value the time and expertise invested in the reviews. In particular, we appreciate the suggestions regarding structure and positioning our contribution within the state of the art. We are committed to addressing the concerns in a substantially revised manuscript.

This response letter is organized in two parts:

1. High-level response to the major issues raised by both reviews.
We first address the overarching points that were consistently highlighted, and outline the concrete steps that we will take in the revision. In particular, our revision will focus on:

1.1. Restructuring the manuscript to follow a clearer organization aligned with NHESS customs;
1.2. clarifying novelty and attribution, i.e., explicitly positioning our methodological contributions relative to established methods and software components, refining related work surveying and adjusting the abstract/introduction to represent our contributions more clearly;
1.3. improving overall presentation, including substantially revising figures, captions, and the exposition of methods and results to enhance readability and interpretability;
1.4. unifying the experimental setup across sections, such that results are comparable and the role of different inputs and assumptions is transparent.

We acknowledge that our simulations rely on some proprietary software, reflecting the current state of the art. As in other cases that use such software, this requires dealing with some accessibility issues; however, this does not prevent full computational reproducibility, i.e., fundamental scientific accountability. In addition, we provide all model inputs, outputs, and configuration details required to verify the analyses presented in the manuscript.

2. Point-by-point responses to all reviewer comments.
In the second part of this letter, we respond individually to each comment raised by Reviewer 1 and Reviewer 2. For each item, we describe the planned change and indicate where it will be reflected in the revised manuscript.

Overall, this will result in a substantially improved manuscript. We hope that the planned revisions will address the reviewers' concerns, making it appropriate for the readership of NHESS.

Sincerely,

Michael Perk, Phillip Keldenich, Tobias Wallner, and Sándor Fekete

1. High-level response to the major issues raised by both reviews

1.1 Manuscript structure and organization

Both reviewers considered the current manuscript structure to be unconventional, making it difficult to follow the workflow and to distinguish between data description, methods, experimental setup, and results. We propose a revised manuscript structure as follows.

1. **Introduction**: Motivation, followed by a revised Related Work / comparison to state of the art, and then a clear statement of the challenges addressed and objectives of this study (separately for hydrodynamic modeling scalability and for sensitivity analysis).
2. **Study Area**: retained in essence, but streamlined (in particular, reducing meteorological/climate background that is not directly required for the subsequent experiments).
3. **Methods**: a dedicated methodology section describing (a) the domain subdivision and dependency-aware scheduling problem formulation, (b) the quadtree-based refinement approach and sub-grid sampling.
4. **Data and Simulation Setup**: a unified section that (a) introduces SCALGO and clearly specifies which data are obtained from it, (b) documents all input data used for the hydrodynamic model (including a consistent and transparent precipitation definition), and (c) describes and motivates the design of each subsequent experiment.
5. **Results**: presentation of results strictly aligned with the experiment definitions from Section 4, avoiding intermixing of setup/methodological details with findings or discussions.
6. **Discussion**: interpretation of results in the context of prior work, including explicit treatment of limitations and threats to validity; still following the two main research blocks “simulation scalability" and “sensitivity analysis”.
7. **Conclusion**: concise summary of findings and implications.

This restructuring is intended to reduce redundancy, improve readability, and ensure that readers can clearly trace (i) which methods are proposed or adopted, (ii) which data and assumptions are used in each experiment, and (iii) what conclusions are supported by the presented results.

1.2 Novelty, attribution, and positioning relative to prior work

Both reviewers noted that the current manuscript could more clearly separate established building blocks and third-party tools from our own contributions, and that parts of the abstract and introduction misrepresent novelty. In the revision, we will (i) tone down the claims in the abstract and introduction, (ii) explicitly attribute all software and algorithmic components, and (iii) state our contribution more precisely in the Methods and Discussion.

In contrast to previous watershed partitioning approaches, which primarily split stream networks at confluences or based on reach length and rely on flow accumulation only to balance drainage areas, our work focuses on optimizing the dependency structure induced by the static stream network. We introduce a partitioning heuristic that seeks to minimize unnecessary chains in the dependency graph. This allows more room for parallelism. We couple this heuristic with an *optimal* scheduling strategy for a fixed number of machines. Our revised version of the manuscript will make this clear and will note that the multi-objective problem of jointly optimizing partitioning and scheduling deserves more attention, opening avenues for future work. Moreover, our framework integrates watershed partitioning with an adaptive grid (quadtree) and adaptive time stepping. We are unaware of studies that combine all aforementioned strategies to enable large-scale simulations at the 1m resolution achieved in our work.

There is another, fundamentally important aspect of our contribution: We are not simply refining individual components, but putting everything to use for an overall framework that is capable of achieving (to the best of our knowledge) previously unmatched performance. That does require a substantial amount of refinement and integration, which we present, with a crucial outcome: It is now possible to perform *practically scalable* simulations and analyses for large-scale regions at high resolution within manageable time. That opens the door for other scientific and practical aspects, like the sensitivity analysis that we provide, with immediate implications for real-world resource management.

1.3 Overall presentation and figures

Both reviewers highlighted that several figures and parts of the exposition are difficult to read and interpret (e.g., map panels too small, colors hard to distinguish, insufficiently self-contained captions).

We plan to substantially improve the presentation. In the revision, figures will be redrawn with larger panels, less data series per panel, clearer color scales, and more informative captions that define all symbols, units, and thresholds. Where map-based difference plots are not sufficiently informative at the domain scale, we will complement or replace them with clearer figures or quantitative summaries so that the reported differences and trade-offs can be assessed more directly.

Beyond the figures, some detailed reviewer comments point to places in which descriptions are perceived as too brief, ambiguous, or insufficiently connected (e.g., definitions, parameter choices, and references). These issues will be addressed by tightening the writing, adding missing explanations and citations, and ensuring consistent terminology. The revised general structure of the manuscript, as proposed above, will help to increase the clarity. Detailed point-by-point responses to the reviewers' comments are provided below.

1.4 Consistency of data and experimental setup

Both reviewers noted that several experiments use different rainfall inputs and configurations, potentially making comparisons difficult and, in places, possibly obscuring the purpose of individual experiment sets. We will conduct additional experiments to produce a consistent, clearly defined baseline setup, with any deviations introduced explicitly as controlled sensitivity cases (and justified as such). The revised manuscript will include a consolidated "Data and Simulation Setup" section that documents input datasets, rainfall definition (duration, total amount, temporal pattern, and spatial distribution), key modeling assumptions, and the mapping of each figure/result to the corresponding experiment definition. Where required, additional simulations will be run; given the reduced runtimes we achieved. These reruns are feasible within the revision process and most of the simulations have already been computed at the time of writing this response - again demonstrating the scalability of our overall approach.

2. Point-by-point responses to all reviewer comments

Here we address all individual comments by both reviewers. The original remark by the reviewer will be formatted in *grey and italic*. Where appropriate, we refer to the high-level changes we propose above.

Comments by Reviewer 1

General comments:

- *Introduction: Numerous complex concepts are mentioned without leading to the own research question. A lot of direct quatations. Should be significantly shortened. Redundancies should be removed. Ends with numerous examples of the state of the art, which is rather unusual. It would improve the structure if the study objectives would follow the state-of-the-art instead.*

  The introduction will be rewritten as proposed above.

- *The structure should be improved. E.g. datasets are mentioned in "1D/2D modelling" (l.183 ff.) and "Refining the mesh" (l. 274). Use the standard structure and describe your dataset in a section "Data" and your software and new code implementations in "Methods" as usual. Mention all software. SCALGO is not described anywhere. The model is not clearly described. The current structure is confusing the reader unnecessarily.*

  The structure will be revised, see above.

- *The authors present as one of their main innovations what seems to me as a standard algorithm to delineate sub-catchments from a digital elevation model as a basis for parallelization. Is this really something new and has not been done before? Additionally you fail to mention that the performance increase is highly dependent on the order of nodes. E.g., if you have a lot of dependent sub-division the potential for parallelization is significantly limited. Furthermore, the error increases with subdivision.*

  As discussed above, we will explain the novelty of our work and its difference to related papers more clearly.

- *The second innovation (quadtree refinement) turns out to be an add-in which you did not develop. However, in the abstract it sounds as if you had developed this ("We present significant progress…. Breakthrough…). Again, this might be caused by the structure and I am not able to distinguish which methods you developed and which pre-existing ones you "just" implemented and tested.*

  We will tone down the abstract accordingly and highlight the actual contribution of our work, as discussed above.

- *I understand that it is difficult to present the results but the current display is not convincing. Maps and plots are too small, colors are difficult to distinguish*

  We will update the figures and attempt to make them more readable, as discussed above.

- *Because of proprietary software the study is not reproducible.*

  We addressed this concern above and are committed to allow for independent verifiability of our analyses. Unfortunately, we are not aware of any open-source product with the capabilities we require.

Detailed comments:

- *Line 1-34: Streamline introduction? Focus on floods caused by high intensity precipitation. What has erosion to do with the topic? What do “ocean“ floods have to do with the topic?*

  We will shorten the introduction as proposed above.

- *Line 74 – 88: This seems like a rather unorganized collection of snippets all related to the idea how hydrodynamic models can be improved by either disrectisation or other parameters (e.g. roughness, sub-grid porosity). Please try to streamline, summarize and to be more concise.*

  Line 89ff: For my taste this part contains too much“name-dropping” and mentioning of complex methods without further explanation. This does not really help the reader who is not completely familiar with the development of hydrodynamic models. There seems to be redundancy, too. e.g. when talking again about porosity treatments (l. 100). Please streamline the Introduction.

  We will shorten the introduction as proposed above and highlight only the work that is highly relevant for our study.

- *Figure 1: You could probably combine plots top left and bottom.*

  We will work on improved figures as proposed above.

- *Line 155: I assume "g" is the standard gravity but please list it below.*

  No problem.

- *Line 186: What is SCALGO?*

  We will explain what SCALGO is and what data we obtain in our proposed “Data and Simulation Setup” section.

- *Line 189 ff: This is essential for flood modeling. Please give more detail. What was the input data for the machine learning? Soil moisture products have usually a very low quality, especially on a 1m grid.*

  We will explain this in some more detail. A full elaboration is the subject of a separate paper by some our colleagues and cited in our manuscript. That paper is already publicly available (with a link provided in our paper) and currently under review at HESS. https://doi.org/10.5194/egusphere-2025-3091

- *Figure 2: I think this Figure never gets referred to in the text. Please add more information to the caption. What is cj, what is v? What is the flow direction? Up or down? Why does orange need to be removed? Why is the bottom node on the right not red? Enlarge the symbols for the nodes so one can distinguish the colors better*

  We will work on improved figures as proposed above.

- *Line 207-233: If I understand correctly, this is the description of a typical algorithm that delineates hydrological subcatchments from a DEM. I think it would be helpful to mention this, or to skip this section entirely and just say "We delineated subcatchments". The headwater catchments can be computed independently. The rest needs to be updated accordingly depending on the flow graph.*

  What we provide is more than a set of delineated subcatchments, as we aim to optimize our partition for the subsequent scheduling problem (which we also solve to provable optimality, unlike other work we have seen). We will revise this section and highlight the novelty of the proposed algorithms.

- *Line 210: Is this the total number of cells?*

  This is the case for a 1x1 grid, but the technique can also be applied to grids of other sizes as well as simplified graph structures. We will be more explicit in the text.

- *Line 211: Does the threshold depend on the RAM size?*

  Yes it does. As we use an HPC model, our main concern is VRAM on the GPU. We will state this more explicitly.

- *Line 217: Again, SCALGO not explained*

  See above.

- *Line 222: What would happen if a cell is flat, or evenly flows into e.g. three downstream cells? There are typically different ways to treat this. For one-directional flow typically a random choice is included which does not make the results reproduceable (when not setting a seed).*

  This will be explained in more detail in the SCALGO section. To the best of our knowledge, in these degenerate cases, a random choice is made. For reproducibility, we include the used flow graph (which was used throughout all experiments) for our entire ROI as a supplement.

  The discussion of which static stream network to use for partitioning is interesting; we will extend the paper's discussion of this issue. Of primary importance is to take culverts into account, as they play a crucial role in our ROI. Thus, a node can have edges to non-neighboring cells, too. Furthermore, if depressions are not accounted for properly, the resulting partition would likely not accurately represent the actual direction of flow as our current setup does.

- *Line 225: Why do you do this? How do you decide on the threshold?*

In our scenario, this flow graph contains ~370 million vertices. Even though each partition can be derived with linear runtime, this is still an enormous overhead. The non-filtered graph would allow a more fine-grained partition, which was not necessary for our experiments.

- *Line 233: This basically refers to subcatchments, right?*

  Yes, but the subcatchment from a node further downstream may include subcatchments further upstream, which we remove from the experimental region. Figure 4 (top left) gives an impression how this can look in practice. The entire region in this figure is a subcatchment of our ROI which is further partitioned into 4 subcatchments. Each simulation j with an incoming edge (i,j) replaces part of the subcatchment by a hydrograph representing the entire set of ancestor nodes of i.

- *Line 236: I don't understand this step. Can you please explain further? Why do you need water transfer between watersheds?*

  A subcatchment upstream can still have an influence further downstream. This amounts to replacing parts of a subcatchment by hydrographs that represent the water flowing into a specific part of the region. We will be more explicit in the text.

- *Line 244: Does this step ensure a minimum size for subcatchments?*

  This is a side effect. The main reason is the size of the non-filtered graph, as mentioned above.

- *Line 245: What do you mean with cell-level data?*

  We are considering a subset of nodes, rather than the complete graph, so we are unable to extract the watershed from the graph alone. Thus, we query a highly-optimized data structure that has all data available.

- *Line 245: Now you are talking about watershed extraction, but isn't that the same as you just described before. What is the difference? Here you are you using a algorithm from SCALGO. The previously described method was developed by you. It is really difficult to differentiate where you developed own code and where you used existing methods.*

  We implemented the partitioning ourselves and it theoretically works "as is" on the complete graph. Because we are using a filtered graph to decrease model-building time, we query the watershed from SCALGO instead of generating it directly from the flow graph. This is a technical detail that is not really important for the actual algorithm. We will clarify this in the new data section.

- *Line 251: Describe model in a "Method" section TUFLOW and add a reference.*

  We will follow this suggestion.

- *Line 257: Another preferred flow path are tilled agricultural areas. Wouldn't it also make sense to have a finer resolution in uneven/rugged areas?*

  That is an interesting suggestion. In our experiments, we have not considered it because we do not have any distinction between “grassland” and agricultural areas. We can, however, mention this in the text.

- *Line 274: Hausumringedataset" -- you mentioned it before. What is it? → “Data” section*

  We will describe it there.

- *Line 281: Subgrid-sampling: Can you give some kind of reference for this? Or an example? Its hard to imagine how this works.*

  We will try to provide additional references.

- *Line 289: You based your flow graph delineation on a 30m-DEM and now you have another finer DGM for SGS? Which one? You should consider having a "Data" section and you should describe which data set you use for what and why.*

  No, our flow graph and DGM are both on a 1m DGM. We are not sure of this impression that some 30m-DEM is involved, as we are not referencing this in our text. If there is any mixup, we would be happy to address this.

- *Line 296: "well-established" -- References please*

  We will provide references.

- *Figure 4: Bottom left too small. Maybe make the maps larger and the histograms smaller. Its hard to tell if these dept differences are relevant?*

  We will address the figures as described above.

  What if they happen exactly in the critical spots?

  This is an interesting point. The short answer is that we cannot tell. Of course, even small differences can have a huge impact if the area is particularly important. Our goal was to show that the simulations give similar outputs even after subdividing with our approach. We will try to pick a few locations that might be important and compare the water depths over time for each of the simulations to give a better idea of the accuracy of the models.

- *Line 310: Does this value come from KOSTRA? Is it a realistic structure to have uniform 50mm on 20km²*

  We will redo these simulations to use rain from the KOSTRA dataset instead.

- *Line 314: So in this example you can compute 2 catchments parallel?*

  Yes, with two machines and an optimal schedule (meaning there is no room for improvement in how we use these two machines for this particular carving), we can achieve 7h of simulation time, compared to 13h.

- *Line 317: Maybe a cumulative distribution function would be easier to read?*

  Yes, we will consider this. See above for the discussion on how we will update the figures.

- *Line 319: Is this because of the differently sized cells?*

  No, we do not use differently sized cells yet. Instead, we define an inflow/outflow boundary according to (and perpendicular to) the stream in the flow graph. This boundary has a specified length, but there might be other parts of the boundary where water can escape/transfer, which are not captured by our model, e.g., when water does not follow the static flow graph.

- *Line 322: Does this mean, the more parallel regions I compute, the larger the error gets? How much could I parallelize until the error gets to large?*

  We have not analyzed this. In general, we assume that the error does not increase with the width of the tree, but with its depth (meaning that multiple simulations pass water to the next simulation along a path in the dependency graph). If each boundary transfer incurs an error, this error will accumulate. We will add that to our discussion and 'threats to validity' section.

- *Line 324: Can show an example, why this impact is not significant?*

  We will try to derive one from the "important locations", see above.

- *Line 333: You mean, you use your algorithm but still don't parallelize? Is the speed up then solely to the smaller memory requirements for the individual subcatchments?*

  The speedup is due to multiple GPUs working on the same output in parallel. Even when performing all simulations on a single GPU, one can still derive what would have happened if more GPUs were available. To do this, we solve a scheduling problem (whose inputs are available *before* running the simulations) and derive an optimal schedule from the solution. The total runtime is then computed from the schedule and the individual simulations' actual runtimes. Part of our contribution is to provide an optimal schedule for a given dependency graph, regardless of the number of machines. Often, only a few, say k, machines (in this case 2) are sufficient to get the optimal schedule for all k' >= k. In this case, even with an unlimited number of GPUs, the presented subdivision cannot be simulated faster than within ~7h. We will clarify this point by extended explanations in the methods section and the discussion.

- *Line 346 Why do you change the rainfall amount here compared to the previous example with 50mm?*

  We will redo the simulations here to use the KOSTRA rain instead.

- *Figure 5: "lower is better" -- why is this in the caption of each subplot? Bars are really narrow and hard to read. Colors (e.g. shades of brown) are hard to distinguish.*

  We will work on improved figures as proposed above. The term "lower is better" was used to avoid misinterpretations for readers who only look at the plot and do not read the full text. In our experience, this is helpful to some readers, and does not really impede others.

- *Line 347: The quadtree increases simulation time, right? Then how do you weigh simulation time against precision?*

  The quadtree with a base resolution of 8m increases simulation time and precision compared to a 2D model with 8m resolution. However, we do not compare only against 8m, but also against 1m resolution. Compared to the 1m model, the quadtree model often saves more than 90% of the runtime while being more accurate than the 8m/16m models. Our goal is to derive a model that is both fast and accurate in areas that matter.

- *Line 370: "spatial variability of rainfall". I think that this correction is just reducing the rainfall amount but it stays spatial uniform rainfall.*

  It is uniform in space but not in time, as we used hyetographs derived from the KOSTRA dataset in our simulations. We will make this clearer in the data section.

- *Line 375: 1D and 2D model are part of TUFLoW? I think a proper model description is missing. Which model is it?*

  We will make this clearer in the data and simulation setup section.

- *Figure 6: So, here its just about quadtree and not about parallelization, right? Couldn't you directly use quadtree on the whole grid? zoom in?*

  Using a quadtree on the whole area could be interesting, but is unfortunately infeasible due to the memory requirements. We aim to achieve a 1m output (meaning a DGM1 model would require 370 million cells). Our preliminary experiments revealed that with our hardware and the amount of precision required by the refinements we make to the grid, we can expect to be able to simulate models with 20-30km^2 without running into memory issues. Note that quadtree and sub-grid sampling both significantly increase the simulation's memory usage, as they require additional data structures and management overhead.

- *Figure 7: The maps are so small, that the differences are hard to distinguish. Maybe you could and not annotated in first subplot Pi Pj*

  We are not quite sure about "you could and not annotated". We will replace the figures and rather show detailed views of certain areas at the expense of not being able to combine all simulations in one figure. Please let us know if you had a different suggestion.

- *Line 408: Can you explain, why you opted for this setting?*

  We will discuss this in the new discussion section.

- *Line 413: It would be interesting if you could give an estimate for the memory requirement for the 1m-simulation on one core*

  We can try to derive numbers here. We do use a GPU-accelerated model, so we have both RAM and VRAM requirements.

- *Line 418: A linear cascade of dependend subcatchments would then significantly slow down the parallelized computation. You should mention that somewhere.*

  Yes, we will mention this explicitly, as stated above.

- *Line 437: I am wondering, why you don't parallelize on many more cores? For a sensitivity analysis you typically need to run the simulation many times. Is the speed up so significant that you could now run a complete sensitivity analysis?*

  Each individual simulation is already massively parallelized. We did not carry out further subdivision, as this was not necessary for each individual simulation. In principle, a subdivision would be possible for even further runtime reduction, but would require additional error analysis, as discussed in Section 5.2.

- *Line 458 ff: It is difficult to see this in the Figure 8. Generally, in Figure 8 the different layers are overlaying each other. Is there a better way to visualize this? Maybe one plot for each scenario even though that would mean a lot of plots… About which inundation depth we are talking? Or do you just count everything above 0 cm? Can you assess this e.g. by the amount of inundated raster cells to visualize the tipping point?*

  This is a good suggestion. We will think about visualizations like this. The plot actually does not overlap, but shows what "new" areas are flooded compared to the previous run, as increasing the amount of rain only increases the extent of the flooded area.

- *Figure 8: Top left could be smaller, while the maps of the flooded areas should be larger. E.g. top left, I can't find any yellow part (highest rainfall). Brown and orange are very hard to distinguish. Scale bar missing. Mention which rainfall amount you chose.*

  There are actually several yellow areas in the plot, but they are very hard to see without zooming in. We will try to find other ways of visualizing this.

- *Line 472: Did you mean "middle right"?*

  Yes, we will fix that.

- *Line 472: Does this mean that you applied 150 mm in three hours uniformly once for each subcatchment? What would happen if the rainfall affects two or more subcatchments partially but simultaneously?*

  Yes, but we will adjust the rainfall amount based on the subcatchment's actual size. The volume is always the same. If it affects two or more subcatchments, we cannot tell what would happen, presumably a combination of both.

- *Line 481: Are you implying that your experiment is 7 times fast than the 1m resolution? Wouldn't it be a "slow-down" factor? Maybe its just a matter of formulation.*

  This may be a misunderstanding: 7x speedup here would mean 1/7 of the simulation time. Should we clarify?

- *Line 500ff: You did not really do this, did you? Why not? How long would it take? What would be the memory requirements?*

  We did get a 1m output (using quadtree + SGS) but can not simulate the whole region within a single model. We cannot provide exact memory requirements, but we know that our machine (48GB of VRAM) could simulate ~50 km^2 with 1m resolution. So we would assume the memory requirements to run a 1m model on this scale to be about 8x what we have available in our machines. We are not aware of any hardware in existence with that much VRAM.

- *Line 505: "Our quadtree refinement strategy" Didn't you use a pre-existing library? Or did you develop this method?*

  As stated above, we will make it clearer what our contribution is. With "refinement strategy", we indicate "what and where" to refine, not the actual HD engine that runs the simulation on this grid.

- *Line 517: Could you give any information on how high these margins could be? Do you have any info on this?*

  We do not have any information on this. What we are aware of are heavy rain events in our ROI where the rain data publicly available (e.g. from DWD) does not represent the actual amount of rain that came down at this time (validated by sensor data we have available in the area).

- *Figure A1: Top left: Headline missing. Generally the maps are too small. Maybe you should supply them as shapefiles.*

  We will add this to the caption. The plot resembles the amount of water in the 1m simulation. We provided all data used to generate these maps as .tif files.

- *Figure A2: "lower is better" -- I would not write this on top of every plot*

  We can remove this.

Comments by Reviewer 2

General comments:

- *General structure of the paper – in my opinion the manuscript lacks a clear structure. After the introduction the authors describe the study area. Then they include a section on the basics of 1/2d-hydraulic models basically presenting text book knowledge which I would question if this is needed in the detail presented. This section is then followed by a section which could be called a method section. Afterwards some results are presented, but also details on the simulation setup are given. This is then followed by a section on stability – again partly methods, partly results. And finally there are some summaries, although they are called discussion and conclusions. So I would suggest to introduce a method section as well as a section clearly defining whatkind of simulations have been conducted (including setup, input data etc) and for what purpose. I additionally would strongly recommend to streamline the simulation setup, since in the current version of the manuscript basically each part uses different precipitation characteristics. And finally please introduce a section where the results are presented and not a mix of method, simulation setup and results.*

  We will revise the manuscript structure and organization completely, as proposed above.

- *Introduction – this section certainly needs to be streamlined to be more focused on the target of the study. In large parts it reads like a series of more or less random quotes of a set of papers without a clear link between them.*

  We will streamline the introduction, as proposed above.

- *Presentation of results – there is a general need to improve the description and presentation of the findings of the conducted analysis. For most of the figures, I have a hard time to see any differences between the simulations (e.g. Fig 4, lower left panel, Fig 5, upper panels, Fig 7, Fig 8 almost all panels, Fig A1). I would suggest to rethink if maps depicting the absolute/relative difference are of any use considering the scale of the model domains.*

  We will work on improving the overall presentation, as proposed above.
  Specifically:
  Fig 4: We will remove the lower left panel.
  Fig 5: We will plot the differences in water depth in a different way that makes it easier to read. Specifically, we will not distinguish between the three different regions but unify the plot and also unify the different quadtree configurations into a single bar.
  Fig. 7: This will be replaced or removed. We will try to either show detailed results for specific locations or remove this figure altogether. The main purpose was to convince the reader that we obtain results that can be combined into a unified output.

- *Discussion of results – the discussion section is not a discussion section but a summary (actually followed by a further summary named conclusion). This definitely needs to improved. The study is neither the first study working with the parallelization of model subdomains nor the first study using local grid refinement. Hence, there is a need to place the outcome of this study in the context of the findings of already existing studies. So to make it short – the outcome of the discussion section should give a clear answer to what actually the 'breakthrough' is, which the authors claim that the present study has made and which was not already there before.*

  We will revise the structure and discussion content as proposed above, revise the related work survey, and state our contributions in the abstract and introduction more precisely.

More detailed comments:

- *Lines 134 ff: What is the added value of mentioning mean climate characteristics of the study domain. I do not think that this information is of any relevance for the presented study, hence I would remove it completely.*

  Lines 143 ff: Again, what is the added value of the CatRaRe events? All simulations are conducted based on relatively short (3h) design storm events. What information does the reader gain knowing that there has been a series of events with a duration of 24h or even 48h – if these events are not within the scope of the study.

  We will streamline this section and focus on information that is relevant to our study.

- *Lines 155 ff: I have the feeling that eq 1 and 2 are identical, describing the flow in X direction. I actually anyway would have only expected two equations.*

  We will remove the duplicate formula and shorten this section.

- *Line 186: Although I know what SCALGO is, there is a need for a reference and a short description here.*

  We will explain this in our proposed "Data and Simulation Setup" section.

- *Line 310: 50 mm in 3h is about a 20 yr event when using KOSTRA (V 2020) as baseline. What was the motivation for this rain amount and duration? Would the deviations in the model results increase with increasing precipitation?*

  Lines 345 ff: Here 30 mm in 3 h (~3yr event) are used. Why this number and not again 50 mm in 3h or why not always using a 100yr event like it is used later? Again, what would be the impact of the input on the deviations in the model simulations?

  We will conduct new experiments using KOSTRA rainfall data to resolve both issues.

- *Fig 5 - upper panels - maybe scatter plots would be better. In section 4.2 its mentioned that the mesh refinement was conducted on streams, roads, buildings, culverts and model boundaries – why only refinement of streets and streams is included in the figure? Additionally, what do the authors refer to when they state as panel heading: 'Lower is better'? (actually also included in almost all other figures). If we look at the computing time it certainly seems that the runs with the lower (coarser) resolution are performing better – but I doubt that this is the case for the other results.*

  The refinement of culverts, buildings and model boundaries is always necessary to be able to reliably run our comparison. So all models presented in the plot have a high resolution in these areas. We will make this clear in the text. In particular, (1) the culverts are required to accurately represent the points at which 1D and 2D simulation interact, (2) the buildings are excluded from the simulation, thus if we don't use a fine-grained resolution, we may have holes much larger than the actual buildings due to a misaligned grid and (3) refinement at inflow/outflow boundaries is required to accurately represent a neighboring inflow/outflow.

- *Lines 457 ff: The description of simulation setup and inputs is a complete mess. According to KOSTRA, a 100yr event for 3h roughly is about 65 mm.*

  The simulations for the current section 6 were not intended to present "real-world" rain events, but instead to analyze the ROI and how stable the outputs of our simulations are when changing different parameters of the input. For instance, in the first set of experiments, we vary the amount of rain to identify tipping points after which a substantial area is newly flooded that can be analyzed to identify weak points for disaster mitigation.

  Where do the other inputs (up to 150 mm/3h) come from. What is actually the temporal assumption of the rainfall input.

  The rain amount 150mm/3h was intentionally set very high to find areas that are at risk of being flooded when a specific region experiences heavy rainfall. The intention of this section is to show that our improved models allow fast simulation times even in these extreme scenarios, which allow new means of analyzing risk and planning of countermeasures.

  Additionally, what about the role of soils. Looking at the large scale of the simulations (as well as the substantial part of the domain covered by forests and agricultural land) what were actually the assumptions of infiltration. And how is the variation on infiltration ability of soils due to different preconditions reflected in the simulation setup?

  The soil is important for our simulations. We used the Horton model and will add a description on which infiltration parameters we used in the new data section.

- *Fig 8 – Looking at the upper right panel, I have the feeling that the largest rain amounts (yellow color) are leading to the least flooded areas. Does this in anyway make sense or what actually is shown in the respective panel? Bottom row – left panel: Are the regions the same as the regions identified by the subdivision process?*

  The plot actually only shows the "newly" flooded areas, as they would otherwise overlap. So in a 100mm/h event, all colored areas (including the yellow areas) will be flooded. With up to 50mm/h all blue, orange and green areas will be flooded. We will try to make this clearer in the caption and text and potentially replace this figure, as mentioned above.

  Right panel – if I understood the setup correctly, 150mm/3h were applied to different subregions. This is a huge amount, hence I would expect to see many flooded areas marked in blue/turquoise color - where the precipitation input was provided over almost all of the domain – why is this not the case?

  There may be a misunderstanding that we should address. The legend labels like "0+1+2" indicate flooded areas when the rain came down in either (0 or 1 or 2). We will make this clearer in the caption and text. 150mm/h is indeed a lot, but it was only applied on a (relatively small) subarea of the ROI. The dark and light blue regions seem to not flood any area uniquely (meaning they flood no areas that are not flooded by another rain location), but both indeed flood several areas of the city, e.g. shown in neon blue, grey and yellow.

- *Lines 484ff and Fig 9 – Here the authors use 50 mm over 4h. Why was this chosen and how would the result differ, if the simulation setup would be similar to the previous setups. Additionally Fig 9 indicates only one critical point when using base resolution 8 and 1m, respectively. Is this the only critical point in the domain? What, if a grid refinement method would have been used (which actually was the scope of the study) – would the result be the same as for the 1m base resolution case?*

  We will address this by running additional experiments with precipitation from the KOSTRA data set.

  As for the particular importance of this critical point, we were able to validate this in direct communication with local experts: The reason is that major parts of the inner city flood when water runs across the street instead of through the culvert into the stream. It is likely that the results would be the same as for the 1m simulation when using a simulation setup with an adaptive grid that uses high resolution at this intersection only.

- *I do not provide detailed comments on the discussion and conclusion sections, since both should be completely worked over.*

  We will revise these sections completely, as proposed above.