Editor decision

Dear Yue Zhu, Paolo Burlando, Puay Yok Tan, Christian Geiß, and Simone Fatichi ,

Thank you for submitting your manuscript on 'Improving Pluvial Flood Simulations with Multi-source DEM Super-Resolution' and for answering the reviewer's comments.
Two reviewers have examined the changes. As you can see from their comments, one is not fully convinced about the revised version.
Some issues remain, which I kindly ask you to take into consideration. Among other things, this refers to i) clarification of the scope of your study, ii) additional explanation of the simulation experiment.
As a next step, please respond to the reviewer's comments and upload the revised, marked-up (track changes) manuscript version. Additionally, you have to upload a clean new version of the manuscript.
If you have any questions regarding the procedure or interpretation of the comments, please do not hesitate to contact me for clarification.
Please also note that this decision does not necessarily imply acceptance of the manuscript in the journal NHESS. It depends on your edits to the manuscript based on the referees' comments and the editor's comments on the revised version.
I look forward to receiving the revised version of your manuscript.

Best regards
Kai Schröter

Dear Prof Schröter,

We sincerely appreciate the time and effort the reviewers and editor invested in evaluating our manuscript and providing insightful comments. We have carefully considered each comment and have made corresponding revisions to the manuscript, including i) clarification of the scope of our study, ii) additional explanation of the experiment, iii) adding a README file to the openly shared dataset and codes, and iv) correction of Figure 6.

To facilitate the review process, we are including a detailed point-by-point response to the comments, indicating how each issue has been addressed. In the below point-by-point response, our responses are in *italic*, and the text in the revised manuscript is marked with quotation marks. Changes made to the manuscript are highlighted in the track-changes file.

We believe that these revisions have further enhanced our manuscript and hope that it will be considered favourably for publication. Thank you for considering our revised submission. Please do not hesitate to contact me if further information is needed.

Sincerely,
Yue Zhu

Report #2

Thank you for your revision. The manuscript has been substantially improved, and the figures are much better now (with the exception of Fig. 6). However, most of my comments have not been adequately addressed and still stand:

- With respect to the relevance to NHESS and the overall structure of the paper: I'm unconvinced that NHESS is the best journal for your work. (Yes, you have found—and now included—some more NHESS references; however, these are not DEM super-resolution focused, and clearly these were not foundational to your work.) Is the paper focused on improving DEM super-resolution? Then clearly it should be in a more ML-focused journal. Or is the paper about how simple DEM performance metrics fail to capture pluvial model performance? If so, then multiple pluvial models should be considered and a broader set of DEMs should be used—and probably an article like this would be better suited to a journal like HESS. Both of these would be interesting articles. However, combining these topics into a single article as you have confuses things.

We thank the reviewer for the thoughtful comment. We appreciate the opportunity to clarify the scope of this manuscript, and to explain why we believe NHESS is a very fit and impactful venue for this work.

While this study investigated machine learning methods for enhancing the spatial resolution of DEM data, its primary objective is to demonstrate how these methods can improve the resolution of terrain characterisation to a level, which allows to enhance significantly the accuracy of flood simulation. In this sense, the study sits at the intersection of natural hazard modelling – pluvial flood in the specific case – and processing of geospatial data needed for natural hazard assessment. We believe this aligns very well with the interdisciplinary scope of NHESS journal, which prominently includes the enhancement of modelling and assessment of natural hazards. Although the NHESS references we added were not purely focused on DEM super-resolution, they explore and highlight the influence of input data characteristics on hazard assessment outcomes, thus making the case for the foundational scope of this manuscript, which, ultimately, aligns well with the scope of NHESS.

Therefore, we believe this integration is actually a strength of the manuscript, because it demonstrates how the different performance of DEM super-resolution models can lead to differentiated improvements of their real-world hazard modelling (in this case hydrological) applications. We firmly believe that this has both scientific and applied relevance, because, to the best of our knowledge, the extent to which the quality of DEM spatial resolution – and its enhancement – contribute to the accuracy of pluvial flood simulation lacks quantification in the existing literature. It seems to us that this question must be highly relevant to the scientific community working on earth system science and natural hazards. Accordingly, to better clarify the scope of this study, we revised the manuscript as follows:

(line 105) *"[…], this study aims to investigate the effectiveness and quantification of how pluvial flood simulations can be improved by using a deep learning-based DEM super-resolution construction method, which incorporates multispectral imagery, including the near-infrared band, as additional input. Accordingly, we develop an integrated methodological framework that allows for enhancing input data quality for practical improvements in flood simulation performance,* *specifically quantifying the extent to which the proposed method of DEM resolution enhancement can contribute to improved pluvial flood hazard simulations."*

Regarding the use of a single pluvial model, our intention is, coherently with the scope of the article, not to benchmark flood simulation models, but to compare under controlled simulation settings the effect of different DEM enhancement techniques based on super-resolution methods. The employed flood simulation model is widely established in urban flood risk analysis, as put in evidence by the cited literature, so that we consider it to be suitable for representing a reference model on which to test how input data quality – the DEM resolution – affects its performance. Please also note that in reference to the hydraulic part, most urban flood models are solving either directly or indirectly some form 2D shallow water equations, and differences are often more related to the numerical implementation rather than to fundamental process description. Therefore, as far as the selected model is representative of this model category, it is likely to provide similar sensitivity to DEM as other urban flood models. Including multiple flood models, as suggested, would indeed broaden the scope of the research and enhance the generalisability of the results. However, it would also dilute the main message of this study and generate a significantly different study that goes beyond the scope of this article, as this study showcases for the first time the impact of different hyper-resolution DEM quality on pluvial flood simulations. The inclusion of additional pluvial flood models would be better suited for a follow-up study that focuses, for instance, on the relative performance of different flood simulation models based on a given DEM input quality. However, to account for this suggestion, we added a new point in the discussion section as follows:

***(line 390)** "[…], further tests could assess the effects of super-resolution DEM under alternative rainfall scenarios and using additional flood simulation models to assess if DEM input quality has some level of model dependency. While many flood models share the same fundamental equations to solve for flow processes (Guo, Guan, & Yu, 2021), such extended analyses would likely broaden the scope of the study and enhance the generalisability of the results."*

- Thank you for sharing the code. I looked briefly at the code, and there is no README. This does not facilitate review or reuse and is unsatisfactory in my view. Also, there are many unnecessary cache files that have been included. What dependency versions were used?

Thank you very much for this comment. We updated the dataset by removing unnecessary cache files and adding a README file with information on dependency versions. The updated dataset can be accessed following this link: https://zenodo.org/records/15212783

- I'm still unclear on how the simulation experiment was divided between the two study areas. Were they each trained separately (i.e., separate model weights)? If so, this implies the proposed method only works when the input 10 m DEM is already available (for the training)? What would the practical applications of something like this be? Instead, I think it would be more useful to train on one study area and apply these weights/inference on the other dataset.

Thank you for this comment. We would like to clarify the rationale of the experiment setup, also providing more details of the experiment. The two study areas were trained and tested independently on two different datasets, correspondingly having different model weights. Training and validating models in different datasets is regarded as a widely accepted experiment setup in many deep learning applications for remote sensing products (Huang, Zhao, & Song, 2018; Mou, Ghamisi, & Zhu, 2017; Shao, Zhou, Deng, Zhang, & Cheng, 2020; Yang et al., 2018). We chose this experiment setting because the goal of this study is to demonstrate method robustness across different contexts, not transferability, and evaluate method performance within each context to validate if it works properly under different geographical conditions, allowing the assessment of whether the proposed method could effectively learn terrain-specific representations that improve flood simulation accuracy.

While we acknowledge that training on one area and testing on another may provide insights into cross-site generalisation, such an approach without any adaptation is known to perform ineffectively, particularly in spatial learning tasks where input distributions differ significantly (Tuia, Persello, & Bruzzone, 2016). In this study, the two datasets vary in terms of acquisition sensors and terrain morphology. Applying a model trained on one directly to the other would introduce domain shift, leading to degraded performance and potentially misleading conclusions about model robustness. To clarify the settings of the training process on the two datasets, we made the following modifications in the manuscript:

***(line 205)** "All the models were trained and validated separately in the UK and China datasets, facilitating the evaluation of model performance in learning terrain-specific representations across different geographical contexts."*

Also, the objective of our study is not to develop a universal or domain-invariant model, but to evaluate the utility of super-resolved DEMs in improving flood simulation accuracy within a given area. Demonstrating that this is the case on two significantly different terrain and urban landscapes is in its own a result, which has practical implications in real-world applications: for instance, limited high-resolution DEM data can be partially available for specific regions, and this framework demonstrates that a subset of such data is sufficient to train effective models locally. We agree that model transferability and generalisation are promising areas for future exploration, and this is very likely achievable by training the model with diverse global DEM datasets. We now highlight this in the discussion section as a direction for further research, particularly with pretraining on diverse global DEM datasets and fine-tuning them for local applications, as follows:

***(line 385)** "Moreover, this study trained and evaluated models separately for two different regions characterised by different geographical and terrain contexts. Future work, such as pretraining on diverse global DEM datasets and fine-tuning for specific local applications, could explore approaches to improve model generalizability, thus supporting broader applicability in regions with limited high-resolution DEM data."*

Guo, K., Guan, M., & Yu, D. (2021). Urban surface water flood modelling – a comprehensive review of current models and future challenges. *Hydrology and Earth System Sciences*, *25*(5), 2843–2860. doi: 10.5194/hess-25-2843-2021

Huang, B., Zhao, B., & Song, Y. (2018). Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sensing of Environment*, *214*, 73–86. doi: 10.1016/j.rse.2018.04.050

Mou, L., Ghamisi, P., & Zhu, X. X. (2017). Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(7), 3639–3655. doi: 10.1109/TGRS.2016.2636241

Shao, Z., Zhou, W., Deng, X., Zhang, M., & Cheng, Q. (2020). Multilabel Remote Sensing Image Retrieval Based on Fully Convolutional Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 318–328. doi: 10.1109/JSTARS.2019.2961634

Tuia, D., Persello, C., & Bruzzone, L. (2016). Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 41–57. doi: 10.1109/MGRS.2016.2548504

Yang, Y., Dong, J., Sun, X., Lima, E., Mu, Q., & Wang, X. (2018). A CFCC-LSTM Model for Sea Surface Temperature Prediction. *IEEE Geoscience and Remote Sensing Letters*, *15*(2), 207–211. doi: 10.1109/LGRS.2017.2780843

- It seems Fig 6 was mistakenly replaced with content from Fig 5 during the revision.

Thank you very much for pointing this out. We have updated Fig 6 with the correct figure.