RC2: 'Comment on nhess-2024-207'

We sincerely appreciate the time and effort the reviewers invested in evaluating our manuscript and providing insightful comments. We have carefully considered each comment. 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. Thank you for considering our revised submission.

Attached are additional comments. Thank you for the nice paper, it was a pleasure to read.

We will make the following response to the comments and revisions to the manuscript:

(line 15) we will modify the abstract with a clarification about the research context: "Accurate flood simulation remains a significant challenge in many flood-prone regions, particularly in developing countries and urban areas, where the availability of high-resolution topographic data is especially limited." and about the quantification of performance metrics: "We evaluated the performance of the super-resolution DEM in flood simulations. Compared to conventional methods (e.g., bicubic interpolation), the simulation results demonstrated that our approach significantly improved the accuracy of flood simulations, with a reduction in the mean absolute error of floodwater depth from 0.137 to 0.119 (-13.1%) and an increase in the IoU for inundation area predictions from 0.207 to 0.302 (+45.9%)."

(line 25) We will tone down the introduction as: "The occurrence of severe floods has been on the rise, partly influenced by climate change, which contributes to more frequent extreme rainfall events (Tabari, 2020)."

(line 30) We will add the citation: "open datasets of DEM data with global coverage are predominantly available at raster resolutions coarser than 30 x 30 meters (Marsh et al., 2023)"

(line 36) We can rephrase the introduction to existing methods for improving the spatial resolution of DEM to differentiate with other similar statements: "Methods to enhance the spatial resolution of DEM data have been widely adopted across various geospatial applications to improve risk estimation. These advancements have significantly enhanced the accuracy and reliability of natural hazard mapping, including flood prediction (Löwe & Arnbjerg-Nielsen, 2020; Tan et al., 2024), landslide modelling (Brock et al., 2020), volcanic flow assessment (Deng et al., 2019), and snow avalanche forecasting (Miller et al., 2022)."

(line 46) We can rephrase the introduction to data fusion-based approaches for improving DEM spatial resolution for better clarity: "During the fusion process, tools such as elevation error maps or weight maps are commonly used to assign importance to each DEM source, ensuring that higher-quality data has a greater influence on the final output. However, these methods often introduce inaccuracies by altering elevation values and failing to address edge effects (Okolie & Smit, 2022), such as abrupt transitions or mismatches between overlapping DEM datasets."

(line 52) As suggested, we will remove the acronym SISR for Single Image Super-Resolution due to its infrequent use in the manuscript. We did not adopt the acronym SR for super-resolution, as it consists of only two letters and could be confused with spatial resolution, which is frequently referenced throughout the manuscript.

(line 50) we consolidated the sentence introducing deep-learning DEM super-resolution methods: "The implementations of deep learning-based super-resolution methods have been shown to substantially improve the performance of remote sensing applications (Ling & Foody, 2019; Shang et al., 2022; Xie et al., 2022) and promote the utilisation of data that was previously underutilised due to limited spatial resolution (Zhu et al., 2021), including the applications of enhancing low-resolution DEM data (Demiray et al., 2021a, 2021b; Jiang et al., 2023; Kubade et al., 2020; Z. Li et al., 2023; Yue et al., 2015; Zhou et al., 2023, 2021, 2021)."

(line 60) We will revise the literature review to provide a richer discussion of related studies, elaborating on their methods and performance. Additionally, we explained how our approach improves upon these previous works: "For instance, Demiray et al. (2021) utilized Generative Adversarial Networks (GANs) to upscale low-resolution DEMs (50ft) to high-resolution DEMs (3ft), although this study demonstrated the potential of adversarial training in spatial resolution enhancement, GANs are known for unstable in training, facing challenges such as mode collapse and vanishing gradients (Jabbar et al., 2021). Zhou et al. (2021) introduced a double-filter deep residual neural network, leveraging residual learning to improve feature extraction and enhance the accuracy of reconstructed DEMs. More recently, Li et al. (2023) proposed a transformer-based deep learning network for upscaling DEM across multiple upsampling factors (e.g., ×2, ×4), showcasing the effectiveness of attention mechanisms in capturing long-range dependencies and spatial relationships. Building on the advances of these existing methods, we refine a DEM super-resolution method by employing a computationally efficient architecture with attention mechanisms to achieve accuracy and robustness."

(line 65) We will rewrite the section "1.2. Multi-source deep learning for remote sensing applications" to remove redundant or unnecessary sentences and added related citations. The modified paragraphs are as follows:

"The benefits of integrating multi-source inputs in remote sensing applications have been increasingly recognised, as the combination of complementary data sources enhances the robustness and reliability of model performance (J. Li et al., 2022). For instance, Shen et al. (2019) developed a deep learning-based model for drought monitoring, which employed multi-source remote sensing data as input, including DEM data, and meteorological and soil data. Lu et al. (2022) proposed a deep learning framework taking Google Earth imagery and point of interest heatmap as input data for urban functional zone extraction. Blöschl et al. (2024) integrated additional bathymetric information into the DEM to enhance national-scale flood hazard mapping.

With respect to the input for DEM super-resolution, it can be argued that, solely replying to a single source of low-resolution (LR) DEM input can be an ill-posed task, as high-resolution details can hardly be accurately reconstructed without additional reference information (Yue et al., 2016). Studies have been made to include additional features generated from low-resolution DEM data. For instance, Zhang et al. (2023) calculated terrain gradient maps based on DEM data to guide the optimisation process of a Convolutional Neural Network (CNN)-based DEM super-resolution. Zhou et al. (2023) proposed a terrain feature-based CNN for DEM super-resolution, which extracts slope and aspect from low-resolution DEM data and deploys them as additional features for model inputs and loss function.

Besides generating additional features based on low-resolution DEM, efforts have also been made to fuse different data sources to offer fine-granular details related to terrain features to bring performance gains. One example following this direction is found in Argudo et al. (2018), who examined the feasibility of combining natural colour aerial images together with low-resolution DEM data as input to train a CNN for producing high-resolution DEM, suggesting improved performance compared with interpolation-based methods. Tan et al. (2024) introduced a deep learning-based DEM upscaling network that uses high-resolution optical images to predict elevation differences, and then fuses these predictions with the original DEM data through additional convolutional layers. It should *be noted that these studies mainly employed natural colour images for feature fusion. In contrast, multispectral images can provide further features from non-visible wavelengths, such as near-infrared, allowing for more detailed and specialised analysis. This is supported by Chen et al. (2013), showcasing the effects of utilising multispectral bands of satellite images on improving the performance of an interpolation-based DEM densification method. More recently, a few attempts have explored the effects of integrating low-resolution DEM with remote sensing imagery for DEM super-resolution. Gao & Yue (2024) used the red band of Sentinel-2 images to provide auxiliary high-frequency information for DEM super-resolution training. Paul & Gupta (2024) incorporated 3-band satellite images with low-resolution DEM to develop a GAN-based DEM super-resolution model."*

(line 100) We will rephrase the second main contribution of the study as: "…(ii) the use of publicly open datasets ensures the generalizability of the method, especially for DEM-related applications in data-scarce regions;"

(line 110) Thank you for the comments on revising the overall significance of the study from "offering an exemplary pathway to address the issue of lacking high-resolution DEM for reliable risk assessments in the context of land use planning and disaster management" to focus on emphasising "improve flood simulation". However, given that we intend to keep the section on pluvial flood simulation evaluation in the main manuscript, we would like to argue that the section on quantifying the improvements in pluvial flood simulation also indicates the potential of improving broader applications across various domains that rely on high-resolution DEMs for reliable spatial analysis. Therefore, we think the original statement reflects the broader significance of the study, showcasing its value beyond flood simulations and positioning it as a methodological advancement applicable to multiple disciplines.

(line 110) We thank you for the comment about adding sentences guiding the reader into the method section. Accordingly, we will add a guiding sentence as follows: "To improve the spatial resolution of DEM data for enhancing flood simulations, we proposed a deep learning-based DEM super-resolution method. This method employs the Residual Channel Attention Network (RCAN) (Y.Zhang et al., 2018) as the backbone structure and incorporates a tailored multi-source input block to leverage multi-sourced input data, contributing to improved performance in reconstructing high-resolution DEM data."

(line 120) As suggested, we will modify the last sentence of this paragraph to clarify our modifications on the backbone structure as follows: "However, since RCAN is developed for image super-resolution tasks on single natural colour images, we tailored the structure of its input module to handle inputs from different data sources."

(line 125) As suggested, we will modify section 2.2 "Multi-source and multi-scale input data fusion" to avoid redundancy in method descriptions as follows: "This study adopts a multi-source method for DEM super-resolution, utilizing the RCAN as the backbone structure. The proposed method, referred to as RCAN-Multispectral (RCAN-MS), incorporates a tailored multi-source and multi-scale input module, which is the key distinction from the original RCAN. This input module enables the integration of high-resolution multispectral satellite images with low-resolution DEM data, leveraging the complementary information from both sources to reconstruct high-resolution DEMs with enhanced accuracy. Multispectral satellite images contain information captured across various spectral bands, including both visible light and invisible bands, which offer a wealth of information about surface materials, vegetation coverage, water bodies, and other landscape features (Carrão et al., 2008), making them ideal for compensating for the coarse information in low-resolution DEMs. By combining high-resolution multispectral imagery with low-resolution elevation data, deep learning models can access a more comprehensive feature set, facilitating the reconstruction of detailed topographic information."

*(line 135) Thank you for the suggestion, we will revise Figure 1 to emphasise the location of the multi-source input module in the model structure, which now corresponds to the text: "The tailored multi-source input module is integrated into the model structure before the first layer of the RCAN backbone structure (**Error! Reference source not found.**)."*

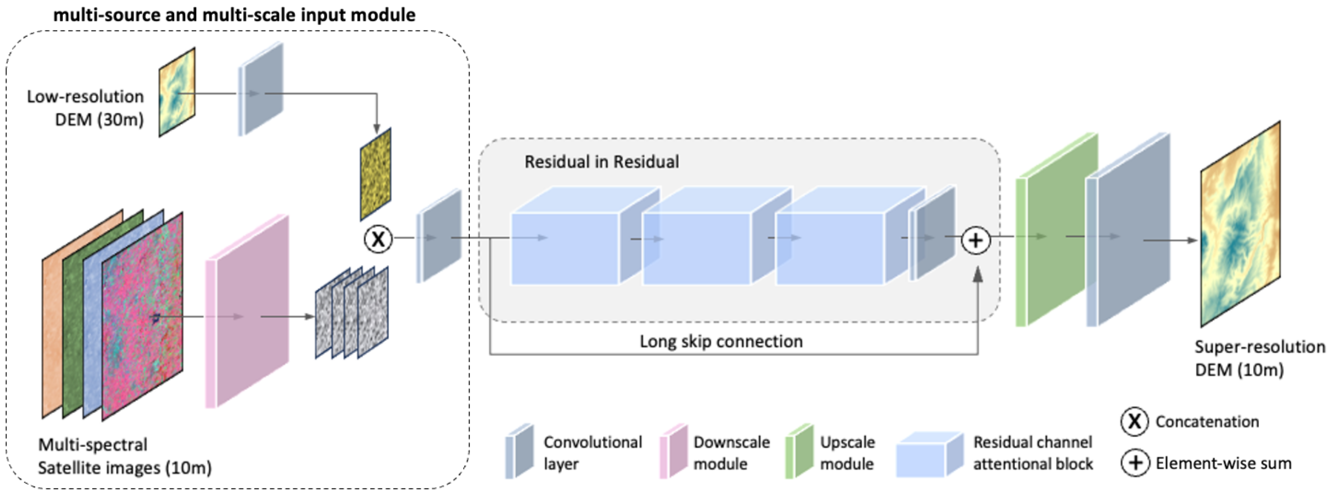

Fig. 1 The structure of the proposed DEM super-resolution model, MS-RCAN. Low-resolution DEM data and four-band multispectral satellite images are fused using a tailored multi-source and multi-scale input module to facilitate the reconstruction of high-resolution DEM data.

Comment: need to expand this section so it is more clear what your contribution/improvements are vs. the original RCAN framework.

(line 145) Thank you for the comment on clarifying that the main difference between the proposed method and RCAN is the multi-source and multi-scale input module. This is clarified in the earlier sections as follows: "The proposed method, referred to as RCAN-Multispectral (RCAN-MS), incorporates a tailored multi-source and multi-scale input module, which is the key distinction from the original RCAN."

(Table 1) Thank you for the comments on Table 1, we added citations to the corresponding dataset in the table, the revised version is as reported here below:

Table 1. Information on the DEM data and multispectral satellite images in two datasets for the tests of DEM super-resolution models

		Dataset 1. England	Dataset 2. Shenzhen & Hong Kong
10m DEM	Collection source	LIDAR Composite DTM 2019, published by UK Environment Agency (2023)	TanDEM-X, provided by German Aerospace Centre (DLR))
	Spatial resolution	Resampled from 2m to 10m resolution using a bilinear interpolation	Resampled from 12m to 10m resolution using a bilinear interpolation
	Acquisition date	2019-09-01	2016-01-13
30m DEM	Collection source	Shuttle Radar Topography Mission (SRTM), accessed from USGS EarthExplorer	Shuttle Radar Topography Mission (SRTM) , accessed from USGS EarthExplorer
	Spatial resolution	1 arc-second (~ 30m) resolution	1 arc-second (~ 30m) resolution
	Acquisition date	2014-09-23	2014-09-23
10m Multispectral Images	Collection source	Sentinel-2A	Sentinel-2A
	Spatial resolution	10m resolution	10m resolution
	Bands	Band 2 – Blue, Band 3 – Green, Band 4 – Red, Band 8 - Near-infrared	Band 2 – Blue, Band 3 – Green, Band 4 – Red, Band 8 - Near-infrared
	Acquisition date	2022-11-25 / 2023-01-21/ 2023-02-13	2023-12-25

NASA JPL (2013). NASA Shuttle Radar Topography Mission Global 1 arc second [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Retrieved from https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003

UK Environment Agency. (2023). LIDAR Composite DTM 2019 – 10m. Retrieved from https://www.data.gov.uk/dataset/8311f42d-bddd-4cd4-98a3-e543de5be4cb/lidar-composite-dtm-2019-10m

(line 160) As suggested, we will add more information on the data sources as follows:

"All the data in these two datasets are collected from publicly open sources, including SRTM, TanDEM-X, and Sentinel-2, which have been widely adopted for remote sensing applications in urban environments (Wu, et al., 2019; Geiß et al., 2015; C. Li, et al., 2021). SRTM utilized dual radar antennas to collect interferometric radar data, which was then processed into digital topographic data with a resolution of 1 arc-second (Farr et al., 2007). TanDEM-X mission uses a single-pass interferometric synthetic aperture radar (InSAR) system to produce 12 m resolution global digital surface models. The Sentinel-2 satellites carry the Multi-Spectral Instrument (MSI), which captures imagery in 13 spectral bands, with the blue, green, red, and near-infrared bands having a 10m spatial resolution (Spoto et al., 2012)."

Thank you for the suggestions for improving Fig.2. We changed the colours of DEM maps to ensure a colour-blind-friendly appearance, and added north arrows, using the same colour scale for similar maps. We did not add a legend for multi-spectral images as they are presented in false colour mode with the composition of three bands. Additionally, we added a world map and marked the location of the two datasets. The modified figure is as follows:

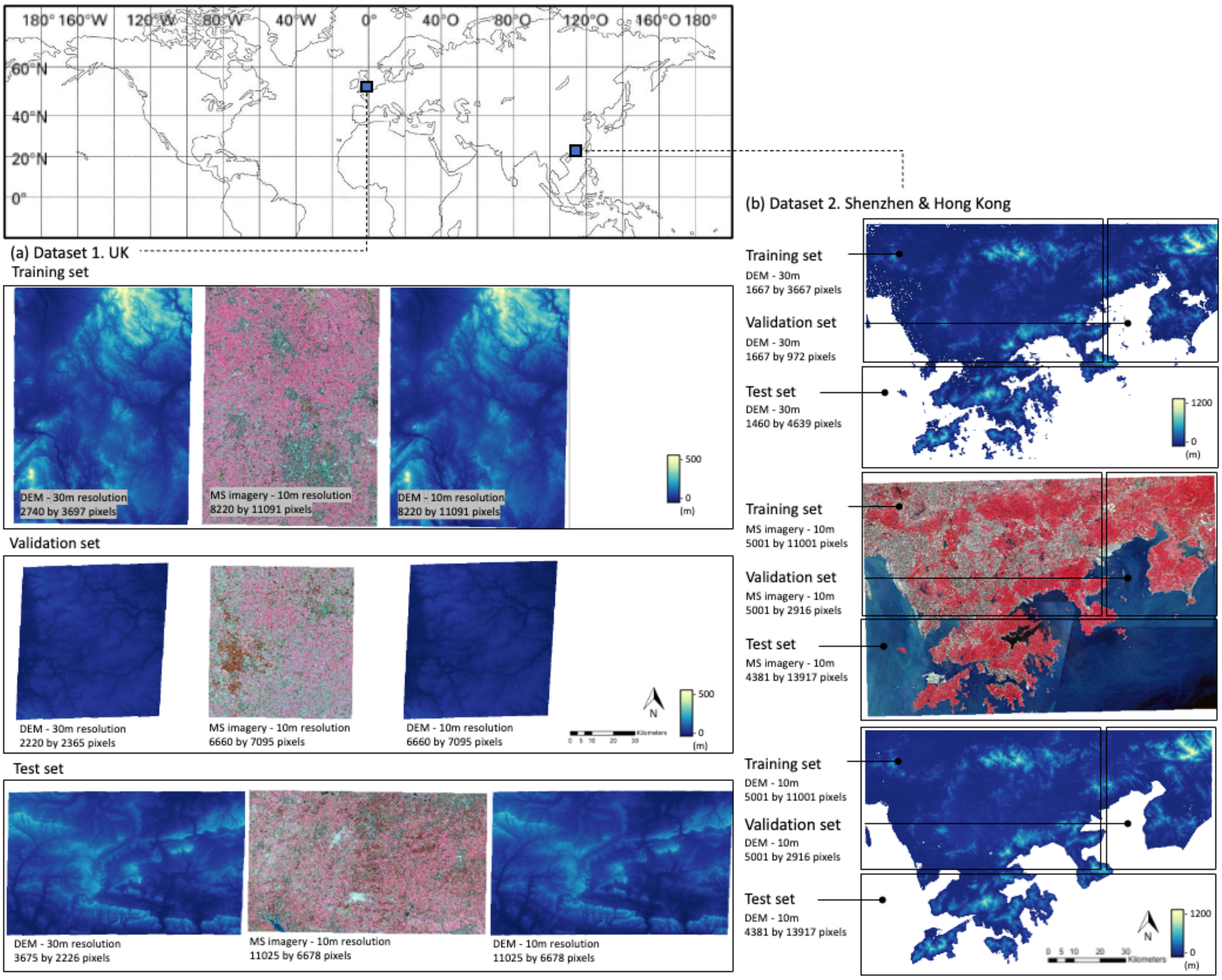

Fig. 2 Overview of the two datasets for DEM Super-resolution. (a) the training, validation, and test sets of Dataset 1. (b) the training, validation, and test sets of Dataset 2 (see Table 1 for data source).

***Comment:* what did you do with the edge/leftover pixels?**

(line 170) We subsample patches in the areas with valid pixel values, therefore no leftover pixels were involved in the training process.

Comment: missing a lot of description that are needed to make your method reproduceable (e.g., hyperparameters, data augmentation, framework (pytorch?), architecture/layer sizes)

(line 175) we have responded in the above main comment section, in which we revised the manuscript to provide details regarding training settings and hyperparameters.

Regarding ***the comment on the necessity of the pluvial flood simulation section****, as addressed in the main comment section, a key contribution of this study is to quantify the impact of the super-resolution DEM, generated using the proposed deep-learning method, on hazard simulations. This quantification offers two main benefits: (i) Since the evaluation metric of super-resolution DEM may not necessarily reflect their effectiveness in geospatial applications, evaluating how DEMs generated through different methods perform in end applications, particularly in flood hazard modelling, can provide more comprehensive performance evaluation. (ii) For practitioners in the field of hazard applications, this section offers insight into whether the proposed deep-learning approach provides a cost-efficient solution for enhancing the accuracy and applicability of flood simulations. We added the above arguments in the manuscript as follows:*

(line 176) "The first stage was centred on assessing the performance of DEM Super-resolution methods in enhancing the resolution of the original DEM data, whereas the second stage was to quantify the effects of adopting the super-resolution DEM on enhancing pluvial flood simulations. This quantification offers two main benefits: (i) provides a more comprehensive performance evaluation on how DEMs generated through different methods perform in end applications; (ii) examines whether the proposed deep-learning approach provides a cost-efficient solution for improving flood simulations."

For the ***comments in section 3.2 Experiment setup****, we provide the answers as follows (line 185):*

- *Regarding the training setup of baseline models, we adopted the same training setup as the proposed method;*
- *The models in the two case studies were trained separately, but using the consistent training setup described in the manuscript;*
- *Regarding the selection of batch size and learning rate, we conducted experiments with various configurations of batch sizes and learning rates. The tests indicated that the chosen configuration achieved the best performance. Additionally, we employed an adaptive learning rate scheduler, which reduced the learning rate by a factor of 0.8 when the validation loss did not decrease for 50 epochs.*

Comment: I'd like to see a plot of MAE vs. epoch for all relevant models on both datasets. I think this will make the methods clearer.

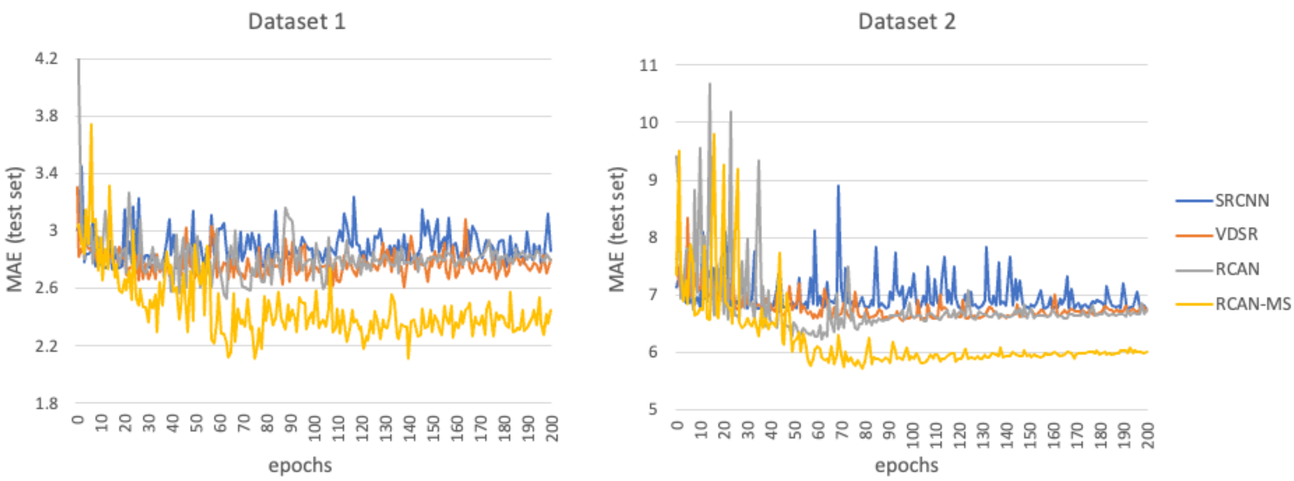

Fig. 3 Changes in the MAE values of all the tested models as training epochs increase for Dataset 1 (left) and Dataset 2 (right).

As requested, we can add a figure to present MAE vs. epoch for all relevant models on both datasets as below:

Comment on section 4.1 DEM Super-resolution: there are too many acronyms and numbers here for me to follow... unless you are pointing out something special.. just leave the info in the table. Or talk about 'percent change' if you want to be more quantitative.

We will revise this paragraph by emphasising just the most important numerical values and adding percentage changes. The revised paragraph is as follows:

(line 215) "For Dataset 1, the RCAN-MS method demonstrates a marked improvement over the Bicubic method, reducing the MAE from 3.0 to 2.2 (-26.7%), and the MSE from 19.0 to 8.7 (-54.2%). This enhancement is also reflected in the values of PSNR (+9.9%) and SSIM (+34.8%), suggesting a substantially improved fit to the target high-resolution DEM. Similarly, Dataset 2 results reveal that RCAN-MS significantly outperforms the bicubic interpolation method, with the MAE sharply decreasing from 9.9 to 5.9 (-40.4%), and the MSE from 186.0 to 67.6 (-63.7%). The RCAN method, serving as the backbone method for RCAN-MS, shows better results than the other deep learning-based methods such as SRCNN and VDSR across both datasets, underscoring the superior performance of the RCAN-based architecture in the task of DEM super-solution. Specifically, for Dataset 1, RCAN posts an MAE of 2.60 and an MSE of 12.9, which are better than those for SRCNN and VDSR. In Dataset 2, RCAN achieves an MAE of 6.4 and an MSE of 83.5, further confirming its robustness. The performance superiority of the proposed RCAN-MS method is evident across all metrics in both datasets, demonstrating its enhanced capability in generating high-fidelity super-resolution DEM data. This is exemplified by the significant reductions in MAE and MSE and the corresponding increase in PNSR and SSIM values, signifying its substantial improvements over the baseline methods."

Comment on Table 2: show units

Thank you for this comment, we will add units for MAE and MSE in the table. However, PNSR is unitless ratio, and SSIM a similarity measure ranging from 0 to 1. The revised table can be as below:

Table 2. Evaluation results of all the tested methods on two test sets with different geographical locations

	Test set of Dataset 1.				*Test set of Dataset 2.*			
	MAE (m)	**MSE (m^2)**	**PNSR**	**SSIM**	**MAE (m)**	**MSE (m^2)**	**PNSR**	**SSIM**
bicubic	3.0078	19.0206	33.4055	0.4621	9.2924	163.0170	35.4505	0.6091
SRCNN	2.7665	15.5027	34.2901	0.5776	6.8153	94.1950	37.8500	0.6794
VDSR	2.6530	13.4866	34.8653	0.5737	6.6412	88.7638	38.1110	0.6811
RCAN	2.5967	12.9453	35.0460	0.5975	6.4150	83.5288	38.3950	0.6838
RCAN-MS	**2.1952**	**8.7102**	**36.7605**	**0.6205**	**5.8181**	**66.6251**	**39.3543**	**0.7411**

Comment regarding experimental results represented in Fig 5 and Fig 6:

- **use the same color pallete as your previous figure**

We would like to argue that, since the previous figure, Figure 2, adopts a color-blind friendly palette, the contrast of values difference is less visually detectable. Unlike Figure 2, which presents an overview of the two datasets, Figures 5 and 6 compare the super-resolution DEMs generated by different models, which contain subtle differences. To make the differences more distinctive, a color palette with multi-hued transitions is more effective. Therefore, we use the 'terrain' color palette in Matplotlib for these two figures.

- **include the SRTM source image (I assume this will be identical to 'bicubic'.. so just ammend the axis title). and an optical image**

We will amend Figures 5 and 6 to include SRTM source images.

- **this figure (Figure 5) has a lot of redundant information as**

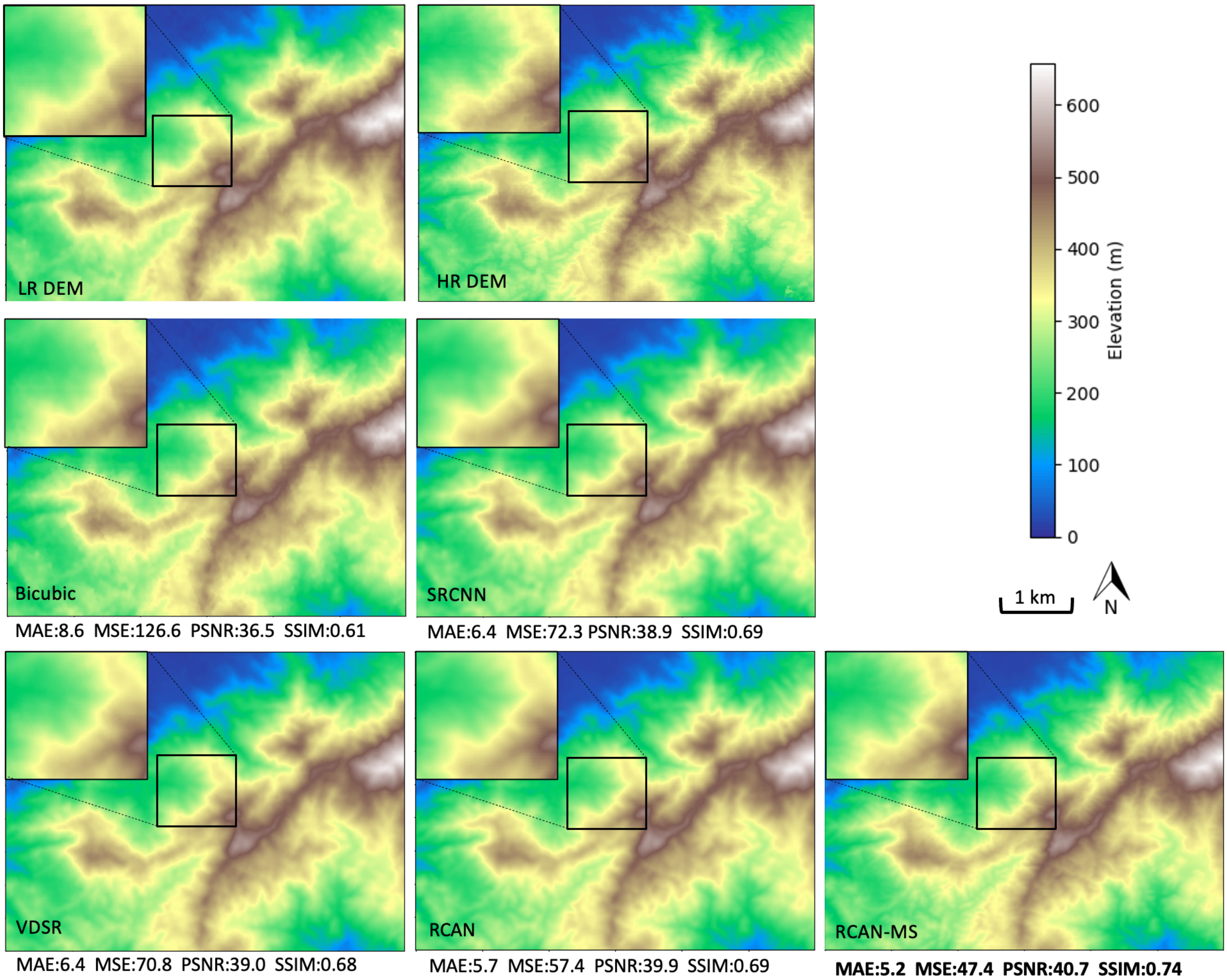

Fig. 4 Comparison of DEM maps in the test set of Dataset 1. generated by the proposed method, RCAN-MS, other baseline methods and the original high-resolution DEM map in Dataset 2.

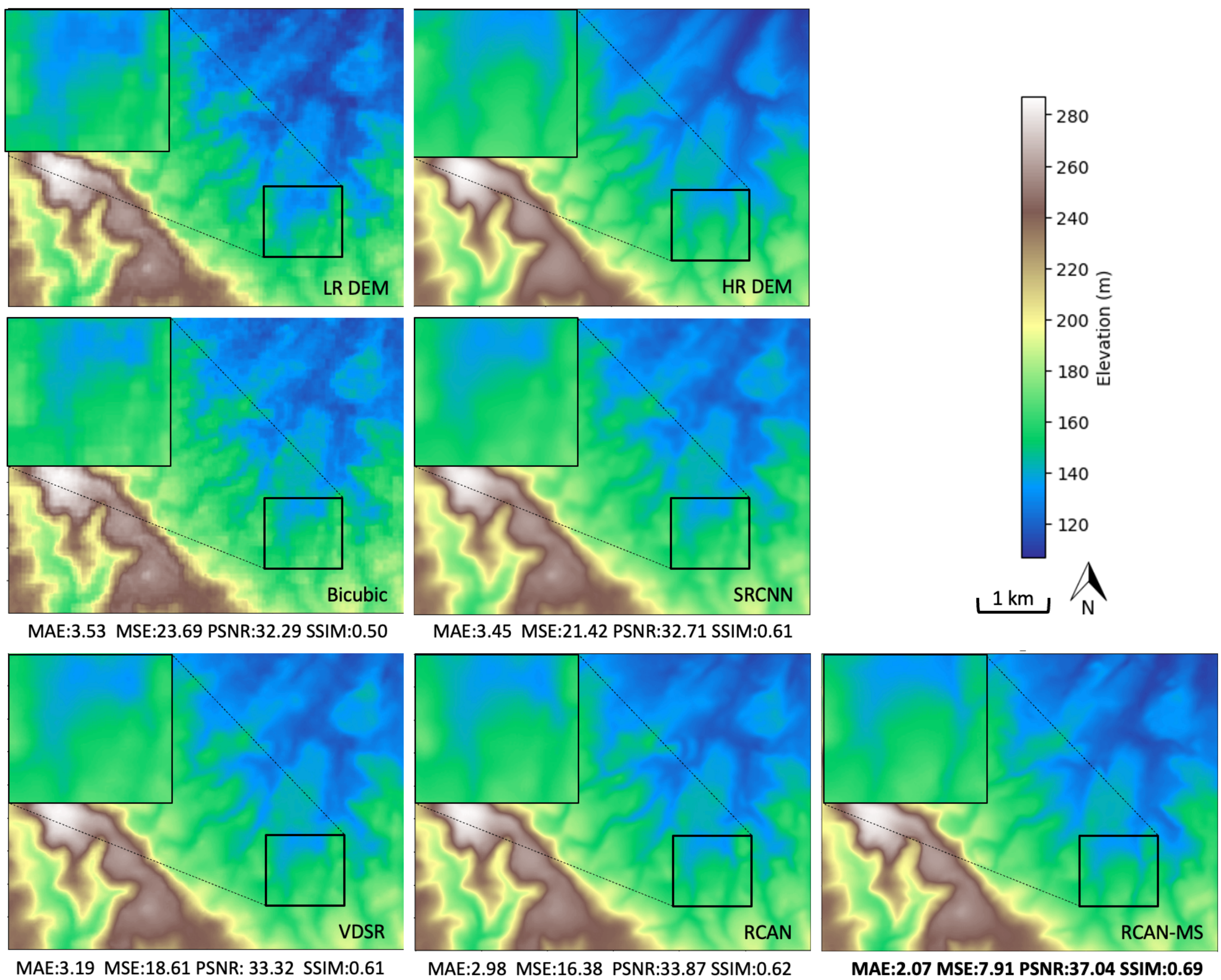

Fig. 5 Comparison of DEM maps in the test set of Dataset 1. generated by the proposed method, RCAN-MS, other baseline methods and the original high-resolution DEM map in Dataset 1.

Thank you for pointing this out. We removed the scale bar on each subplot, and only kept one for all of them on the right side of the figure.

- **for the zoom/call-out, use a different colorscale so we can see what is going on**

Thank you for this suggestion. The primary objective of the zoom-in patch is to enlarge details within the tested area while maintaining consistency with the multi-hued color palette used for the main plots across both datasets. Our current approach ensures that variations remain detectable without introducing potential misinterpretations from an altered color scale. While applying a different color scale to the zoom-in plots might add additional contrast, it could also disrupt consistency and comparability. Given these considerations, we believe our current visualization effectively conveys the necessary details.

- **oh.. I see you are reporting the metrics separately for the patch.. is this needed?**

Table 2 provides the overall performance metrics for the entire test set, which may not capture differences in specific subareas. Therefore, to facilitate a meaningful comparison between these patches, we also report their corresponding metrics. This approach ensures that the patches are quantitatively comparable.

- **I think this whole paragarph could be replaced with "test patch performs similarly"... unless there is some performance difference we should be aware of.... this also seems like it can be removed (paragraph starting with ”The enlarged area of Dataset II….”)**

Thank you for this suggestion, we will remove some numerical descriptions that present consistent performance with the overall evaluation on the test set, whereas we preferred to preserve the description related to the pluvial flood simulation. The revised paragraph is as below:

(line 240) "Fig. 4 and ***Error! Reference source not found.*** *present the two selected patches from the test sets of Datasets 1 and 2 for visual assessment of the performance of the super-resolution DEM maps, in which a subarea of exemplary patches is additionally enlarged for further visual comparison of details. The corresponding reference low-resolution DEM and high-resolution DEM map are also presented for comparison. The values of the evaluation metrics (i.e., MAE, MSR, PSNR, SSIM) are also summarised in the two figures. The ranking of the performance of all the tested methods is aligned with the overall evaluation of the test sets reported in* ***Error! Reference source not found.****, suggesting t hat RCAN offers a larger magnitude of enhancement than SRCNN and VDSR, and RCAN-MS stands out among all the tested methods, recording the lowest MAE and MSE values. These two exemplary patches of the test sets are employed for pluvial flood simulation in the following section.*

*The enlarged area of Dataset 2 is situated at a relatively higher elevation in the patch (****Error! R eference source not found.****). Despite the different geographical locations of the exemplary patches in the two datasets, the results of the DEM super-resolution test on Dataset 2 align with the results of Dataset 1. RCAN gained the second-best performance, and the proposed RCAN-MS also presents in the case of Dataset 2 the best performance among the models tested, highlighting its effectiveness in reconstructing fine-grained information and also capturing the complexity of terrain elevations."*

- **I don't think you're 1 patch comparison is enough evidence to claim that your model performs better in flat terrain. include some discussion of this limitation here.**

Thank you for pointing this out. We would like to clarify that the proposed model not only achieved the best performance on this single patch but also demonstrated superior overall performance on Dataset I, which generally features relatively flatter terrain compared to Dataset II. However, we agree that this does not definitively confirm its generalizability in flat terrains. Therefore, we will tone down the statement and added a discussion of potential limitations.

(line 255) "In contrast to the exemplary patch from Dataset 2 (Fig. 6), the patch from Dataset 1 is characterized by a relatively flatter terrain (Fig. 5). Arguably, flatter areas could pose a greater challenge due to smaller variations in elevation, which are closer in magnitude to the vertical accuracy of the DEM, potentially increasing the likelihood of error. Given the superior overall performance of RCAN-MS in Dataset 1, this suggests its potential effectiveness in handling subtler elevation changes. However, the datasets only represent terrains from two geographical regions, which do not encompass the full diversity of terrain characteristics."

- **Please add an evaluation of cross-validating the two datasets (i.e., use the Hong Kong model weights to make predicitions in England). This will better communicate the methods ability for transferability.**

Thank you for your suggestion. We acknowledge that an extended cross-validation between the two datasets may provide additional insights into model transferability. However, differences in data distribution, sensor resolution, and regional terrain features between the two datasets may introduce confounding factors that require further adaptations. In this sense, applying a model trained on one specific source domain directly to another without any adaptations (e.g., parameter tuning) could *significantly impact performance, particularly given the distinct terrain characteristics of the two regions.*

We understand that extended cross-validation between models trained in different datasets would be a valuable approach for studies prioritising transferability. For instance, a truly generalisable model would require training on diverse datasets that minimize bias rather than relying on a single source domain. However, our primary focus in this study is to assess model performance within different datasets under consistent training conditions, ensuring applicability to various terrain characteristics. The experimental results on both datasets have met these objectives.

Comment on Section 4.2 Pluvial flood simulation: I suggest moving all of this to the supplement

As clarified in previous responses, we consider the pluvial flood simulation section as essential because it quantifies the impact of super-resolution DEMs on hazard modeling, complementing the evaluation metrics of DEM super-resolution and underlying the practical implications in hazard modelling. This section provides indeed practitioners with insights into the cost-effectiveness of the proposed deep-learning approach for improving flood simulation accuracy. Therefore, we prefer to retain this section in the main text.

Comment on Fig 9: need to make it more clear that these are popout boxes; use a different color for zero

Thank you for this comment, we will amend the figures using white color for zero and made them clearer as pop-out boxes. The revised figures are as below:

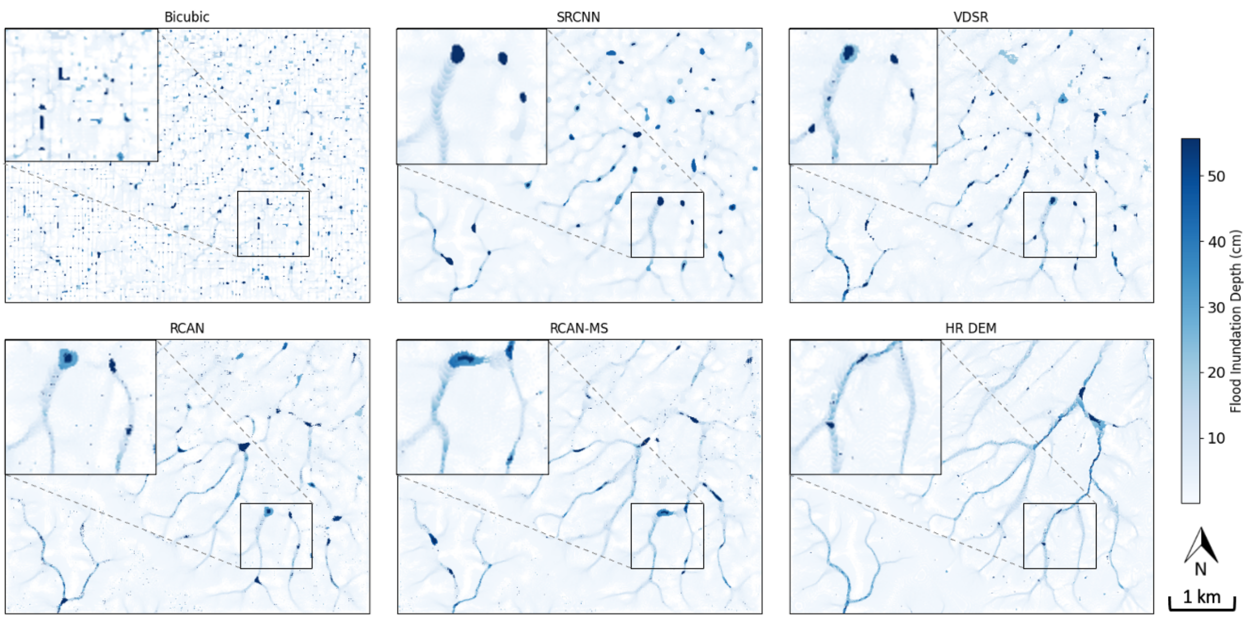

Fig. 6 Maps of pluvial flood inundation depth were simulated using super-resolution DEM data and compared with the original high-resolution DEM data in an exemplary patch of Dataset 1.

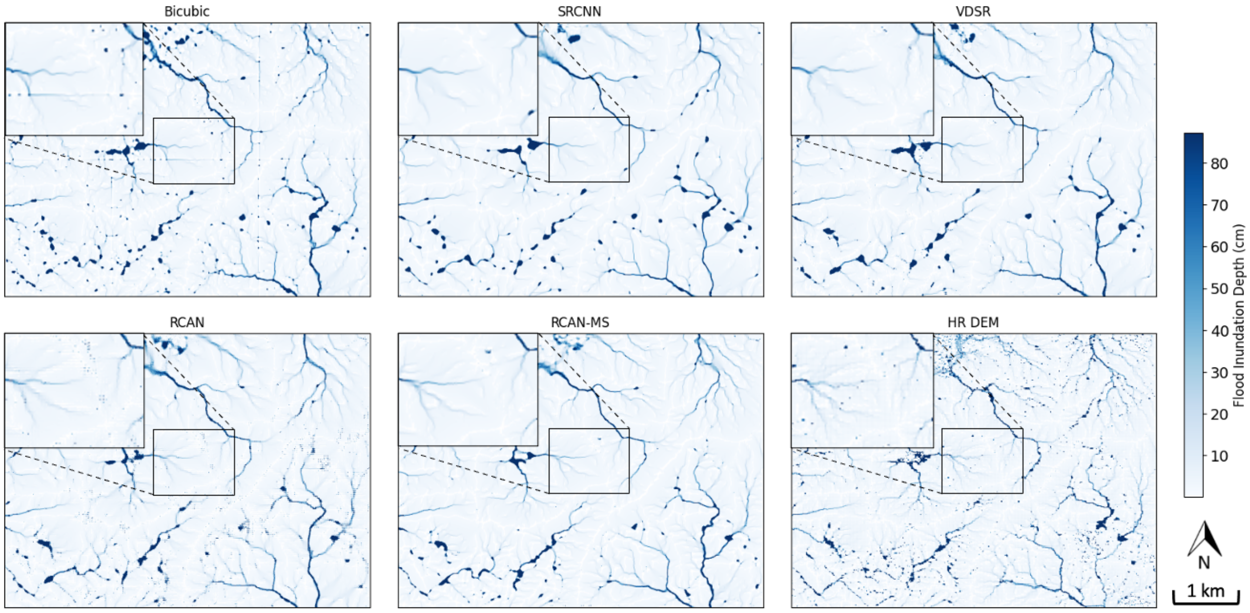

Fig. 7 Maps of pluvial flood inundation depth were simulated using super-resolution DEM data and compared with the original high-resolution DEM data in an exemplary patch of Dataset 2.

Comment on Section 5. Discussion:

- **"pepper-and-salt noise effects" need to talk about this in the introduction**

Thank you for this comment. We would like to clarify that the "pepper-and-salt noise effects" mentioned here refer to issues observed in the super-resolution DEM data generated by RCAN, one of the baseline models. This is not a common issue for this model, nor is it a characteristic of the original data. Therefore, we believe it is more appropriate to discuss this in the discussion section rather than in the introduction. To avoid any misunderstanding, we have revised the sentence for better clarity. The revised version is as follows:

(line 305) "Such superior performance is due to learning from multi-source inputs, particularly the incorporation of high-resolution multispectral satellite images, enabling it to achieve fine-resolution details, while mitigating the pepper-and-salt noises in the super-resolution DEM generated by RCAN but avoiding over-smoothing in SRCNN and VDSR."

- **"in RCAN-MS, the improvement effect of the multi-source input on DEM super-resolution is likely due to the additional feature extracted from multispectral information" didn't you show this? This is why you need to provide more information on your baseline models.**

We wish to clarify that "the additional features" here refer to the multi-spectral satellite images that are employed as part of the input for RCAN-MS, this has been explained in the previous sections. To avoid unambiguity, we will revise this description for better clarity as follows:

(line 306) "in RCAN-MS, the improvement effect of the multi-source input on DEM super-resolution is likely due to the input from multispectral information, which provides additional features to facilitate the estimation of the reflectance of varying land cover types."

- **"in principle better performance in DEM super-resolution does not necessarily guarantee an improvement in flood simulation accuracy" This is an interesting point, and a good**

argument for including some sort of hydrologic metrics in your evaluation (as you did with the pluvial flooding). However, I still think this is a minor point... esp as it only supports the conclusion of your traditional metrics.

Thank you for pointing this out. As clarified in earlier responses to this concern, besides supporting traditional metrics, an important contribution of having the section on pluvial flood simulation is to quantify the extent to which super-resolution DEM generated by the proposed model can improve flood simulation, compared with other baseline models. This not only can highlight the effectiveness of the method in hazard simulation applications, but also provide important reference for practitioners to consider and evaluate the cost-effective of this approach in similar applications.

- **"only incorporated 4-band multispectral satellite images as additional features, other terrain-related features (e.g., slope, aspect) that may bring further improvement to model performance were not tested in this study" this isn't really a limitation.. more about future work**

Thank you for this comment. We will revise this paragraph to merge into the section of discussion on future work. The revised content is as follows:

(line 335) "In future work, further tests could focus on investigating the impact of including additional features on model performance. This study takes advantage of multi-scale and multi-source input data for DEM super-resolution but only incorporates 4-band multispectral satellite images as additional features. Other terrain-related features (e.g., slope, aspect) that may improve model performance were not tested. Thus, future work can explore the impact of terrain-related features on enhancing model performance, as well as examine the performance of the proposed methods with different downscaling factors, where higher-resolution DEM data is available as training targets."

- **"In future work…" this is better placed in the conclusion section.**

We believe this discussion of future studies is better suited for the discussion section rather than the conclusion. As the discussion of future studies here is more connected with the above-described limitations, putting it in the discussion section allows for a critical reflection on the limitations and highlights areas for further exploration. Meanwhile, we tend to have the conclusion section focus on summarizing the key findings and main takeaways. Introducing future work in the conclusion may detract from the focus on the study results.

Comment: Need to have a very good reason to not share your code. Esp. considering all the authors seem to come from publicly funded institutions. I can not complete the review without seeing the code.

As responded in the previous section of the main comments, we agree with openly sharing the code and data, except for the high-resolution DEM data for Dataset 2. This information can be added in the manuscript as follows:

"Except for the data from TanDEM-X, which requires a proposal submission and approval for data acquisition, all other data and codes are openly accessible here: https://zenodo.org/records/14868516"