RC1: 'Comment on nhess-2024-207'

The paper introduces a deep-learning method that combines low resolution DEM and multi-spectral images to obtain a high-resolution DEM that is ultimately used for running a pluvial flood simulation. The authors also compare this approach with other DL and not methods showing its improved efficacy.

The manuscript is well written, clear, concise, and informative. As such I recommend publication with just few minor details that might further improve the quality of the paper.

Thank you very much for your comments, we provide a detailed point-by-point response to each comment as below.

Minor comments:

1. **In the results/discussion section, I would emphasize that the difference between the RCAN and the RCAN-MS is mainly in the inputs used (if I followed everything correctly), thus further proving your point that the extra information coming from multi-spectral images is beneficial, since so far it seemed "just" a difference in method as you have with VDSR, for example.**

Thank you for this suggestion. We agree with this and will emphasise in the manuscript that the primary distinction between RCAN and RCAN-MS lies in the inputs used, as the tailored input layers to processing multi-sourced inputs. This added information supports the advantage of using multi-spectral images and strengthens the argument that they contribute to improved model performance. We can revise the discussion section to clarify this as follows:

(Line 138) "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."

(Line 147) "The tailored multi-source input module is integrated into the model structure before the first layer of the RCAN backbone structure (Fig. 1)."

(Line 400) "The improved performance of RCAN-MS in flood simulation, compared to its backbone method RCAN, underscores the value of incorporating multispectral data. The additional information provided by the multispectral images enhances terrain representation and reduces noise in the super-resolution DEM, thus leads to more accurate flood simulation results."

2. **Could you explain why do the results in terms of flood simulations look more consistent in Dataset II rather than in Dataset I, at least visually? For example, in Figure 6, all interpolation methods seem to produce some sort of accumulation ponds in correspondence of the bifurcations of the rivers and the bicubic approximation results in a noisy pattern. However, that does not seem the case for Figure 8 with Dataset II. Do you have any clue why?**

Thank you for raising this point. A potential explanation for the difference in flood simulation results between the two datasets may stem from the terrain characteristics of the study areas. As shown in Figures 4 and 5, the test area in Dataset 1 is relatively flat, while the second test area in Dataset 2 has a hillier terrain. In Dataset 1, the flatter landscape leads to a more diffuse distribution of floodwater, which can result in less distinct patterns and variability in the simulation results. In contrast, the hilly *terrain of Dataset 2, even with bicubic interpolation, naturally facilitates more concentrated floodwater accumulation in certain areas, resulting in relatively more consistent simulation outcomes across different methods.*

We can include this discussion in the manuscript as follows (line 360):

"In addition, the terrain characteristics can influence the effectiveness of interpolation and super-resolution methods in flood simulation. Specifically, the improvement in flood simulation maps achieved by RCAN-MS is more pronounced in Dataset 1 than in Dataset 2. A key factor contributing to this discrepancy is the difference in terrain between the two datasets. As shown in Fig. 5 and Fig. 6, Dataset 1 features a relatively flat landscape, while Dataset 2 is characterized by hillier topography. In the flatter terrain of Dataset 1, floodwater tends to be more diffusely distributed, resulting in less distinct patterns and greater noise in the simulation results generated by baseline methods (e.g., bicubic interpolation). In contrast, the hilly terrain of Dataset 2 naturally promotes more concentrated water accumulation in specific areas, leading to more visually coherent flood patterns across different methods, even with bicubic interpolation. Therefore, the improvement provided by the proposed super-resolution method tends to be more significant in flatter regions, where its effects are more pronounced."

3. **I think you could also comment further on why is the IoU very low (despite the proportional increase) for high thresholds of water depths.**

Thank you for your question. The low IoU for high water depth thresholds, despite the proportional increase, can likely be attributed to the much smaller extent of deep floodwater areas. At higher thresholds, the areas of flooding become more concentrated in specific regions with much smaller spatial coverage, which may not align well with the predicted flood areas. In this case, at higher depth thresholds, even small misalignments between the predicted and actual flood zones can result in a significant decrease in IoU. While the proportional increase suggests that the model is correctly identifying more flood-prone areas as the water depth threshold rises, the precision and spatial accuracy required to match the predicted and actual flood extents become more challenging.

We can make the corresponding revision in the manuscript as follows:

(line 325) "It can be observed in Fig. 8 and Fig. 10 that, although the proportional increase in IoU indicates that the proposed methods are correctly identifying more flood-prone areas compared to baseline methods, the IoU for high water depth thresholds is much lower than for lower water depth thresholds. This can be attributed to the significantly smaller spatial extent of deep floodwater areas. At higher depth thresholds, even small misalignments between the predicted and actual flood zones can result in a substantial reduction in IoU. While it becomes more challenging to simulate flood extents at higher depth thresholds, flood simulation based on RCAN-MS still achieved the best performance in simulating deep floodwater areas compared to all baseline methods in both datasets."

4. **In terms of metrics you could also consider adding a different metric such as the critical success index (CSI), which has been used in several flood studies.**

Thank you for this suggestion. We incorporated Intersection over Union (IoU) as one of the metrics in our analysis. The formulas for IoU and Critical Success Index (CSI) are mathematically identical in the context of this study. Both metrics measure the overlap between the predicted and actual positive areas (True Positives, TP) relative to the total number of areas covered by both predicted positives (TP + FP) and actual positives (TP + FN), which can be expressed as:

$$CSI = IoU = \frac{TP}{TP + FP + FN}$$

We believe this provides an adequate measure of overlap and performance in our flood simulation results.

5. **While most figures are of high quality, I think Figure 7 and 9 can be better, despite already being informative. Consider changing their style.**

To improve Fig 7 and 9 (which are now Fig 8 and 10 in the revised manuscript), we changed the figure style to bar charts as follows:

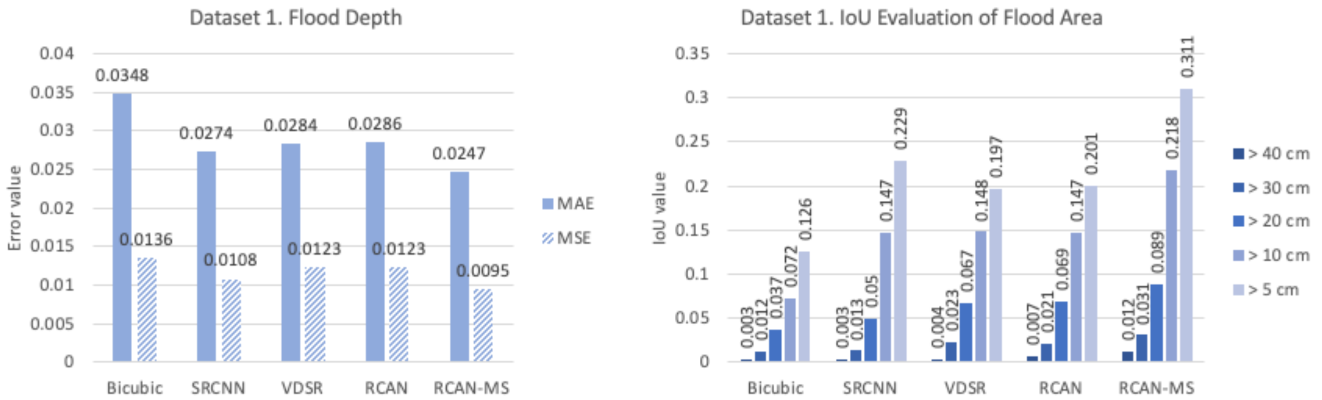

Fig. 8 Performance evaluation of pluvial flood simulations based on super-resolution DEM data compared with the original high-resolution DEM data in the exemplary patch of Dataset I. Left: MAE and MSE comparison of flood depth values; right: IoU evaluation of the spatial coverage of flood area delineated by different depth thresholds from 5cm to 40 cm.

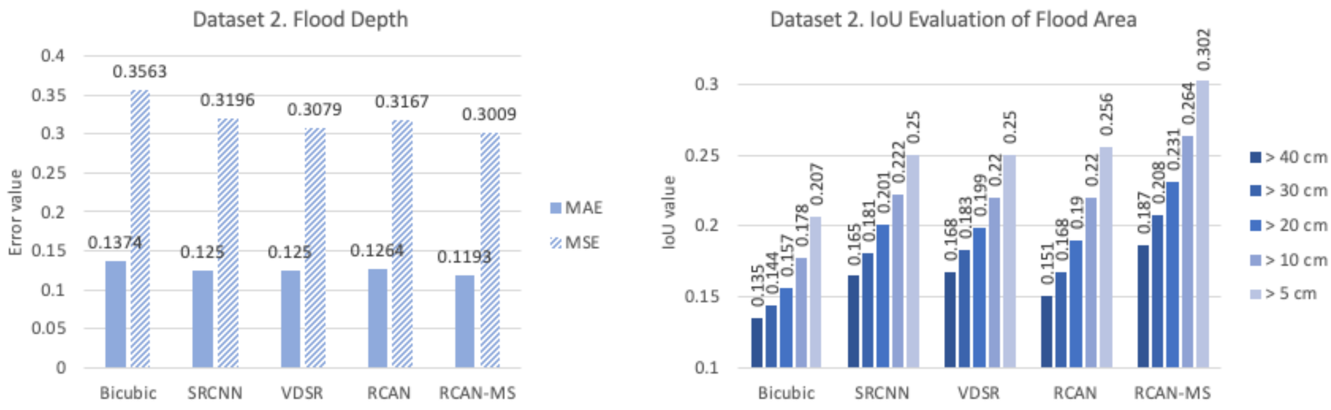

Fig. 10 Performance evaluation of flood simulation maps produced based on super-resolution DEM data compared with the original high-resolution DEM data in the exemplary patch of Dataset II. Left: MAE and MSE comparison of flood depth values; right: IoU evaluation of the spatial coverage of flood area delineated by different depth thresholds from 5cm to 40 cm.