RC3: 'Comment on nhess-2024-207'

This paper presents a new deep learning method to downscale coarse, satellite-derived terrain data to 10m resolution by exploiting higher resolution multispectral image data. The results of the method are validated for two case areas, both through direct comparison against high resolution terrain data, and by comparing pluvial flood simulations with varying terrain inputs. Several benchmark downscaling methods are included in the comparison.

I think this is a very good paper. I very much appreciate the investigation of the effects of downscaling methods on the final application, i.e., pluvial flood simulation, and I think it is well placed within the scope of the journal. I have only some very minor comments that are mentioned below and don't require further review. I suggest accepting the paper.

Comments:

1. Language - please perform a proofread, there are several typos distributed throughout the paper

Thank you for your comments. We performed a careful proofreading and corrected typos in this manuscript.

2. Units - please include units in the results figures e.g. Fig. 7 and 9. Similarly, the scores in Table 2 require units. I believe that the test in Hongkong does not have an average error of 8m, but how should we interpret an MSE of 66???

Thank you for pointing this out. The metrics in Table 2 require proper units. PSNR and SSIM are unitless. MAE is measured in meters (m), MSE is measured in square meters (m^2), and reflects the mean of squared elevation differences, making it more sensitive to outliers. The MSE of 66.6251 aligns with the average MAE of 5.8181 when considering that MSE is more sensitive to outliers. As such, MSE provides a measure of error variability, where large errors have more influence. This highlights the importance of interpreting both MAE and RMSE for a complete understanding of model performance. We can includ these units in the revised table to clarify the interpretations.

Table 1. 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**

3. Figure 1 - please include resolutions in the figure. The entire residual in residual block operates in 30m resolution. In addition, the upscaling module is not described. I suppose this is another 2D convolution. Does it receive a skip connection with high resolution as input?

Thank you for this comment.

In terms of the resolution in Figure 1, we have included resolution information for the low-resolution DEM data, high-resolution multi-spectral satellite images, and the super-resolution DEM output. We argue that, since the data passing through convolutional layers are tensors that may not have explicit physical units, it is more appropriate to represent the height and width of the spatial dimensions as spatial resolution in the figure. Therefore, we can add the spatial resolution (height, width) in this figure, as shown below:

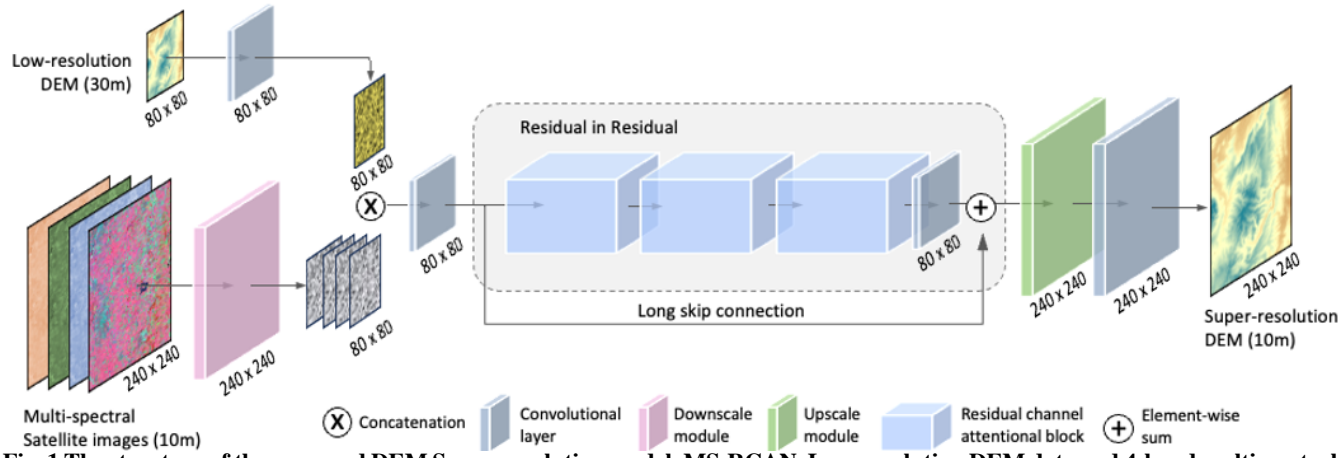

Fig. 1 The structure of the proposed DEM Super-resolution model, MS-RCAN. Low-resolution DEM data and 4-band multispectral satellite images are adopted as the input to reconstruct high-resolution DEM data.

Regarding the upscaling module, we can add a description regarding the upscaling layer as follows:

(line 155) "After that, the concatenated multi-source input is passed through the RCAN backbone structure, which consists of RIR blocks and includes a 2D convolutional layer at the end of the model structure to upscale the data flow to the size of the high-resolution DEM map."

We did not add a skip connection between the input and the high-resolution output, this is because there is a long skip connection at the end of the input module and before the upscaling layer at the end of the model structure, which are just a few layers to reach the final output.