

No	Review A	Response
1	- Urban classification: cross-validation is needed, say 5-fold or 10-fold. - If different training and testing sets are randomly selected, will the results reported be unchanged? - What will happen if a more sophisticated algorithm, Neural Network, is used?	-The reviewer raises an interesting concern. However, because the data used in this research had coarse spatial resolution, we faced with difficulty in classifying mixed pixels. Besides, urban points were chosen based on polygons, one polygon contains one or more points; non-urban points were taken randomly throughout Vietnam's territory by stratified sampling based on GLCNMO v3' classes. Therefore, if we use cross-validation, it will not keep the constraint of stratified random sampling. As a result, we used hold-out validation in our method. - If different training and testing sets are randomly selected, the final result will be unchanged, because we used the stratified random sampling method. - The suggested experiment is interesting, but we feel that it falls outside the scope of this study. As we regarded in paper's conclusion, some sophisticated machine learning methods, such as neural network, support vector machine, and ensemble methods will be used in our future work.
2	Section 3.1: "appropriate thresholds were chosen automatically to separate urban and non-urban points into two distinct parts." : thresholds were given without any clues about how we could find them automatically	We agree with this comment, and this has been addressed in the revised version (in subsection 3.3.1).
3	- Subsections of section 4 must be 4.1, 4.2 instead of 1, 2. -Section 4.1 just "read" all "table" 3,4,5. Explanation is a must but not available yet (why	We agree with this comment, and this has been addressed in the revised version. - As we mentioned in subsection 3.2.3, mean square error, peak signal

	this combination is the best with this dataset, why others are worse...)	to noise ratio, and structural similarity index was used to estimate the effects of resampling methods on spatial data. The combinations was assessed by their MSE, PSNR, and SSIM values. The lower MSE is better and vice versa. The higher PSNR and SSIM are better and vice versa.
4	- Table 6 should be simplified, for example, Worldpop and MOD44W Waterbody could be removed since they are the same for all experiments. - Many values in table 6 are exactly equal, value 97.71 appears 9 times!!! why?	- We had removed Worldpop and MOD44W Waterbody data out of Table 6 . - Because the data used in this research had coarse spatial resolution, and the number of testing points is quite small, therefore, the overall accuracy of some combinations are equal.
	Reviewer B	
5	What the reason to apply only one Aggregation method on Worldpop and MOD44W data? By doing this, you excluded to use them for evaluation.	Because Worldpop data is population count data, therefore, we only used sum aggregation method to resample the data to keep its characteristics unchanged. Similarly, MOD44W data is categorical data, for that reason, we only used majority aggregation method to resample this data. As a result, we had removed Worldpop and MOD44W Waterbody data out of result table (Table 6).
6	In the paper, it isn't clear the choice of the different thresholds. It would be interesting to add more information about these thresholds. In the same way, you should assign a name to the different axis of histograms to help the readers to understand your choice of the thresholds.	We agree with this comment, and this has been addressed in the revised version (in subsection 3.3.1).
7	You state that you modify the proposed classification method in the GLCNMO, but you didn't give any more information about the modifications.	We agree with this comment, and this has been addressed in the revised version (in subsection 3.3.2).

8	In table 2, what means the “Training data accuracy” in the last column?	This is a typing mistake, we have adjusted to the text to be clearer. The “training accuracy” is the accuracy when applying the thresholds on the training data. It shows how good the thresholds are in separating the training data into two distinct parts.
---	---	--