The paper titled "Watershed zonation approach for tractably quantifying above-and-belowground watershed heterogeneity and functions" develops a watershed zonation approach for characterizing watershed organization and function. The authors use multiple high resolution spatial datasets available over the East River watershed to explore this relationship at the hillslope level. Their use of process-driven observations (Annual N export) is an important contribution as it shows that the zonation is indeed connected to the local processes. I recommend that this paper be accepted for publication after major revisions.

Thank you very much for the detailed review and constructive comments. We have revised the manuscript accordingly.

1. The steps used to delineate the watersheds is clear in the current manuscript. However, the step to delineate the hillslopes remains unclear. I am assuming it involves just splitting each watershed into a left/right hillslopes and a "headwaters" hillslope if it exists; however, this needs to be made more explicit. On a related note, the title is misleading since the clustering is performed at the hillslope level. I would suggest that the title be clarified so that it is clear that it is hillslope-level and not simply watershed-level.

Answer: We have expanded the explanation of hillslope delineation in the first paragraph of Section 3.1. We also changed the title to “Watershed zonation through hillslope clustering for tractably quantifying above-and-belowground watershed heterogeneity and functions”

2. "This is consistent with Wood et al. (2011), concluding that the order of 100 m is a sufficient resolution for representing hydrological fluxes" is a stretch. There is a lot more heterogeneity that will matter at scales finer than hillslopes even if it is not captured in your data; furthermore, the role of hillslope hydrology is variable depending on the topographic environment. This sentence is not necessary to show your point; I would remove it.

Answer: We agreed and removed this sentence.

3. Line 180 - The reason why the hillslope-level correlations is better than pixel-to-pixel can be deceptive. This is probably a combination of the processes being more connected at 100 meter scales but also just simply because you are removing random noise from the higher resolution data by aggregating at the hillslope level. I think it would be useful to test how the correlations vary as you upscale the original regular grid maps (e.g., 50 m, 100 m, 250 m...). You would do a pixel-to-pixel comparison for these as well. It would be a strong result if the hillslope approach still wins out. My hypothesis is that it won't and that they will be pretty close. This analysis would be useful within the paper or could be placed in the supplement.

Answer: Thank you for this great suggestion. We indeed confirmed that upscaling the pixels improves the correlations among metrics. We included the pixel-by-pixel correlations at different pixel sizes in the supplementary material (Figure S2).

However, we still think that the hillslopes are still more effective units for capturing the watershed heterogeneity than pixel-based upscaling. In response to Dr. Band's comments, we have computed the across-hillslope variance (i.e., the variance of the hillslope averaged metrics) compared to the overall pixel-by-pixel variance (Figure X1a). In addition, we computed the variance of the upscaled pixel metrics compared to the overall variance (Figure X1b). The ratio of the across-hillslope variance or upscaled-pixel variance decreases as the threshold drainage area or the pixel area increases, since the averaged values do not account for the overall variability fully. However, in the hillslope averaging (Figure X1a), there is a plateau of the area with some metrics like elevation, radiation and peak SWE (which are known to be critical for key watershed processes). This means that there is a certain hillslope size, up to which the variability within each hillslope is limited and the hillslope-averaged metrics can capture the overall variance. Such a plateau does not exist for the case with upscaled pixels, since larger pixels can contain different aspects of hillslopes within. This shows that the hillslope averaging is more effective than averaging within upscaled pixels, which is consistent with Band et al. (1991).

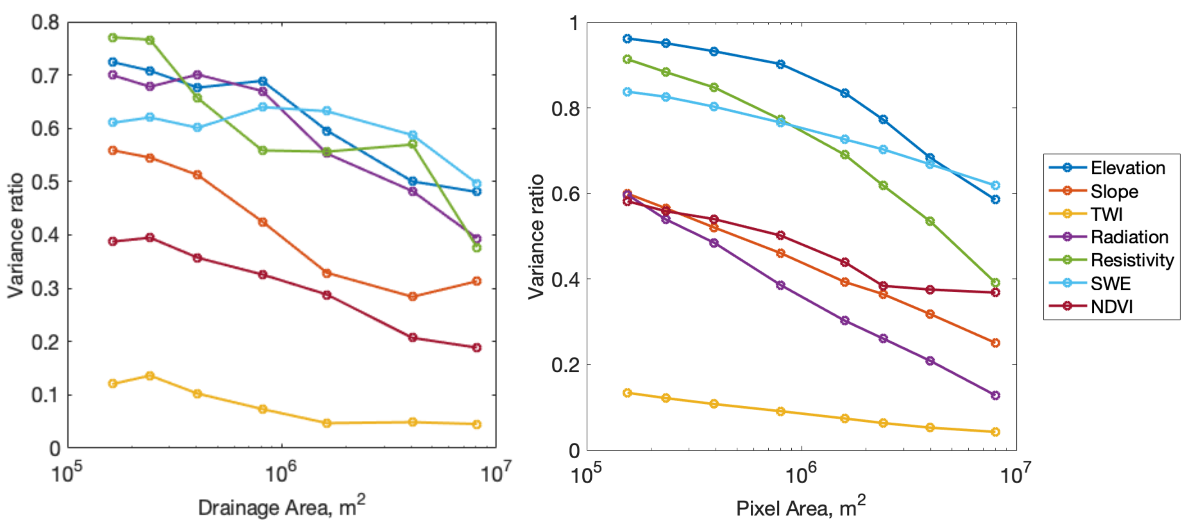

Figure X1 (Figure 2 in the manuscript). Variance ratio (a) between the across-hillslope variance (i.e., the variance of hillslope-averaged metrics across all the hillslopes) and the overall variance (i.e., the variance of the pixel-by-pixel properties) as a function of the threshold drainage area, and (b) between the across-upscaled-pixel variance and the overall variance as a function of the pixel area.

4. The current discussion has a number of disjointed paragraphs and ideas. I would encourage to split it up into discussion and conclusion sections and then to subdivide the discussion section into subsections.

> Answer: We have improved the connectivity between paragraphs in the discussion section, and created a conclusion section in the revised manuscript.

5. Figure 1 - The legend of NLCD is not comprehensive. There are developed/urban areas on the map but they are not referenced in the legend.

> Answer: Thank you for pointing this out. We have revised the legend in Figure 1b.

6. Figure 3 - The 10 on the colorbar is cut off

> Answer: Thank you for pointing this out. We have changed Figure 4 (formerly Figure 3)

Citation: https://doi.org/10.5194/hess-2021-228-RC2