4 February, 2022

This is a review of "On the similarity of hillslope hydrologic function: a process-based approach" by Maina and colleagues submitted to the Hydrology & Earth System Science. This manuscript details hydrologic classification at the regional catchment scale focused on groundwater parameters gleaned from modeling, but also including a full range of related hydrologic and geomorphologic characteristics. The ability to classify hydrologic function across the landscape is important for large-scale hydrologic modeling, and in turn, for predictions of future water resources. To my knowledge, this work is novel particularly for the focus on including groundwater dynamics in the classification effort. This manuscript is within the scope of the journal and will be of value to the readership. Overall the writing is clear, and the figures are well drafted. My general and specific comments are provided below.

General Comments:

The manuscript may benefit from some improvement in organization. For example, the core purpose of the manuscript appears to focus on classification, however classification methods are not described or justified in the Methods section.

As far as I can tell, groundwater dynamics are only modeled and direct measurements are not used to calibrate or validate the model results. There appear to be some groundwater data in ESS-DIVE – would it be possible to use these to support the modeling results?

> *Williams K ; Carroll R ; Dong W ; Versteeg R ; Tokunaga T (2020): Water Level Data from Wells PLM1 and PLM6 for the East River Watershed, Colorado. Watershed Function SFA, ESS-DIVE repository. Dataset. doi:10.15485/1818367*
>
> *Dafflon B ; Dwivedi D (2020): Groundwater level elevation and temperature at the Lower Montane in the East River Watershed, Colorado. Watershed Function SFA, ESS-DIVE repository. Dataset. doi:10.15485/1647040*

I had difficulty interpreting the 'distribution plots' Figure 6-9. Perhaps it is because I am unaccustomed to interpreting this type of plot, or perhaps it is because the x-scale is squeezed to fit all of the lines on, but it could be more clear what type of patterns the reader should be looking for and how to draw specific conclusions from these graphs. Perhaps considering an alternative way to present this data would be useful, or maybe include an explicit explanation for how to interpret them for the reader.

The authors may wish to consider splitting section 3 up into separate 'results' and 'discussion' sections to improve organization and help guide the reader to where data are presented vs where they are contextualized. In general, the discussion content of this section builds primarily examines patterns within the results of this study and builds only limited links to past work – more complete referencing in the discussion may help to improve contextualization of this research within the broader body of scientific work.

The 'summary and conclusions' section could be condensed by removing the summary and focusing on the core conclusions of this research.

Line Comments:

Line # | "quote from manuscript" , Comment   .

69-71: Perhaps provide a reference as evidence that would call this assumption into question?

85:     "300 mm" Does this mean 'within a single hillslope'? Where does this number come from?

85-86: Where do the order-of-magnitude numbers come from?

98-100: "In this study...functional zonation." As I understand this, the authors are defining "functional zonation" as the seasonal change in groundwater levels. Later, on line 110, the authors appear to state that 'hydrologic function' is an equivalent term of 'functional zonation.' (my apologies if I have misunderstood this). If this is the case, why not just stick with the term as originally defined and be consistent, rather than introduce a synonym that may add confusion? Additionally, it is somewhat unclear how the 'integrated hydrodynamic response' can be effectively captured by simply the seasonal changes in groundwater level – this would seem to ignore any unsaturated zone dynamics that do not directly affect the water table such as storage, partitioning, plant water use, etc. I recognize the argument present on line 89 that groundwater is linked with unsaturated zone processes, however Maxwell and Condon found this on a continental scale using 1km model cell resolution, and it is not clear that the same relationship is robust at the hillslope scale. I am not suggesting that groundwater level is unimportant, but rather that it's unclear if it is truly

appropriate as a proxy that 'integrates' the whole hydrologic "story" of a hillslope or watershed.

Figure 2: The y-axis tickmarks could be improved: WTD only 1m and 3m are labeled, and tick marks of apparently .3333 m are provided, which is a bit awkward as an uneven number. on the SWE exist, the 200mm and 800mm labels don't appear to line up with any tick marks.

242-246: I am unclear: is the WTD plotted in Fig 2 a measured value or a model output? from this sentence it appears to be a model output, but it is unclear.

248:    "beginning of snowmelt (i.e., May)": based on Figure 2, it appears that snowmelt might begin in April or perhaps earlier?

249:    perhaps this is nuanced, but the timing of events on Figure 2 is slightly different than noted in the text: apparently, the groundwater level begins to rise somewhat before the SWE begins to decrease substantially. Maybe the issue is just that the text is describing snowmelt, while the graphic is illustrating SWE (i.e., there may be substantial snowmelt occurring before SWE begins to decrease.

249:    "peak discharge is mostly observed in June and July when the snow melts" Again, my apologies for being pedantic, but based on how I read the graph, it appears that SWE goes to zero by around May 13th, so somewhat before peak discharge.

253:    deltaP1 & P2: are these parameters defined for the first time in this paper, or is there a reference that could be cited with a more precise definition? "This variable indicates the ability of the hillslope to release water" This seems vague: would it also be dependent on inputs, antecedent conditions, etc.? it appears to carry units of "meters" so it's unclear how it quantifies the ability of a hillslope to release water. Similar comment WRT delta P2 "contains information about the storage and the recharge capacity"

264-267: reference needed or more complete explanation?

268:     "Figure 3 shows the classification" Is this really showing anything 'classified' –
         the caption seems to be more accurate "spatial distribution of average values.
         Perhaps I am misunderstanding and the text could just be clarified.

280-281: "These two patterns are different from each other, and they are also different
         from the ones associated with the land surface processes..." It may be helpful to
         provide a brief characterization of how these patterns are different.

283-285: "...complementary information, with areas with high ΔP1 having low WTD
         because the strong changes in groundwater levels, as quantified by ΔP1, lead to
         a deep WTD." Suggest rewording in a more straightforward way to improve
         clarity. Also, it is not immediately clear to me why "strong changes in
         groundwater levels" should result in deep water table. Why could two systems
         not have the same mean value with different standard deviations?

Figure 4: Overall this is a nice figure, but a possible suggestion to improve readability
         would be to color-code the bottom left boxes with a red-white-blue color scale
         scaled to the strength and sign of the correlation to make it easier to digest at a
         glance. Also, some type of lettering/numbering scheme may make it easier to
         draw the reader to the correct part of the figure when discussing the figure in
         the text.

297-298: "...the two variables provide the same information." This is inconsistent with
         earlier in the text where ΔP1 and ΔP2 are defined (253-259) as containing
         different information.

298-299: "TWI, AI, SWE, WTD, and ΔP1 are significantly correlated with elevation."
         What is the threshold for significance? The correlation between Elevation and
         % Bare is **not** included on this list (correlation coefficient = 0.8), while the
         correlation between Elevation and TWI is included (correlation coefficient = -
         0.76).

302-304: "A high correlation between the percent of forests and the elevation is found in
         the mid-elevation whereas grassland shows a high correlation in low and high
         elevations" I am unsure how to read the figure to interpret different correlations
         at different elevation ranges, as suggested by this text.

304-305: The sentence feels repetitive, suggest rewording.

305-306: "$\Delta P1$ is, in general, well correlated to all these variables" Does this suggest that a correlation of -0.24 and -0.35 indicates well-correlated variables? Suggest stating the metrics used for deciding if correlation is strong or not.

306-307: "...the selected variable contains valuable information about these variables." Suggest rewording to improve clarity.

312-314: "Regions with shallow WTDs have the highest ... changes in WTD do not impact ET." I am having trouble discerning the indicated relationships from Figure 4. Specifically, the exponential behavior and threshold are not clearly visible.

325-316: "classifications...zones" I am slightly struggling with how 'classification' and 'zones' are being used here. What are identified as "zones" appear to me – as a reader – to be classes assigned to the underlying polygons. Perhaps it would be helpful to more explicitly define these terms.

317-318: "...grouping was made based on the manual selection of natural grouping in the

318   "probability density function." This is unclear. Perhaps the method could be elaborated on in the Methods section? The explanations between 318-366 are helpful, but they generally come across as arbitrary: for example, why are elevation cut-offs at 3000 and 3500 m used? Should this be based on some statistical property of the dataset? This applies to all categories except "clustering."

316-366: Seems like this could be in the "methods" section.

378-379: "...zones with the least variability..." It's unclear why this is 'an important metric that provides a degree of performance for the method's ability to delineate zones.' Also, this statement could benefit from a reference to support it.

401-402: " the essence of that classification" Unclear what this means.

402:     "excellent index for identifying hillslopes with similar elevation" It's unclear why you would want to do this? why use these indirect observations when elevation is directly available?

408-411: I find it difficult to follow the logic here. Why is it desirable to 'distinguish zones of similar elevation?' How can similar results also indicate that they yielded the same results?

412:     "average percentage" – I'm not sure I follow: the table does not have any numbers expressed as percentages. I think they are possibly reported as fractional values, and just updating the numbers to percentage would make this clearer.

413-420: I am unsure how describing the contents of the table here is helping the reader to take away any particular conclusions. For example: "The selected classifications lead to similar conclusions, hillslopes associated with zone 1 have mainly grasses..." So in many (but not all) of the classifications, there is more than 50% grassland for zone 1... but what does this mean? what does this tell us about the classification or the hydrology? Furthermore, this statement is misleading because two of the classifications have grassland <40% for Zone 1. Similarly: "...zone 2 have mostly identical percentage of forest and grasses..." It's unclear what 'mostly identical' means since the numbers are not equal (i.e., they are not identical). The remainder of the paragraph describing Table 2 is similar – it's unclear how to interpret these results, or what they mean.

429-430: "The classifications based on elevation and AI allows clearly distinguishing the hydroclimate associated with each zone" I am unable to interpret Figure 7 in such a way to understand how the information provided can 'clearly distinguish the hydroclimate.' I see the brief explanation provided in the following lines 430-432, however I still am unable to see how this information or interpretation is represented in the figure.

440-441: "this type of classification mainly describes how a given hillslope release water based on its topographic structure" It is unclear what this means or how it is interpreted from the results presented.

451-452: "A hillslope hydrologic function should aim to describe how a hillslope partitions, stores, retains, and releases water." Great – this is useful, however perhaps it could be presented in the Introduction to set up this concept for the manuscript.

Also, it should be supported with references and specific definitions. What is the key parameter of interest for each of these process functions? timing? volumes? locations? all?

453: "...are simultaneously occurring..." occur simultaneously

475-476: "As a result, the land cover based classification performs well at delineating hillslopes with similar ET rates (Figure 8b)" I'm just not sure how to interpret this from the figure.

478-479: "To some extent, the TWI and elevation classifications poorly distinguish hillslopes with similar ET." Why?

495: "regrouping" Unclear what it means for hillslopes to be 'regrouped' during classification.

496: "Because the TWI approach describes water transfer..." Based on the description of TWI provided on line 344, it is unclear why TWI would describe "water transfer" or what the definition of "water transfer" is.

498-499: "The $\Delta P1$ based classification has one of the lowest averages of CV..." This statement is misleading at best. Looking at the figure, the CV of $\Delta P1$ is 0.12 – there are also two other classifications that achieve the same CV, one classification that achieves a *lower* CV, and the other three CV's equal 0.13, which appears to be only very slightly higher than 0.12. Perhaps this is "one of the lowest" however all CVs are very low and very similar, so it is unclear how 0.12 brings any significance to the argument (or that the low CV is due to the connection between GW and soil saturation as is claimed later in the sentence).

508: "Groundwater storage is mostly quantified in terms of WTD." Support with a reference?

512: "intermediary" Not sure this is the best word.

604-605: "...transcending the uniqueness of place inherent in traditional classifications..." Unclear, suggest rewording with more direct language.

A. Parsekian