Please find our responses below in black font to each review point in red.

Review #1

Briggs *et al.* (HESS) review

First, thank you for the opportunity to review this paper. I really enjoyed reading it. In their paper, Briggs *et al.* collated geophysical surveys, remote sensing data, and stream temperature and discharge loggers to reveal the role of bedrock depth, in catchments underlain by low hydraulic conductance bedrock, on stream dewatering and thermal resilience. The authors have done an excellent job in highlighting the role of fine-scale hydrogeological setting on the aforementioned hydrological processes. Further, the authors revealed that global scale datasets of depth to bedrock (DTB) largely overestimate this critical parameter, by as much as 12 m. Ultimately, this piece is very timely, and adds a nice story to the hydrology puzzle. I especially applaud the authors in identifying some complex processes that really drive home the importance of surface water -groundwater interactions.

I am happy to recommend this paper for publication after what I consider minor to moderate revisions. This dataset is incredibly rich, and whilst I understand it is not possible to do everything, I do think there is some bandwidth for the authors to dig a bit deeper into what is at play in Staunton. I think given the density of these data, a few things may be conceptualized. For instance, wider valleys will have high solar loading, and the recharge will be spread over a larger area. What role might this have on thermal and discharge regimes?

Thanks again, and nice work! Antóin O'Sullivan

Hello Antóin,

We sincerely appreciate your thoughtful and insightful review that was clearly informed by your own substantial research on similar physical watershed topics. The time you took to develop new plots of our data (e.g., valley width vs bedrock depth) is especially appreciated. You are correct in that we did not dive into some of the large-scale stream valley structural controls in detail. That choice was made in part to help keep the reader's focus on the extensive depth to bedrock and stream dewatering data, and in part because Johnson et al (2017) (doi.org/10.1016/j.jhydrol.2020.124929) explored numerous physical valley attributes in conjunction to explain stream temperature/groundwater influence patterns in these same catchments. However, we realize our current understanding of these systems could benefit from some additional analysis regarding physical valley controls on the shallow suprabedrock aquifers across the study catchments, including utilizing available LiDAR data. However, this type of analysis will need to be addressed as part of future work. The Shenandoah watersheds are extensively forested so we do not expect direct solar warming of the land surface to be a major factor in summer. In a revision of this discussion paper we plan to more thoroughly explore the relations between bedrock depth and valley width, valley width variability, and slope, both at local (to HVSR measurements) and average valley scales. We also note that the Johnson et al (2017) conducted in these same catchments found and inverse relation between valley width and their metrics of groundwater influence on stream temperatures, which agrees with the findings of the current study.

Apologies for grammatical errors herein.

Specific comments:

L 134: True, however, could this also be a function of bedrock *K*? For bedrock with a relatively high *K* (karst for instance), a duality may exist where a portion of the water is driven laterally - as stated - whilst another portion may be recharging the bedrock aquifer. These mechanisms are also likely temporally dynamic. In the setting of this study, I agree that lateral flow with bedrock shallowing seems most likely given low *K*. However, in the introduction, it may be best to speak exclusively to the conceptual controls in general.
Great point, as we want this introduction to be broader in scope than the hydrogeological setting of Shenandoah NP. This sentence was revised to include ‘..... and bedrock permeability.’

L 162: Not sure if it is worthwhile stating consolidated sediments, e.g., clay, may dampen the signal too (see Haefner, R.J., Sheets, R.A. and Andrews, R.E., 2010. Evaluation of the horizontal-to-vertical spectral ratio (HVSR) seismic method to determine sediment thickness in the vicinity of the South Well Field, Franklin County, OH.). If one assumes macro-pores in the clay are limited, this may also lead to low storage capacity in areas of shallow depth to low hydraulic conductance layer?
The HVSR technique is sensitive to the shear wave velocity used and in heterogeneous unconsolidated sediments there may be uncertainty associated with the technique. In the study cited, the influence of clay on depth to bedrock measured with HVSR occurs in an environment where 5-meter thick horizontal clay layers are present. In the study, these clay layers represent heterogeneity in the vertical shear wave profile. The authors note that "It is likely that the largest errors in sediment thickness arise from variability of geology (shear-wave velocity) in the subsurface". It is true that in the presence of a relatively impermeable clay lying above bedrock, one could estimate a depth to bedrock that wouldn't represent the depth to no-flow layer. However, we located our HVSR measurements within alluvial sediments and, based on available boring logs, feel confident that clay lenses such as this are not typical. We changed the sentence around L162 to include this new text: "While insensitive to variations in unconsolidated sediment permeability (i.e. identifying relatively impermeable clay layers), the HVSR method is effective at identifying the depth to distinct unconsolidated sediment/bedrock interfaces (Yanamaka et al., 1994)."

L 187: this is an awesome study site overview. nice job! Thank you
L232: maybe spell out the acronym here as it is the first time it appears in a caption. done
L244: nice study sites! Thank you
L253: this progression seems logical As this was a multiyear study, we were able to adapt the study design year to year as we learned more about the geologic system.
L270: typo “bedrock”. This aligns to my prior comment on the assumption the signal to being changed by bedrock. Hardpan or clay may also do this? This does not distract from the study, just worth a nod to the potential limitations of passive stratigraphy mapping
The HVSR method assumes a single shear velocity is representative of the depth profile; significant heterogeneity in subsurface material can lead to uncertainty in the depth estimates. In general, the impedance contrast between sandy soil and clay is not typically sufficient to lead to a resonance peak (even in the case of 5m thick horizontal clay layers as in Haefner, Sheets, and Andrews (2010)), so we believe we are justified in interpreting the resonance peaks to the bedrock interface, particularly as lenses of fines have not been documented for the Blue Ridge Mountain watershed systems previously. Additionally, our empirical and direct (active seismic) measures of shear wave velocity have been quite consistent across the NP.

L 271: and here is the answer, awesome :) Could there be an *n* shown for number of these boreholes? Looks to be *n*=6 from Goodling et al. report?
Correct- ‘six boreholes’ was inserted into this line, thank you.
L 290: typo ‘teams’ rectified
L301: nice! Yes, and it was great exercise.
L329: are these gauges co-located with temperature sensors as displayed in Figure 5? The gages are located a bit further downstream,
Figure 4 – this is a serious dataset, folks. Nice study design! Thank you
L352: smallest seems like an odd word for describing DTB, maybe lowest? This is not a game-changer, just a style thing. Yes there does not seem to be the perfect adjective here... but ‘largest’ seems to work for larger bedrock depths so ‘smallest’ seems the appropriate match for that term. ‘Lowest’ might be potentially interpreted as ‘deepest’.
Table 1 - a quick regression of valley width to median bedrock depth illustrates a power law relationship (R2 ~0.62). Given how poor the broad scale depth to bedrock maps were at predicting bedrock depth, it may be useful to ill
ustrate here that using high res LiDAR and valley morphology as controls on bedrock rock may provide a more realistic view of bedrock depth in these areas. See Figure 1 below. Something for appendix maybe, but I think you have shown in a remarkably clear way that we need better geophysical data. Awesome stuff.
Thank you for explicitly pointing out the relation between average bedrock depth and valley width; in a revision we would add such a plot to our appendix material and discuss that finding in the main body text

Figure 5 – nice Thank you
L386 – might rephrase this sentence for clarity. done
Figure 7 - this is a powerful figure. It brings a lot of questions to mind. I wonder how this would look if one added another 2 panels that plotted the same dewatering observations and sub DTB with valley width? The reason I suggest this, the valley is 3D, by accounting for this 3D space, and given the authors have this amazing data set, it may point towards a more robust understanding of x,y,z space on these hydro processes. See Figures 2 and 3 below.
In a revision we would explore local variation in valley width along these two focus HVSR study reaches. One complicating factor are tributary confluences, which can substantially increase valley width at the ~100 m down valley scale. Both of these focus study reaches include tributary confluences.
L 435 – nice Thank you
L482 – typo ‘HVSR’ rectified
L483 - this is an excellent finding. I think even more important given the findings of bedrock depth controls in this study. Yes, given the average bedrock depth across all study watersheds was 3.4 m or smaller, this offset from the global scale dataset is stunning. Reviewer #3 correctly points out that we might not expect the global dataset to perform well in mountain regions with few borehole controls, and the authors of that dataset state as much in their paper; regardless, such interpreted bedrock data are currently being used to populate large-scale predictive models. Our work shows that approach is likely to be problematic if modeling goals include mountain baseflow dynamics and stream dewatering predictions.
L545 – this echoes my prior comment about 3D Valley composition. Agreed.
L559 - I think the authors have enough data within this study to conceptualize a 'why'. Why does Staunton not dewater? I compiled some simple plots to illustrate potential interactions of

interest. For instance, Staunton has both the deepest median DTB, but also the most confined valley (see Figures 1 to 3).

An additional plot of dtb, valley width, and volume of deposit in valley (dtb*width) reveals a negative correlation of temperature with dtb, a positive correlation of temperature with width and a strong negative correlation with volume. As such, it would seem the authors have stumble upon some zone of width/dtb ratio that offsets dewatering? I encourage the authors to dive a bit 'deeper' here, as I think they may have something novel to report here. Please see our response to our L585 comment below. There are some tantalizing potential research directions indicated in the existing data regarding bedrock depth and valley width, but we believe the hillslope recharge/storage dynamics must also be evaluated for such a more in-depth analysis of how these various physical stream valley controls interact to generate baseflow. Hillslope bedrock depth transect measurements are currently planned for the 2022 field season for the three focus subwatersheds (Paine, Piney, Staunton). Please stay tuned!

L574 – this is awesome! Thank you.

L585 – okay, this is what I was speaking to earlier. We have added to this statement regarding baseflow supply in Staunton: This apparent conundrum indicates the importance of bedrock depth (suprabedrock aquifer thickness) in facilitating spatially persistent baseflow generation during dry times.

L597 - also, Paine run has median valley width 5 m > Staunton. This detail was added to the sentence in question

Figure 11 - might say GW influence 'prediction' in the legend too. Agreed.

Figure 1 plotted relationship between depth to bedrock (DTB) and valley width for the Briggs et al. study. Where a strong negative correlation is found between DTB and valley width. I understand valley width is taken ~ 2 m above the valley floor, but this is still a meaningful measure.
Thank you for taking the time to plot these data together and for highlighting the apparent negative relation between the physical variables. We used a valley width of 2 m above valley floor so the measure would be less sensitive to fine scale topographic variation and instead better identify the true valley walls.

Figure 3 I encourage the authors to think about the relationship between volume (storage) and the incised natured of the coldest streams, such as Staunton. Given these rich datasets the authors have generated, I think there is bandwidth to conceptualize what may be at play here. This point is appreciated, but based on the lidar data from these catchments, we have not found systematic patterns in stream incision as related to summer temperature. We made some early attempts to predict measured bedrock depth by streambank height (ie channel incision) and that did not work for our test reaches, indicating stream incision is not directly related to bedrock depth in these coarse/rocky colluvial sediments. We might expect incision to be better related to bedrock depth, and therefore baseflow supply, in headwaters with fine valley sediments such as glacial till though fine sediments inherently have lower permeability and may inhibit groundwater exchange.