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

Review #3

Comment on hess-2021-622
Anonymous Referee #2
Referee comment on "Bedrock depth influences spatial patterns of summer baseflow, temperature, and flow disconnection for mountainous headwater streams" by Martin A. Briggs et al., Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2021-622-RC3, 2022
Review of HES-2021-622 https://doi.org/10.5194/hess-2021-622
I appreciated the opportunity to review this interesting paper by Briggs and coauthors entitled 'Bedrock depth influences spatial patterns of summer baseflow, temperature, and flow disconnection for mountainous headwater streams'. The work addresses important questions regarding the description of connectivity and interaction between groundwater and surface water in mountainous catchments. The authors develop in their paper an interesting vision at the interfaces between geomorphology, hydrology and hydroecology (principally fish habitats). They performed systematic measurements of depth to bedrock along stream corridors in eight headwater streams in Shenandoah National Park (Virginia USA) using passive seismic technics along with identification of wet/dry segments and measurement of river temperature.
They highlight 3 main important outcomes from these measurements:
that measured bedrock depths strongly deviate from the ones available in global-scale geologic and soil dataset. permeable streambed thickness is highly discontinuous along the stream channels. On zones with important depth to bedrock, the authors identified localized disconnection of stream flow channels during extended period of droughts. mean stream temperature during summer is negatively correlated with depth to bedrock suggesting preferential connectivity with groundwater with implications for stream aquatic ecosystems and habitats.
This paper has been carefully prepared and is well written. The introduction presents the context, state of the art and main questions in a comprehensive manner. The results are
interesting and their interpretation are well supported by a robust analysis. The discussion and conclusions will definitely trigger the attention of the readers of HESS. I have only raised few general points and made suggestions that could be helpful for the authors to develop the discussion and conceptualization of their results.
We greatly appreciate your time and thoughtfulness on this review. You have accurately summarized the three main high-level points we intended to make in the manuscript, and are gratified you believe these findings will be of strong interest to the HESS readership.

I have some concerns regarding the comparison between measured depth to bedrock and the one compiled in global databases. I agree with the authors that such databases might not be suitable to capture local properties of soil types or depth to bedrock along the river corridor. Nonetheless, there is a major difference in representative scales between the geophysical measurements and the estimates that are compiled in those databases. The depth to bedrock database from Shangguan et al. (2017) provides data over a spatial resolution of 250m, while the data presented here integrate a few cubic meters around the instrument (is the measurement scale actually mentioned in the manuscript?). I believe that it is still interesting to mention but I would recommend the authors to minimize its importance in the manuscript and acknowledge the main differences and complementarities between both datasets.

The HVSR data essentially represent a point measurement of bedrock depth, which as the reviewer correctly points out, is perhaps not directly comparable to the 250m scale global bedrock depth layer. However, we do believe that in aggregate the 191 HVSR measurements here are appropriate to compare to the corresponding 250m gridded data to assess general agreement in the data types. We have also shown in Figure A4 how bedrock depth is systematically overestimated by the global scale data layer. The point vs grid scale offset is difficult to escape, as the Shangguan bedrock depth model was created via using soil boreholes and well drilling observations, which also only integrate a limited spatial extent and could also be considered point measurements of bedrock depth.

In section 5.4 of the Shangguan and others (2017) paper, they note that the DTB predictions should be used with caution, particularly for shallow depths to bedrock and in mountainous regions. This is due to a reduction in variance (under prediction of deep DTB and overprediction of shallow DTB) resulting from the ML regression and due to a sparse number of observations in mountain ranges. In the discussion section, we now incorporate the Shangguan and others self-recognized limitation of their bedrock depth layer mountain settings such as Shenandoah NP>

We believe our study points to the additional work needed to characterize bedrock depth for headwater streams, and our comparison with the Shangguan dataset simply highlights problems that are likely for any global/regional model of bedrock depth based on limited borehole data. This ‘point’ is timely, as such large-scale geologic datasets are currently being used as templates for a range of predictive models that include baseflow generation dynamics.

It also remains unclear to me to what geomorphological processes/features of the landscape the measured depth to bedrock are assigned to: preferential erosion, fracturation/weathering, sediment accumulation, all of them without distinction? I believe that it would be important to link the measured stream corridor depth to bedrock and streamflow behaviors to some knowledge of local catchment-scale geomorphology/geology. This could help identifying generic information to be transferred to other catchments (or at least provide guidance). For example, in table 1, it seems that there is an inverse correlation between valley width and DTB. Also, one would expect that DTB impacts drainage density (dd~K*DTB) and intermittency (through aquifer volume available V~ 2*DTB*river length*hillslope length ~ DTB*river length/dd).

Exploring such generic relationship would help to conceptualize the results and increase the impact of the paper in my opinion.
The Reviewer is correct in that we generally did not attempt to identify the geologic mechanisms that controlled bedrock depth variability along the study streams or between them, except for Paine Run where there were more clearly pockets of alluvium and colluvium built up along a shallow, oft exposed, bedrock surface. We also speculate that the 20m+ deep apparent trough in the bedrock surface found along Piney River is an unmapped fault zone. We did not originally highlight the negative relation between measured bedrock depth and valley width but do so in the revised text. Assuming that the unconsolidated material along the valley floor is generally sourced from eroding hillslope colluvium, it does stand to reason that more narrow valleys show thicker deposits. While we understand that the omission of a more in depth analysis of bedrock depth controls for the study site may be unsatisfying to readers with geologic and geomorphic background and that the development of a transferrable relation could be useful to the community. In a revision of this paper we would explore rank order correlations between the physical valley variables already mentioned (bedrock depth, width means and variance; channel slope means and variance).

Some references:
Litwin et al 2021 https://doi.org/10.1029/2021JF006239 Great paper, we will add the citation to a revised version of the manuscript and take the findings into account when discussing our study.
Luo et al. 2010 https://doi.org/10.1130/G30816.1
Warix et al., 2021 https://doi.org/10.1002/hyp.14185 Great paper, we will add the citation to a revised version of the manuscript and take the findings into account when discussing our study.
Ilja van Meerveld et al. 2019 https://doi.org/10.5194/hess-23-4825-2019 We intended to cite this related paper in the original submission but that was somehow dropped along the process. We will add this paper to a revised version of the manuscript.

I believe that this work brings very interesting insights and data for our understanding of the impact of depth to bedrock to flow continuity and groundwater-surface water exchanges in mountain regions. I recommend the paper to be published in HESS.
Thank you.

Please
also consider few minor points listed in the following.
Specific comments:
l145-149: likely to be biased by the location of wells preferably implemented downhill and where more productive aquifer maybe be identified.
Agreed, this sentence was added: ‘In more typical headwater systems, existing wells may be preferentially installed to maximize the production of water and not broadly sample the true range of bedrock depths.’
l167-169: do you mean in context where the water table is close to the surface? i.e. when

K/R (R=recharge) is low? The text ‘for mountain stream corridors’ was added to this statement for context, as when there is a stream present through permeable mountain sediments the water table is inherently close to/at the land surface (at the stream).
l234: how to differentiate sediment accumulation from weathering/fracturing development that can also enhance K? Sediment accumulation on a bedrock surface is expected to result in a clear/interpretable HVSR measurement, while heaving fracturing/weathering of the bedrock surface is expected to result in HVSR measurements that are of low confidence or are unable to be interpreted. That is because the method depends on distinct vertical changes in acoustic impedance. Therefore, our HVSR dataset is weighted toward evaluating zones of sediment accumulation on a low permeability bedrock surface as described generally in this paragraph.
l320: I did not understand how atmospheric effects were filtered here.
l326: providing the equation of BFI would help the readers that are not familiar with this Index BFI is not determined by a single equation, but through by connecting points along the hydrograph recession that are not expected to be impacted by storm/quickflow with straight lines and then calculating the daily ratio of baseflow to stormflow as distinguished by the lines. A reference to Barlow et al 2014 was added to this methods statement to aid the interested reader.
Figure 4: why showing depth in log here? I think it masks the actual variability of your Dataset. That is true regarding the deeper bedrock depth anomalies, but we found the log scale important in showing variation among the shallow bedrock depths that dominate the dataset.
Figure 4: 1 m seems to be the minimum depth measurable, correct? Is it mentioned in the manuscript? You may be correct, but we have not rigorously evaluated < 1m sediment thicknesses with the method and are not aware of other published results that address this. You do point to a potential issue, in that some number of measurements that we were not able to interpret may result from < 1m true bedrock depths, and those points would be preferentially dropped from the analysis.
table 1: it seems that there is an inverse correlation between valley width and DTB. Do you see correlation between drainage density and DTB? Since dd ~ K*DTB. It would be interesting to assess the relationship between landscape topography and measured DTB to identify generic relationship that could be transferred to other catchments.
Absolutely, as noted above we plan to more explicitly describe and discuss the negative relation between valley width and bedrock depth, thank you for pointing that out. In this response we are assuming the reviewer is referring to "drainage density" (length of stream/drainage area) in the same way USGS sometimes uses drainage density as a basin characteristic when developing peak flow or low flow statistics. The first example that comes to mind is Bent and Archfield, 2002 (https://pubs.usgs.gov/wri/wri024043/pdfs/report.pdf) where they found that drainage density was inversely related to the probability that a stream flows perennially in Massachusetts in a wider/unconfined valley types. The study catchments are confined valley settings there is a relatively small range in drainage density between study sites; we feel that any relationship between DTB and drainage density that emerged might be applicable to only to small percentage of the otherwise wide range of drainage densities

possible in headwater streams but we will explore this topic in a revision of the manuscript.

l398: how is this analyzed/filtered? Text added: 'Paired air and water annual temperature signals exhibited a spectrum of shallow groundwater influences as indicated by extracting fundamental sinusoids from each multiyear temperature dataset per methods described by Briggs et al. (2018). Observed phase shifts between stream and local air annual temperature signals ranged from approximately 5 to 30 d with a mean of 11 d.'

Figure 8: I did not fully understand how this graph is interpreted. A complete interpretation of Figure 8a does potentially necessitate some additional background provided by Briggs et al 2018 and better cited in the revised manuscript.

l423: I did not fully understand what this means? Did you remove an outlier to improve Statistics? No, what we mean to indicate is the significance of the relationship is driven by the Staunton River data point, such that if that site was removed a linear relationship determined for the remaining seven sites is not significant.

Figure 9: it would be useful to add the confidence interval on this plot. We are not clear if the Reviewer suggests a confidence interval be added to the regression or the HVSR (or temperature) data points.

l450: I do not understand why "(low permeability)" is added between parenthesis here. Please clarify your meaning. These parentheses were removed. Our intention is to indicate that in settings of high bedrock permeability, bedrock depth may be a less-important control on shallow groundwater flowpath dynamics.

l479: they concern different spatial scales. Not sure how we can interpret this result. Please see our response to a similar point made above

general comments.

l495: I fully agree with this statement. However, the resolution of this database is way lower than the scale you are interested in. In consequence, it may appear obvious that differences exist. Yes, the disparity is perhaps expected/obvious, but in reality (in our experience) these types of large-scale geologic datasets are currently being used to inform models of subbasin scale GW/SW exchange processes. As our study is unique in measuring bedrock depth at relatively high spatial resolution along several mountain streams we feel these comparisons are worthwhile. We do not mean to negate the general value of the large-scale datasets but instead point to important challenges in specifying the geology of mountain stream networks.

l619: I find hazardous to compare two different years with different recharge records. The BFI is integrative of full baseflow period, but may not be representative of the punctual measurement performed. Could you clarify this point? We calculate BFI on seasonal timescales for this study and report the values from 2015 and 2019 here to show the similarity in calculated baseflow fraction such that these two years were comparable from a baseflow perspective for the purposes of discussion.