Reviewer 2, Phil Benson

This is very nice work, and a very interesting paper. The authors have collected a huge rock mechanics and rock physics dataset on volcanic rocks collected from Mt Unzen (Japan); the site of extensive activity in the early 1990's. It is always pleasant to see new datasets presented and published, as these data are hard to collect and will benefit the community for many years. I have only a few general comments, and some minor points for the authors to consider.
We thank the reviewer, Dr. Benson, for his synopsis and comments, which are addressed individually below.

General comments:
1. How were the blocks selected in the field? Was this opportunistic, or were the sites selected via some form of criteria? This could, of course, be as simple as to cover a range of rock physical properties in 'accessible' locations, but it'd be nice to have that directly stated.
We thank the reviewer for the comment. The answer is rather complicated, but in short, yes, we aimed to select a range of porosity blocks which were both isotropic and anisotropic. We actually selected around 14 blocks to bring back to the UK for a range of distinct studies. In the field these were selected from block and ash flow deposits due to accessibility, initially for favourable size and shape (to ensure we could get enough samples for repeats but also that we could carry and ship them) as well as a couple of criteria, for example we wanted them to appear uniform across the whole block (no large porosity contrast, etc) and not have any block-scale damage like through-going fractures that would impact ability to prepare samples. Once we had selected a number of samples, we examined them to compare to what we had seen during our time in the field (prior to sampling we had been working on the lava dome), and we also made some basic density estimates and measured some samples using tinyperm (though this was not very consistent as weather was poor and measurements do not work on wet surfaces). We then tried to fill any gaps in our sampling, knowing we ought to represent approx. 1.5-2.4 g/cm^3 density (see Kueppers et al., 2005- https://doi.org/10.1016/j.jvolgeores.2004.09.005). For example it was difficult to locate mid-porosity samples (20-25 %) which did not bear a fabric, though this might represent a slight bimodal distribution which has been previously observed at Unzen (Kueppers et al., 2005). We also knew that much of the dome contains sheared magmas (e.g. Wallace et al., 2019), so selected the block with visible shear textures on the surface. When we were able to bring the samples into the lab we cut each to see the internal texture, prepared a slice for thin section and a core to measure porosity. After that we selected which samples could be used for each study (e.g. this study, and Coats et al., 2018, etc.) and compared to our previously selected samples (e.g. Hornby et al., 2015). Because the purpose of the present study was to observe the impact of porosity across the spectrum of materials, we selected samples that spanned the density range measured at Unzen previously (Kueppers et al., 2005, compared to our Supplementary Table 1), and used the sample with cataclastic banding collected specifically for the study to compare the role of anisotropy. Our existing description in section 2.1.1. sample collection, has been modified to describe our sampling rationale:

"During a field campaign in 2015 a suite of blocks, each > 15 kg were collected from block-and-ash flow deposits on the eastern and north-eastern flanks (Fig. 1b). The samples were assessed in the field to ensure representative texture and estimated densities that matched the known range of physical attributes of Unzen lavas (cf. Kueppers et al. 2005). The target was to select 4 blocks for this study which spanned low (UNZ14), medium (UNZ1) and high (UNZ13) porosity, plus an additional block that displayed an anisotropic cataclastic fabric (UNZ9) as the summit lava dome is pierced by shear zones (see, e.g. Wallace et al., 2019). Blocks UNZ1 and UNZ13 were also used for the study by Coats et al. (2018) which examined the role of temperature, alteration and strain rate on the rheological response to deformation at high temperature and defined a failure criterion for porous dome rocks and lavas."

2. Somewhat covered by the earlier review: Permeability is easily one of the most tricky parameters to measure and discuss, particularly in the field, and in terms of spatial variation. Many years ago a NERC scheme (micro to macro) identified that permeability needed a measurement every few metres to see such variations, compared to 100's of metres and km to resolve parameters like elastic wave velocity and conductivity. That's just one example, but perhaps this type of 'challenge' is worth reinforcing when introducing and discussing the general nature of heterogeneity inherent in volcanic deposits of all kinds.
We thank the reviewer for this suggestion, we tried not to venture too far beyond the bounds of our dataset, since we do not tackle upscaling directly, yet we wanted to stress the importance for the volcanological community to 1) use real values and data collected and 2) develop approaches that can actually upscale these values in a meaningful way. We modified some of the discussions around permeability due the comments of reviewer 1 also, and we made sure to mention the difficulties of upscaling permeability, which is a vital parameter in interpreting volcanic unrest. We really need, as a community, to tackle upscaling of permeability in a meaningful way.

Minor queries:
3. Line 137: Typo, "ultransonic" should be 'ultrasonic'.

Thanks, changed.

4. Line 14: For a recent report on damage and Vp changes in volcanic rocks see, for example: - Harnett, C.E., P.M. Benson, P. Rowley, and M. Fazio (2018), Fracture and damage localization in volcanic edifice rocks from El Hierro, Stromboli and Tenerife, Scientific Reports, 8, 1942, doi: 10.1038/s41598-018-20442-w.
We thank the reviewer for this suggestion, which we have now cited this relevant manuscript in a number of places, in both introduction and discussion.

5. Line 300: A pore pressure differential of 1.1 to 1.5 is actually fairly high considering the confining pressures of 5.5-13.5 MPa. Leading to what is, in effect, an 'effective pressure differential' across the length of the sample (rule of thumb being dP of around 10% of Pc, so 1.3MPa for the 13.5MPa experiment). Might the author comment or add a few words here? I suspect this protocol was adopted simply due to the low permeabilities of the rock types investigated, but it'd be good to have this confirmed by the authors.
We apologise that this was actually imprecision in our description of the methodology. We actually varied the flow rate until the inlet and outlet pressure stabilised, we targeted an outlet pressure between 1.1-1.5 MPa (which we chose as a target – this was the reference to the setpoint in the previous version of the method), at which point we locked the pressure of both inlet and outlet (i.e. set the pressure differential) and let the flow rate equilibrate again before taking the permeability measurement. This is a slightly unusual approach, chosen as the samples span quite a range of permeabilities. The average differential pressure of our measurements was actually 0.28 MPa (not 1.1 to 1.5 MPa as was wrongly insinuated, which was the outlet pressure), all the pressure and flow rate data for each test are shown in Supplementary Table S3 (previously S2). We thank the reviewer for highlighting the confusing description of the methodology, which we have now rectified in section 2.1.6 (Confined water permeability).

6. Line 375: I wouldn't call this sub-section heading "Acoustic emissions - active": surely you mean simply "elastic wave velocity"? Or perhaps "active surveys"? Rather a minor quibble, but I do think the use of AE is implied as passive only and this is well established in the literature.
We agree that "active surveys" is a good title, our original terminology was due to wanting to clarify the distinction between section 2.2.4 and 2.2.5 but is not needed, so we have changed it.

7. Lines 695 (figure 8): What is the error on the velocity changes? Apologies if it is in the text, and I missed it when reading up to this point.
This is a good question, and not one we can fully quantify, unfortunately. The acoustic emissions we recorded were at KHz, so it might be reasonable to say that there was an error of at least 0.001%. However, we expect it could be higher due to potential misalignment of the pulses during the stacking stage. Even misalignment of 1 sample would introduce errors, and because tests are noisy this is possible. So, we anticipate maybe the error would be on the order of 0.01% velocity change, i.e. approx. 1 % of the measurement. As this is speculative, we added a comment as to the accuracy of the values in the discussion (as opposed to the method which states the sampling rate).

8. Lines 935-945, and a few other places: Do the authors note any differences in the character of the AE with regards to the dry and saturated experiments? This is a well known phenomenon in volcanic systems with the inherent fluid-rock coupling, for example: - Fazio, M., P.M. Benson and S.V. Vinciguerra (2017), On the generation mechanisms of fluid-driven seismic signals related to volcano-tectonics, Geophysical Research Letters, 44, 734-742, doi:10.1002/2016GL070919. - Fazio, M., Salvatore Alparone, Philip M. Benson, Andrea Cannata, Sergio Vinciguerra (2019), Genesis and mechanisms controlling Tornillo seismo-volcanic events in volcanic areas. Scientific Reports, 9, 7338, doi: 10.1038/s41598-019-43842-y I leave it to the authors as to whether they think it is worth including, or out of the scope of their study.
We thank the reviewer for the comment and we also would very much like to explore this further. However, we have not been able to make this distinction, unfortunately - due to the active surveys we had a lot of noise in our wet samples, and we were not able to reliably exclude the survey pulses from the passive AE recording, hence our omission of passive AE in the saturated tests in the manuscript (note we were able to extract enough pules with confidence to track the velocity, but not all of them so as to reliably trust the passively recorded data).