

CVK Test-Retest Investigations

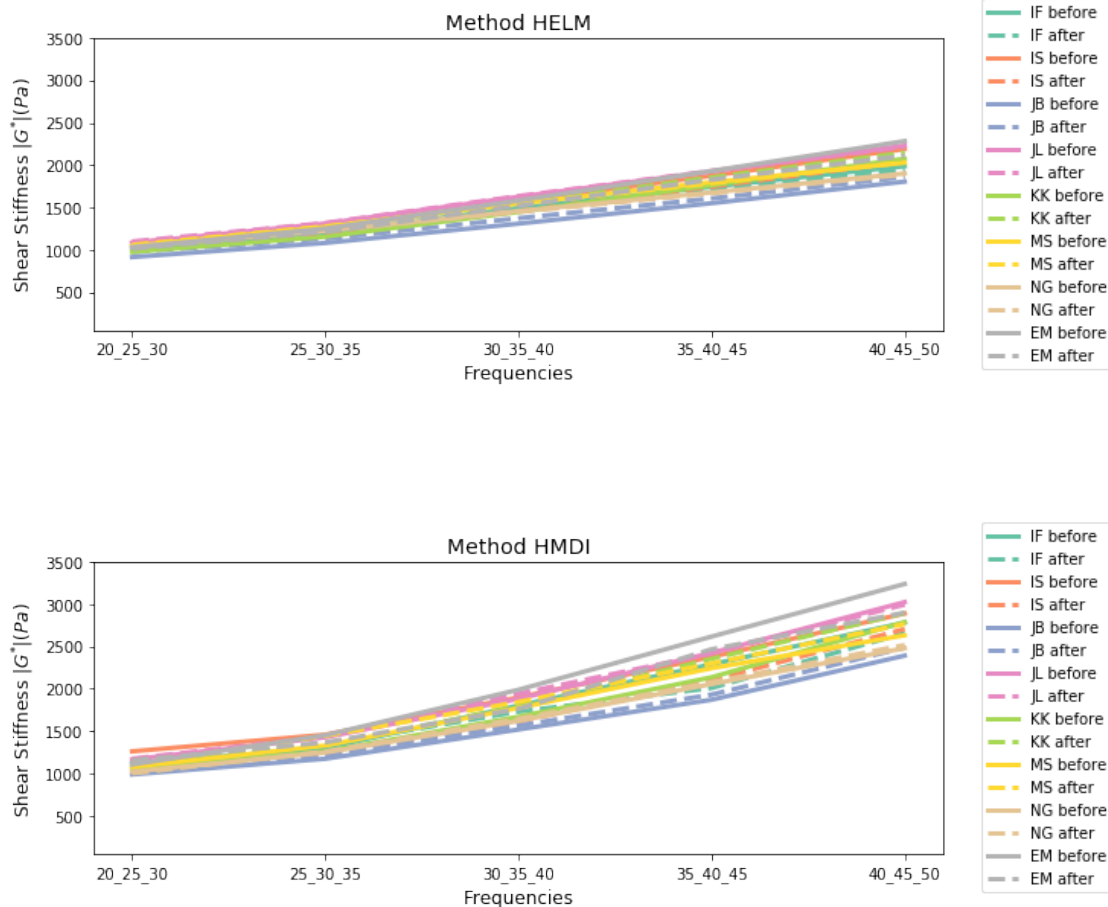
May 8, 2018

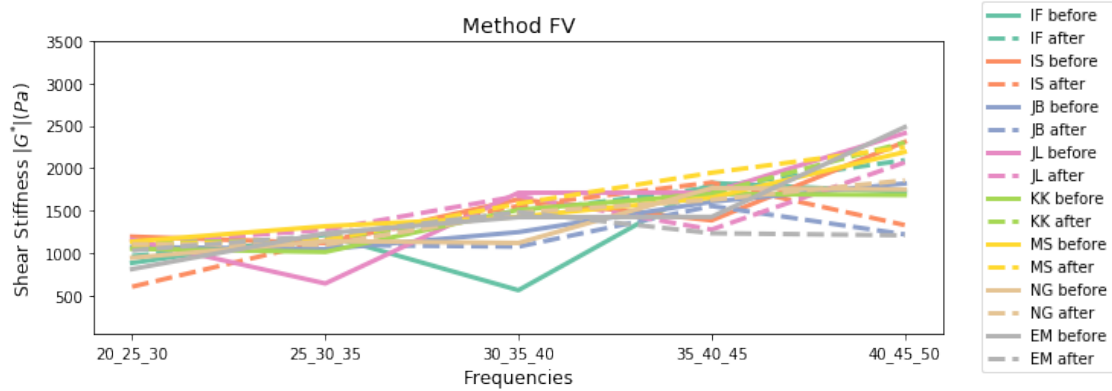
Here I look at by-frequency sliding window elasticity results for the CVK Virchow test-retest set. I compared HMDI and MREdge-MDEV (called here HELM). The only difference between them was inversion method -- all other pre-processing was identical.

1 Wideband sliding windows

Elasticity results for each subject at each sliding frequency window are shown below. Test is continuous line and retest is dashed.

Elasticity results for CVK test-retest cohort



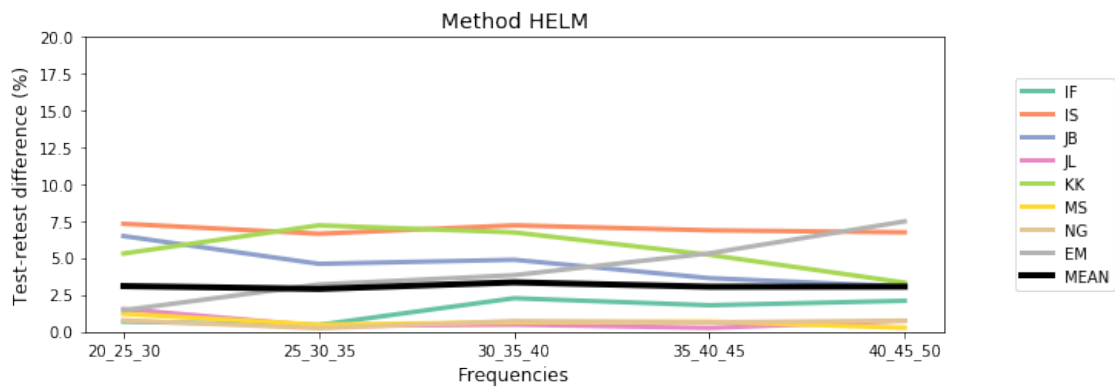


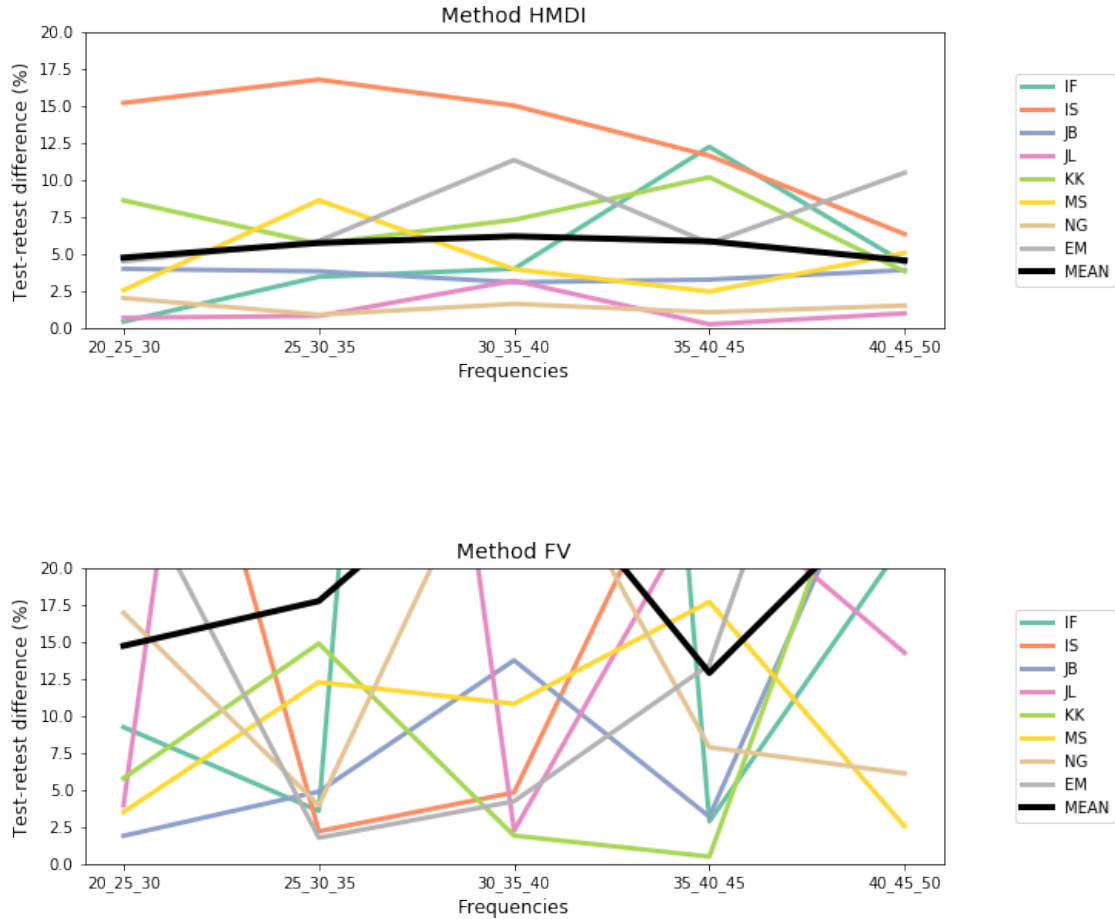
Like with the HMDI / four-groups paper, there is good consistency across frequencies for each subject (stiff subjects stay stiff, soft subjects stay soft) supporting our findings that the range is from natural variation, and our methods are robust.

HMDI and HELM also agree very clearly on who is stiff and who is soft. They also agree on values at 25 Hz. However, as the frequencies go up HMDI gets much higher than HELM. Again, all the denoising is the same: so is HMDI just more noise-robust?

Below are the test-retest differences broken out by subject and sliding frequency window:

Test-retest differences by sliding window for CVK test-retest cohort





2 Just to be clear, we are not looking at stiffness increasing or decreasing with time, but consistency between test and retest getting worse.

3 main observations:

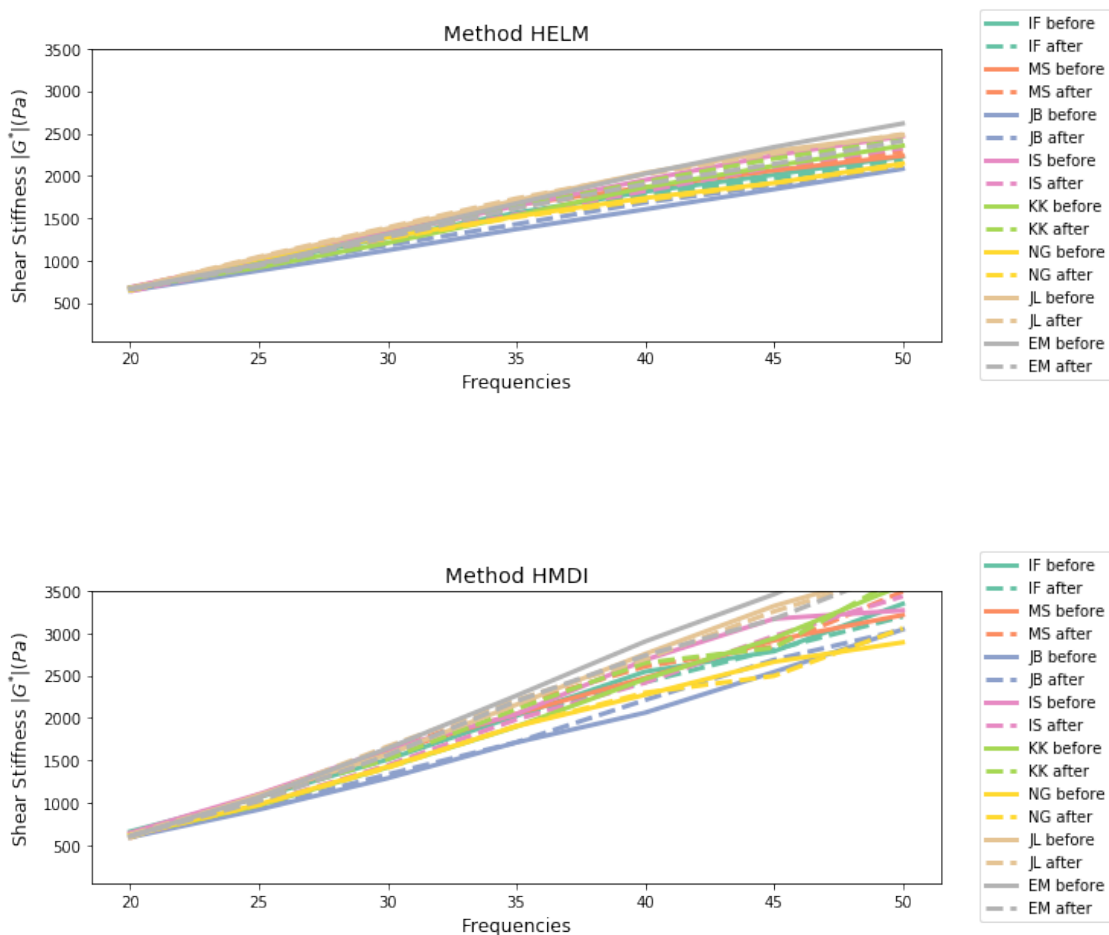
- MREdge-HELM is clearly stabler than HMDI here. Keeping every other aspect of the inversion exactly the same, the HMDI inversion is sensitive to something that HELM is just not. This is particularly surprising since HMDI can easily be shown to be more noise-robust than HELM in simulation testing. So what is it?
- Overall the lower frequencies are stabler than the higher frequencies, but there is some zig-zagging. The frequencies were acquired [20, 25, 30, 40, 50, 35, 45]. From these data we cannot sort out in these charts whether it was acquisition order (more movement as the acquisition went on) or frequency that causes the growing instability.
- And, obviously, the FV method still has stability problems and will be dropped from the remaining analyses until regularisation is in place.

My guess so far is, just like with abdomen, movement by the patient causes the images to blur and the values to drop. Since the only difference is inversion method, I think that HMDI's global solver is affected by patient motion more than a purely local solver. (If it gives you hope for the future, recall that the finite-volume solver is the most local of all.)

3 Wideband individual frequencies

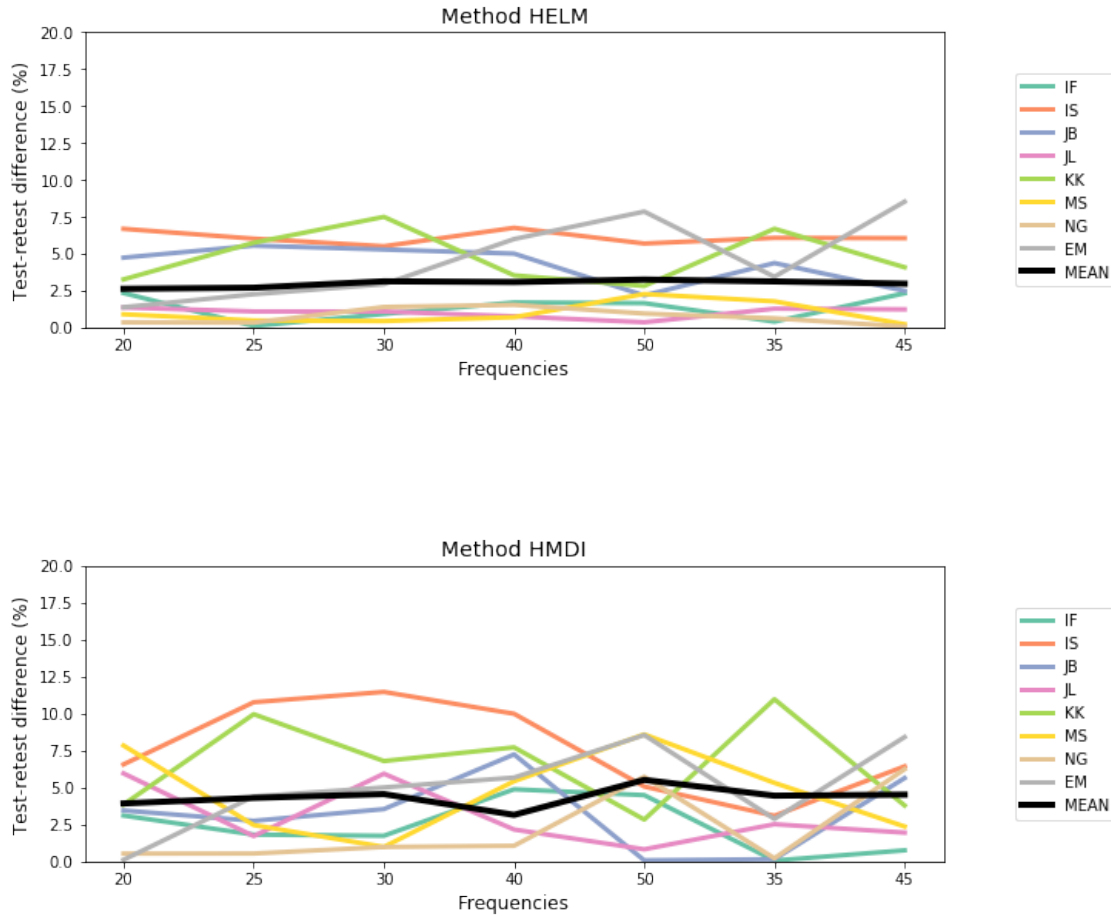
First we plot elasticity results by frequency. Again, HELM steadiness over the frequencies is very impressive. Again, HELM and HMDI agree on who is stiffest and who is softest, but HELM is more stable. Also, HMDI appears to be getting too high at 50Hz:

Elasticity results for CVK test-retest cohort



The HELM image is particularly stable and convincing. This is good news for the MREdge pipeline going forward, as some groups request that in multifrequency MRE, the values at each frequency are reported. This chart shows that the new protocol and pipeline can do this in a very stable way.

Test-retest differences by sliding window for CVK test-retest cohort



The test-retest differences appear to show no particular relation to frequency or lateness in the cycle.

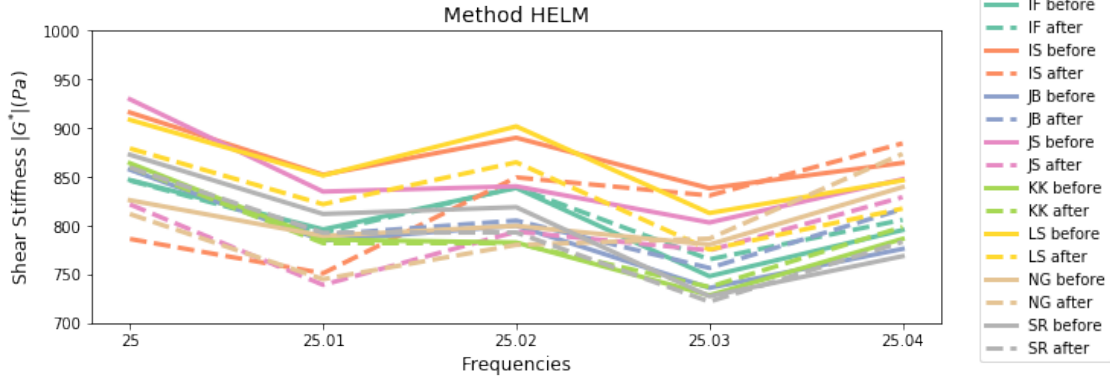
The mean line in the HELM acquisitions is now essentially linear. The wideband data thus strongly suggest that test-retest accuracy is dependent on *frequency* rather than *placement* in the acquisition sequence.

4 Narrowband

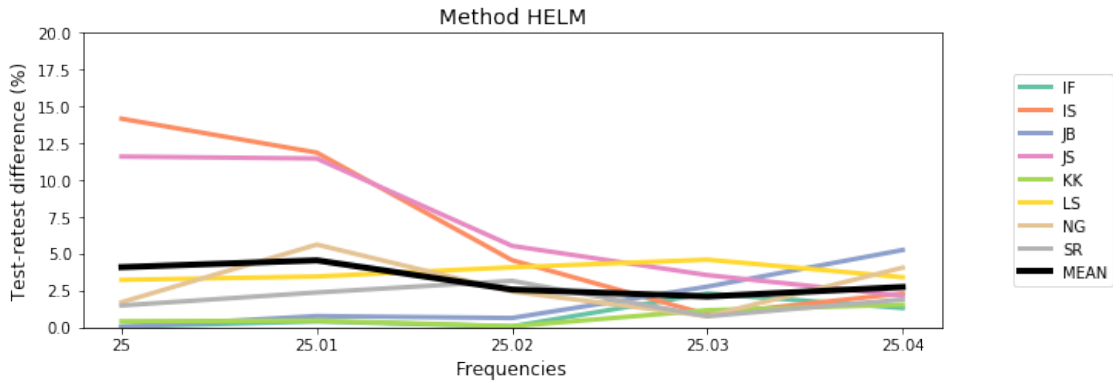
We now look at narrowband data, in which the same frequency was acquired 5 times in a row. For a labeling convention, the five acquisitions are labelled as 25, 25.01, 25.02, 25.03 and 25.04 respectively; however the true frequency was always the same, 25 Hz.

In this case there is no meaning to the sliding windows or to the sorting of frequencies, so we just look at the data set in the acquired order, and stick with HELM:

Elasticity results for CVK test-retest cohort



Test-retest differences by sliding window for CVK test-retest cohort



I see the following conclusions:

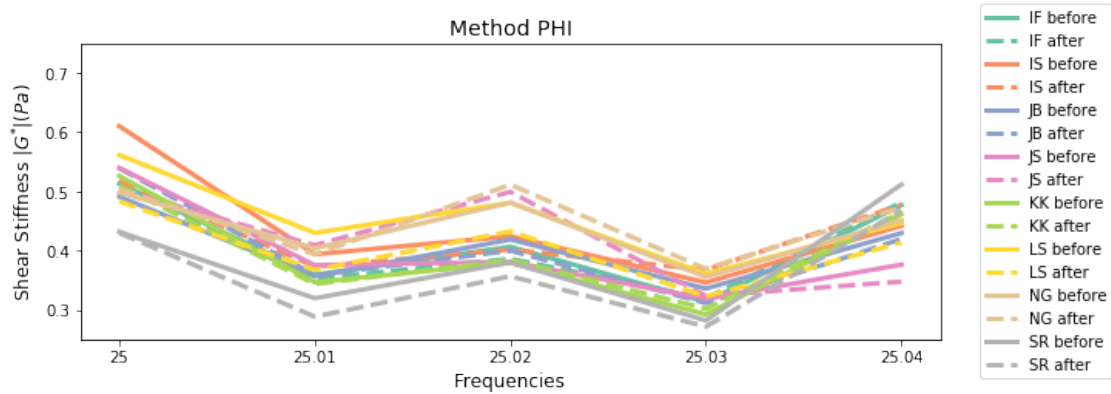
- The oscillation of the narrowband elasticity results across the five acquisitions is fascinating. From other recent experiments, we are not surprised by the oscillation. But why would the particular up-down pattern be the same for all the subjects? That is, whatever drives this natural oscillation, laying the subject in the scanner appears to put them all "in phase".
- Here we see that test-retest results controlling for frequency are mostly extremely consistent, while two subjects show the *opposite* trend from above, that is, the reproducibility increases over the course of the acquisition and reaches a steady state.

5 Stability of ϕ and G' variables

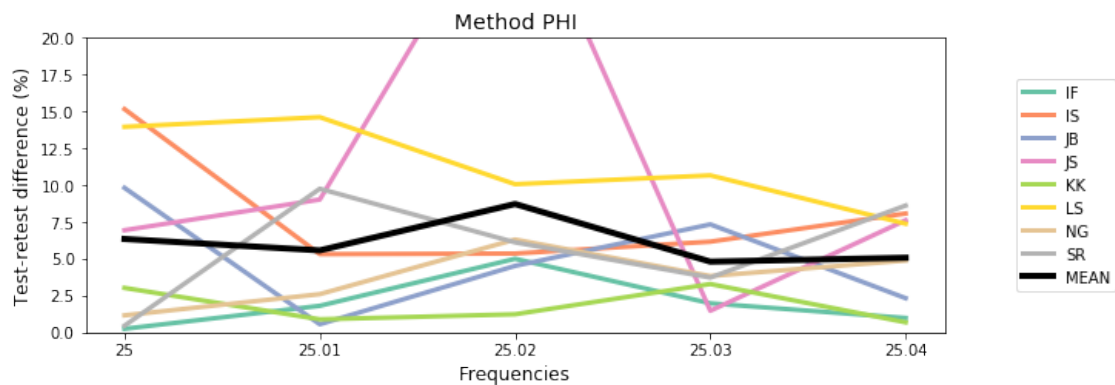
One possibility for the greater stability of HELM was that it is evaluated locally, however the FV method was not stabler despite it being evaluated locally. Another possibility is that the mathe-

metics of HELM -- handling real magnitudes rather than complex values -- has a large stabilizing effect. I thus decided to look at Phi values as well as a storage modulus (G') value reconstructed from the combination of $|G^*|$ and ϕ .

Elasticity results for CVK test-retest cohort

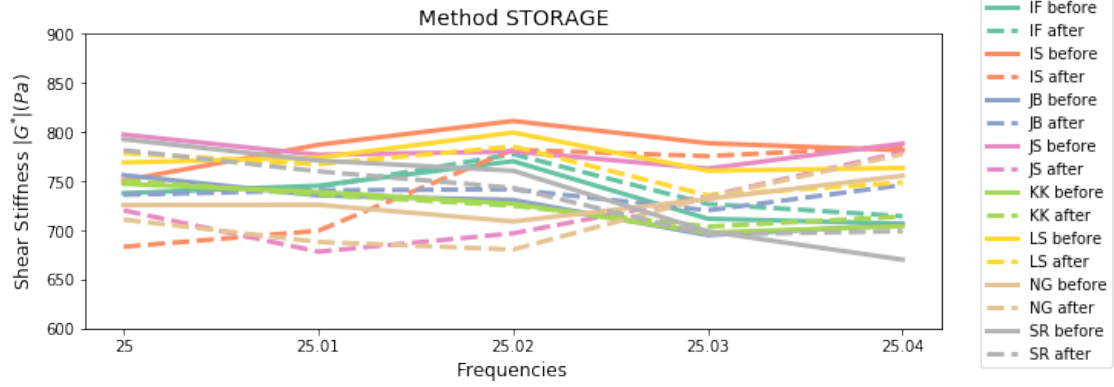


Test-retest differences by sliding window for CVK test-retest cohort

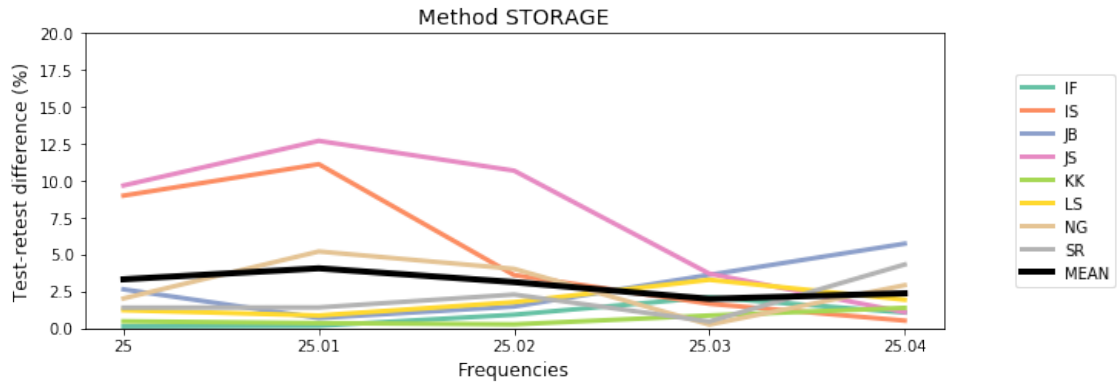


ϕ shows similar oscillations to G , although more pronounced. The test-retest profile is clearly quite different, the outliers are not all the same or at the same time points. Thus some of the test-retest differences in elasticity may be physiology rather than just motion. Also, the normalized test-retest differences are considerably larger, which could either be due to the instability of the variable, or the phase angle being more physiologically variant. Next I plot storage modulus G' :

Elasticity results for CVK test-retest cohort

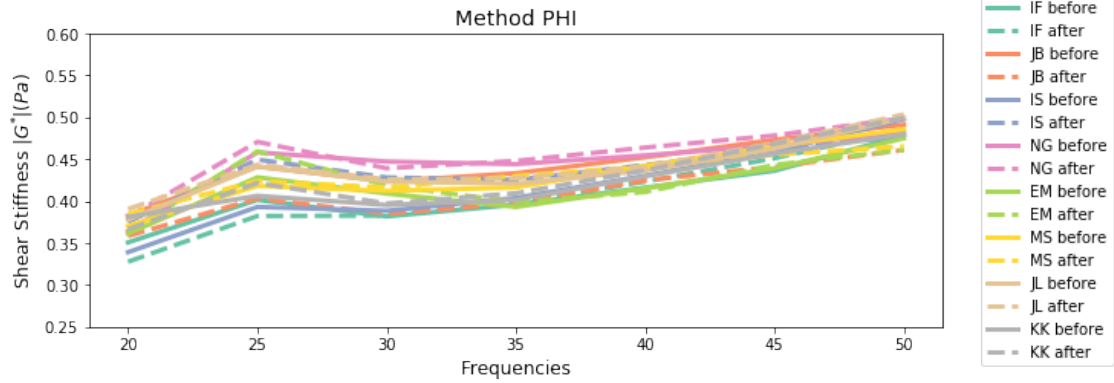


Test-retest differences by sliding window for CVK test-retest cohort

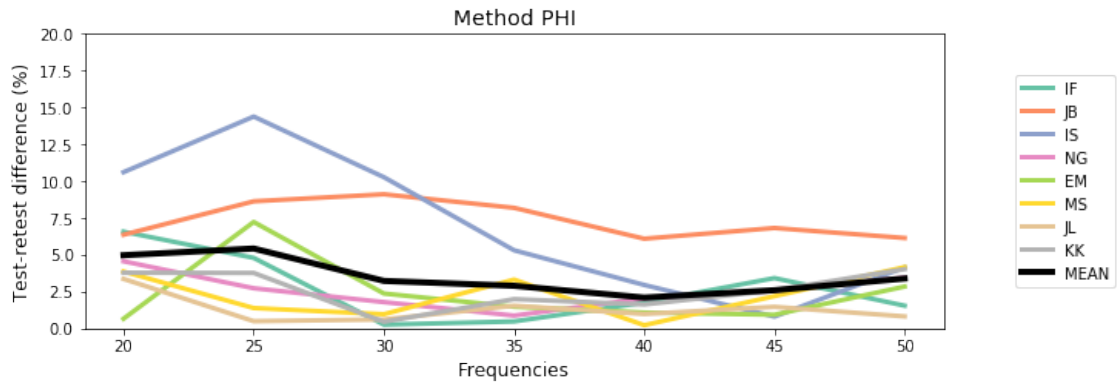


Storage modulus in the narrowband case seems less stable. The next sections will look at the wideband results for ϕ and G' .

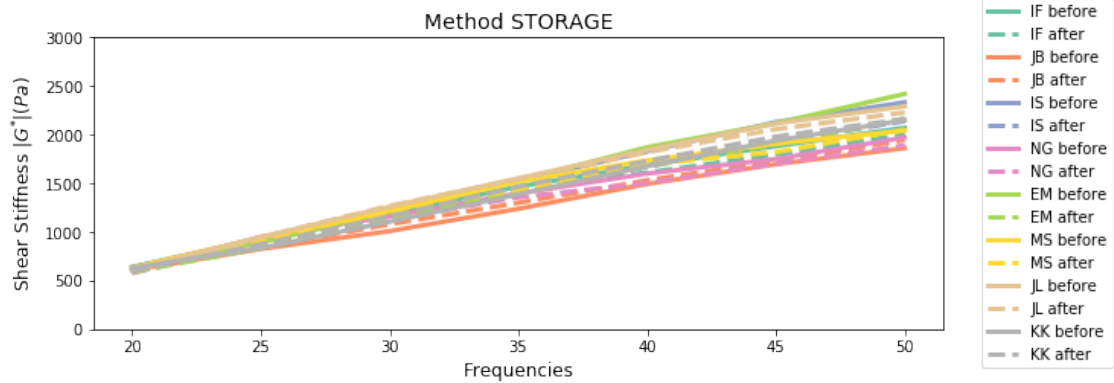
Elasticity results for CVK test-retest cohort



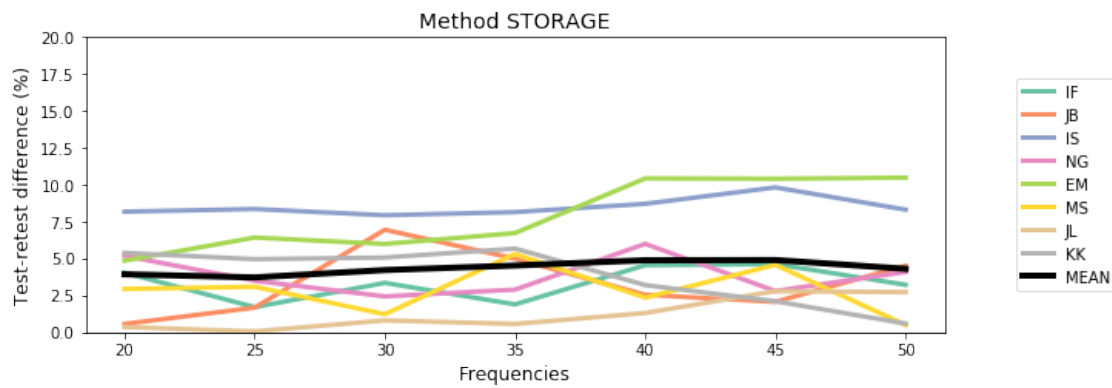
Test-retest differences by sliding window for CVK test-retest cohort



Elasticity results for CVK test-retest cohort



Test-retest differences by sliding window for CVK test-retest cohort

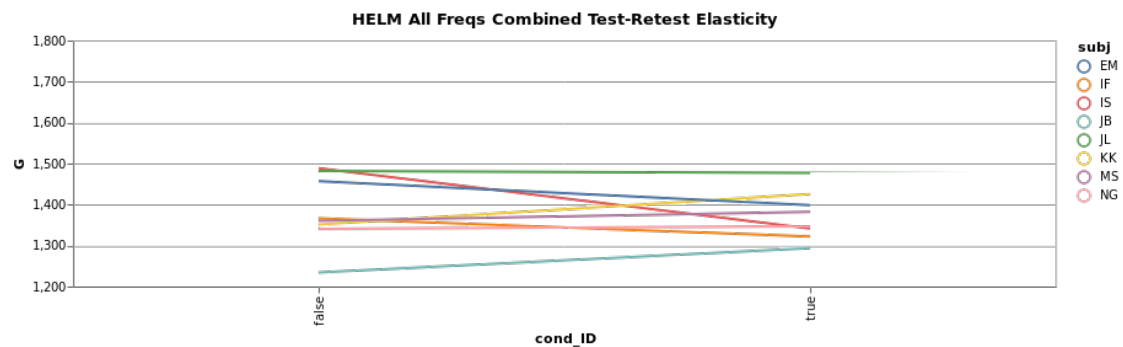


Storage modulus is slightly less stable than $|G^*|$, but not dramatically. It shows the same frequency-dependent test-rest discrepancies, slightly worse.

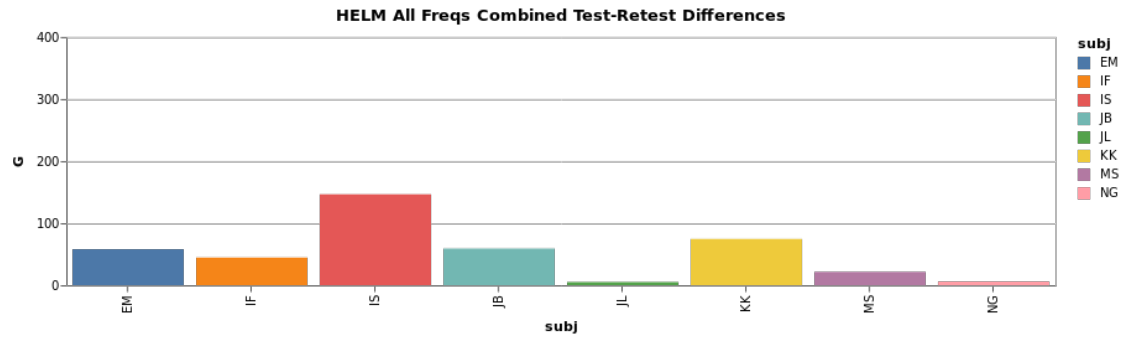
6 Test-Retest for combined broadband

Here we look at whether the combination of all frequencies improves test-retest discrepancies and what the ranges are.

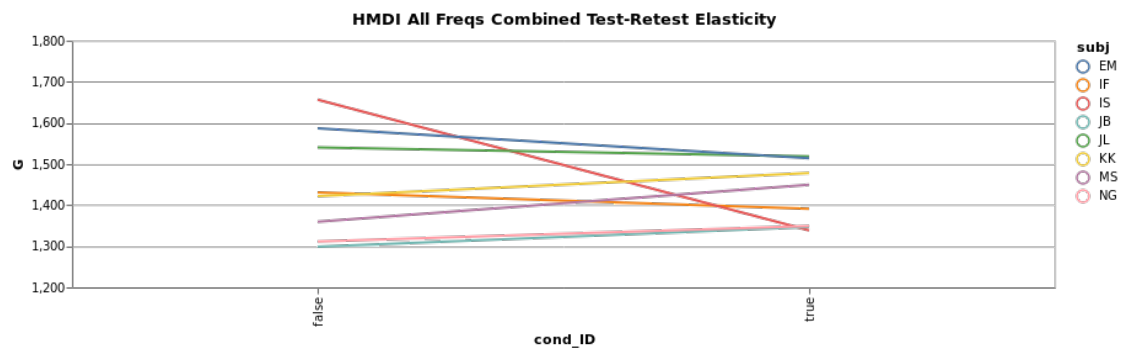
Out [413] :



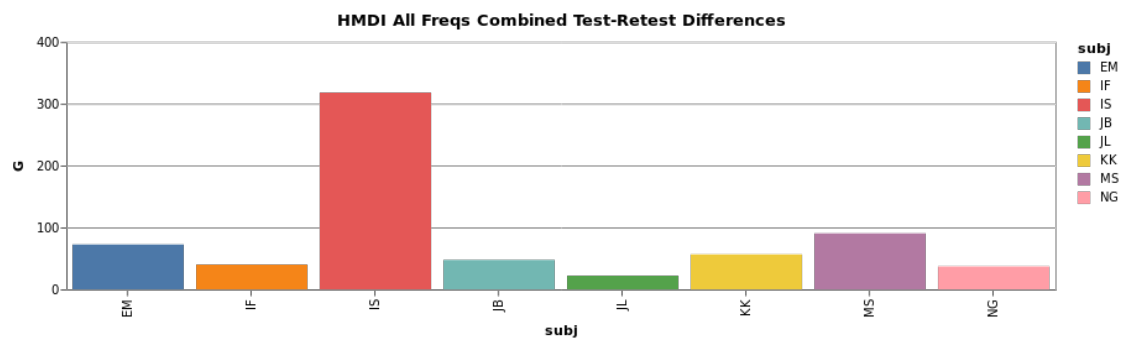
Out [414] :



Out [415] :



Out [416] :



Other than one very bad outlier (the mysterious subject IS), HMDI with all frequencies is comparably stable to MDEV, with medians nearly identical:

Mean HELM test-retest difference: : 51.944

Mean HMDI test-retest difference: : 85.440

```
Median HELM test-retest difference: : 51.429
Median HMDI test-retest difference: : 51.976
```

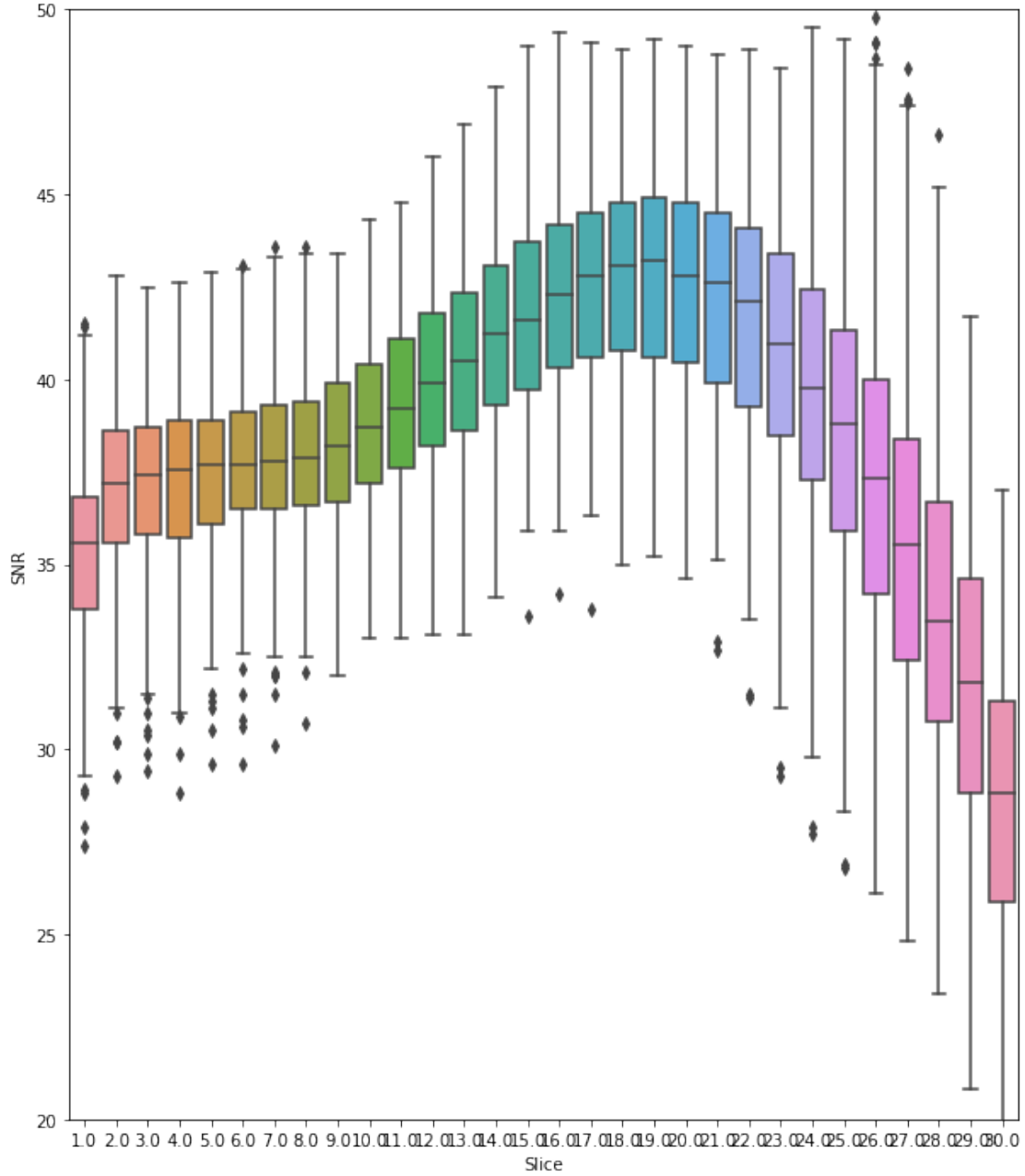
Broadband MRE does indeed seem to do its job of increasing overall reliability of the acquisition as compared to a single-frequency acquisition.

7 Slicewise SNR for the acquisitions - 3D Denoising

Here I investigated whether slicewise differences in displacement SNR are consistent across all acquisitions. The wideband data was investigated as there seemed no reason for narrowband to be different.

Below is a boxplot, showing SNR by slice for all acquisitions and frequencies lumped together. There appears to be a smooth overall trend, however with substantial errorbars. Tests of individual frequencies showed individual results to be similar to this plot.

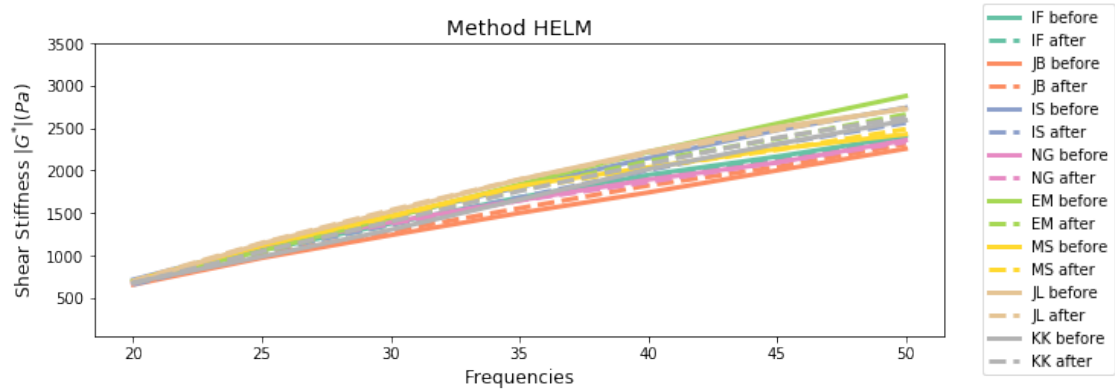
```
Out[590]: [(20, 50)]
```



8 Slicewise Denoising Then Dejitter (SDTD): SNR Results

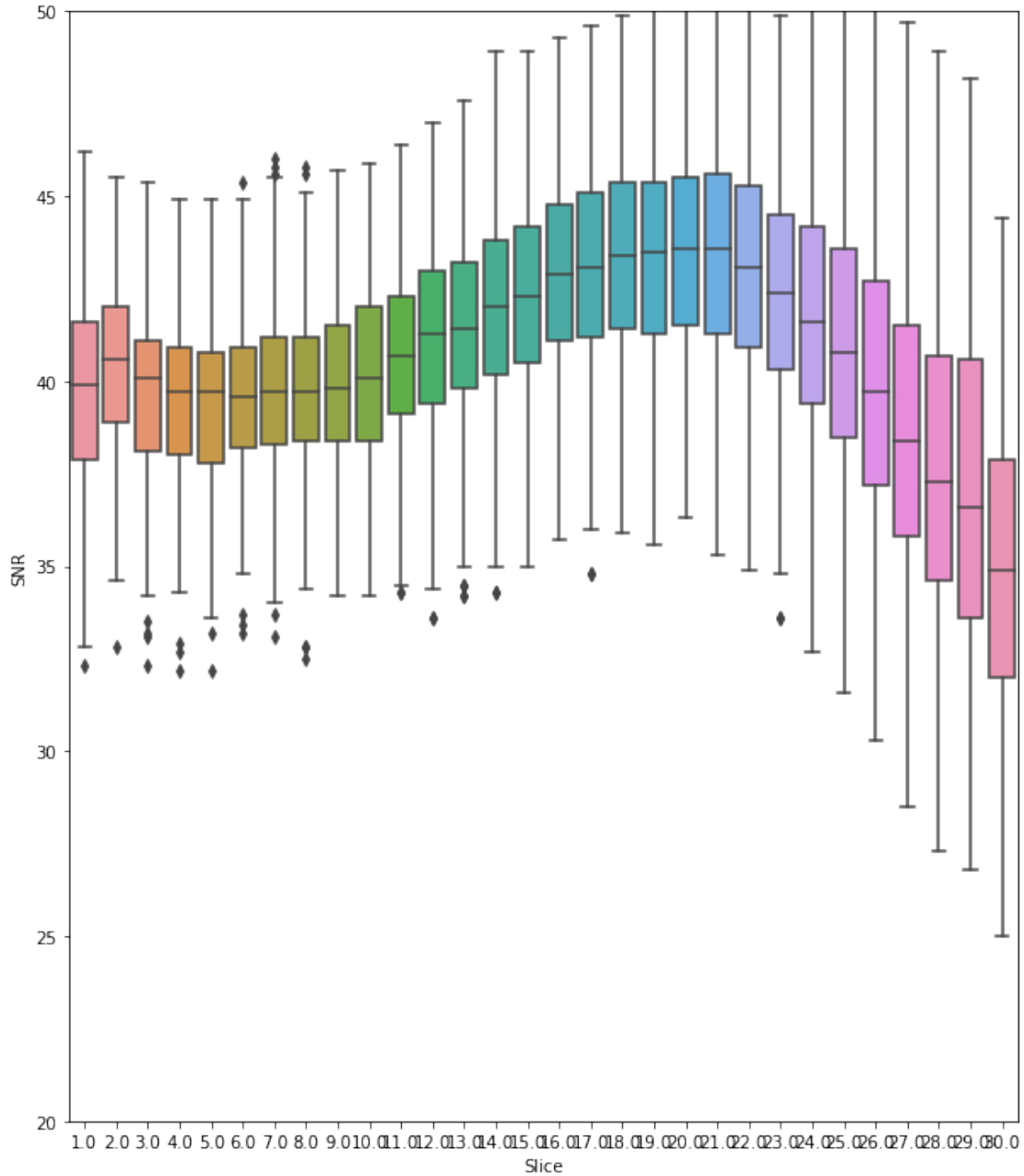
First we look at elasticity and test-retest results using the SDTD pipeline order, to make sure the values are comparable. We plot HELM:

Elasticity results for CVK test-retest cohort



Results are very similar to the 3D denoise, though the range on top extends slightly higher (more noise-robustness?).

Out [605]: [(20, 50)]

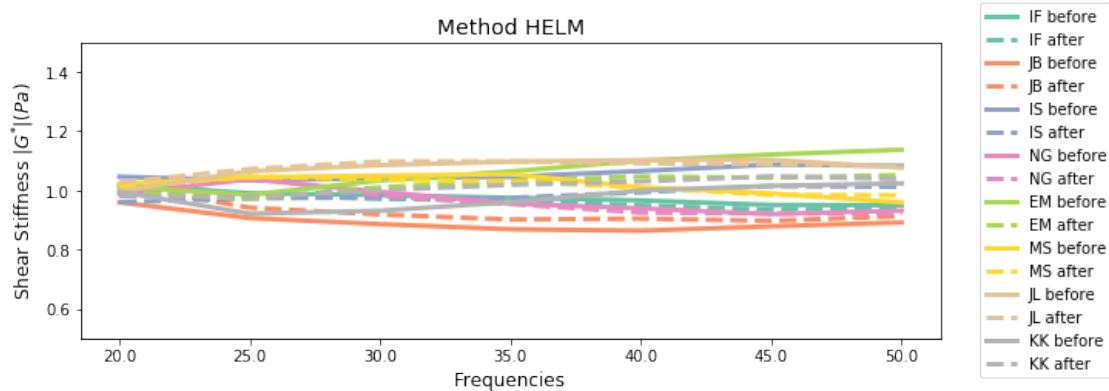


This shows a much improved snr on the lower end of 35-45 as opposed to the 28-45 above. This explains the higher values at higher frequencies; it is indeed because the 2D denoising appears more noise-robust. Interestingly, the pattern is effectively the same, so whatever causes the shifting SNR is not effected by this improvement. Consequently the 2D followed by dejitter is both faster and gives improved SNR, and this should be adopted in the pipeline.

9 Does inter-subject variability depend on frequency?

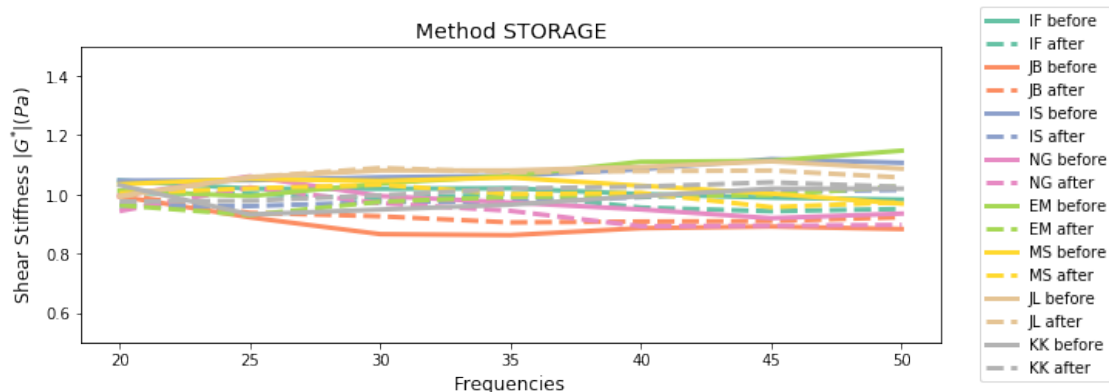
Here we see if there is greater inter-subject variability by frequency, by normalizing each value of each frequency to the mean of the values obtained at that frequency. Note: shear modulus values on the left are normalized.

Elasticity results for CVK test-retest cohort



This suggests that starting from approximately 30Hz, inter-subject variability does not change appreciably by frequency. However at the very low frequencies, particularly 20Hz there does seem to be a tighter grouping. Recalling that $|G^*|$ is a combination of elasticity and viscosity, one possibility is that viscosity changes are driving most of the inter-individual differences. We can check the storage modulus to see if this idea is supported:

Elasticity results for CVK test-retest cohort



Results are almost exactly the same. So either this isn't it or storage is not a good proxy here.