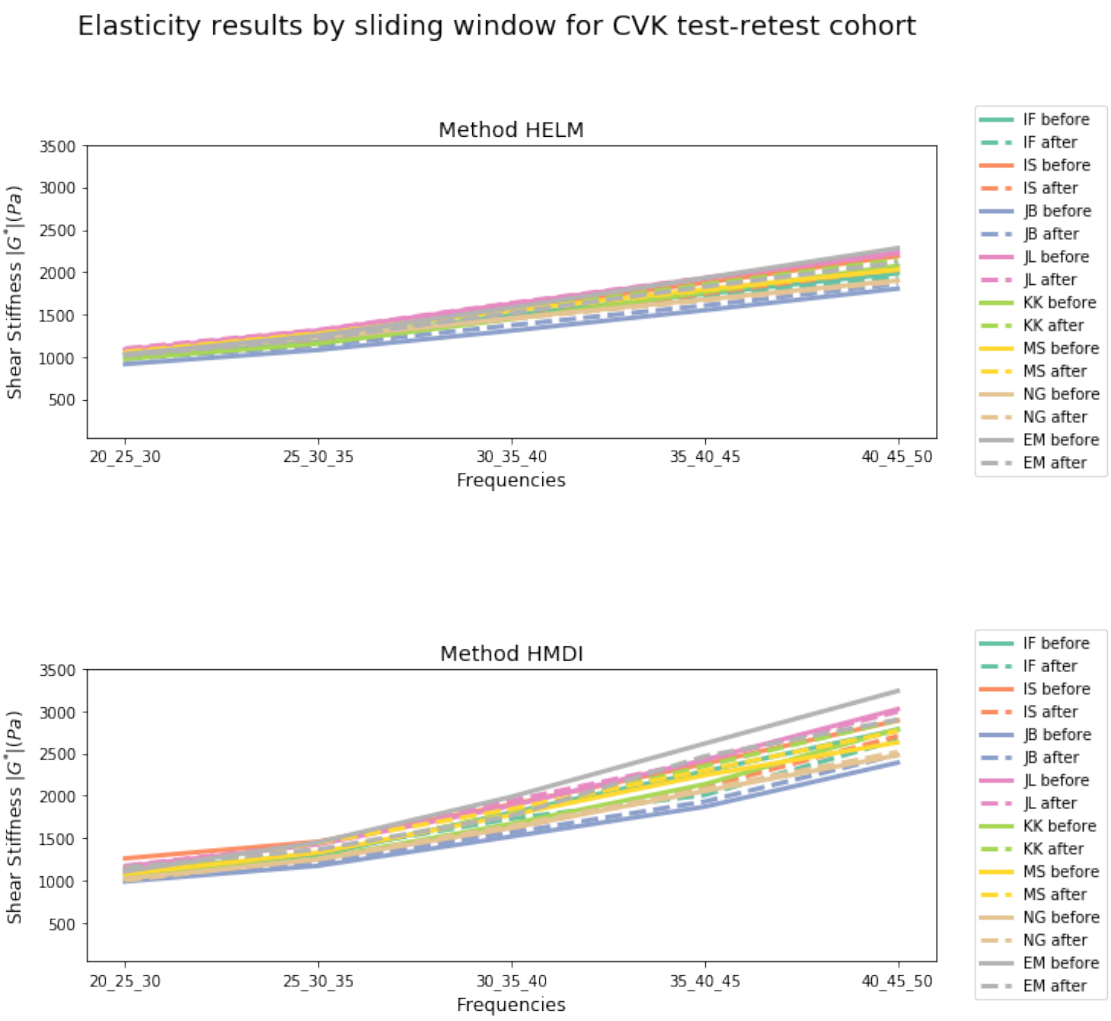
CVK Test-Retest Investigations

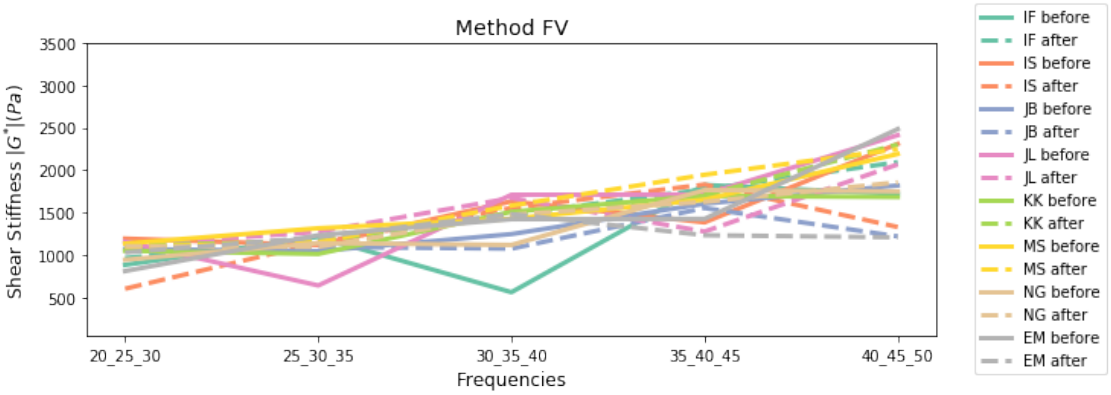
May 2, 2018

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.



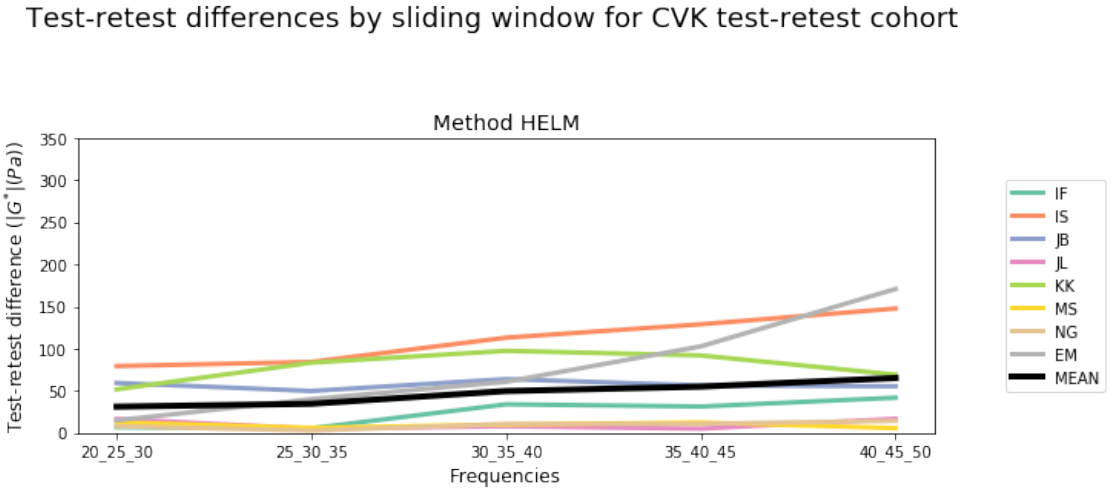


Like with the HMDI / four-groups paper, there is good consistency across frequencies for each subject (stiff subjects stay stiff, soft subjects stay soft) supporting our findings that the range is from natural variation, and our methods are robust.

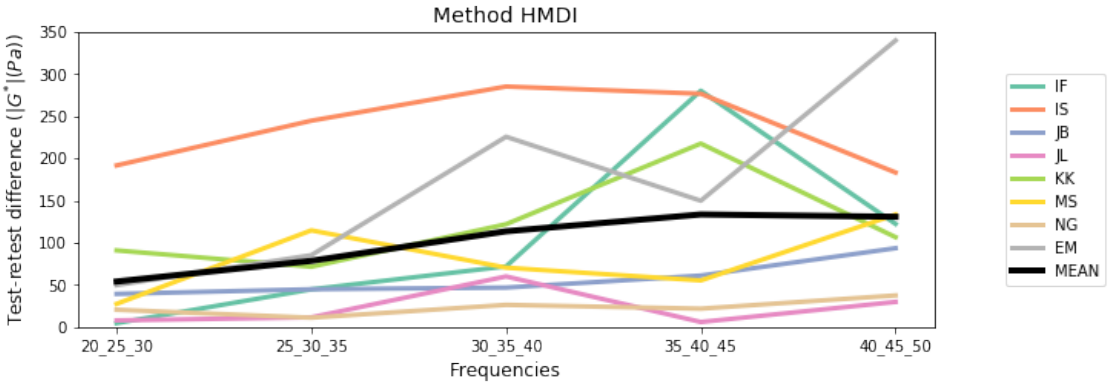
HMDI and HELM also agree very clearly on who is stiff and who is soft. They also agree on values at 25 Hz. However, as the frequencies go up HMDI gets much higher than HELM. Again, all the denoising is the same: so is HMDI just more noise-robust?

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

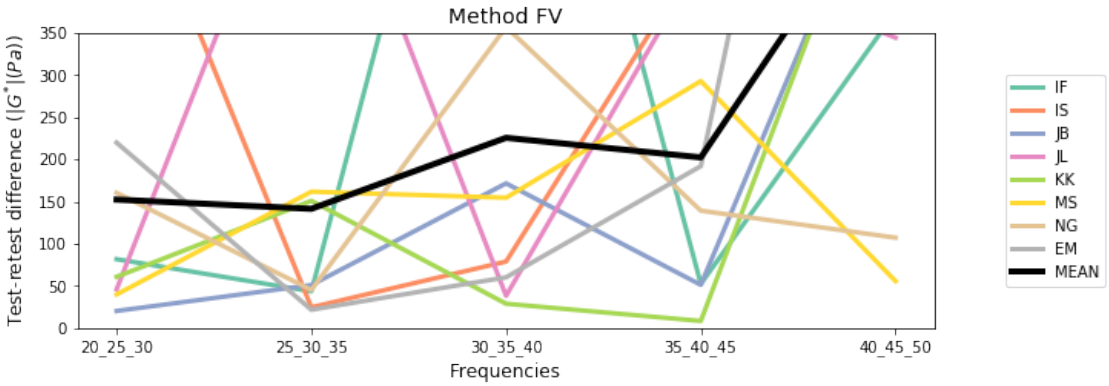
```
Mean line:  [31.4893875 34.94275   50.00735   55.2705125 65.5275625]
```




```
Mean line:  [ 54.1678375  78.69165   113.696725  133.6924625 130.9328875]
```




```
Mean line:  [152.4776    141.523775  225.7565125 202.4733625 544.3749125]
```



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

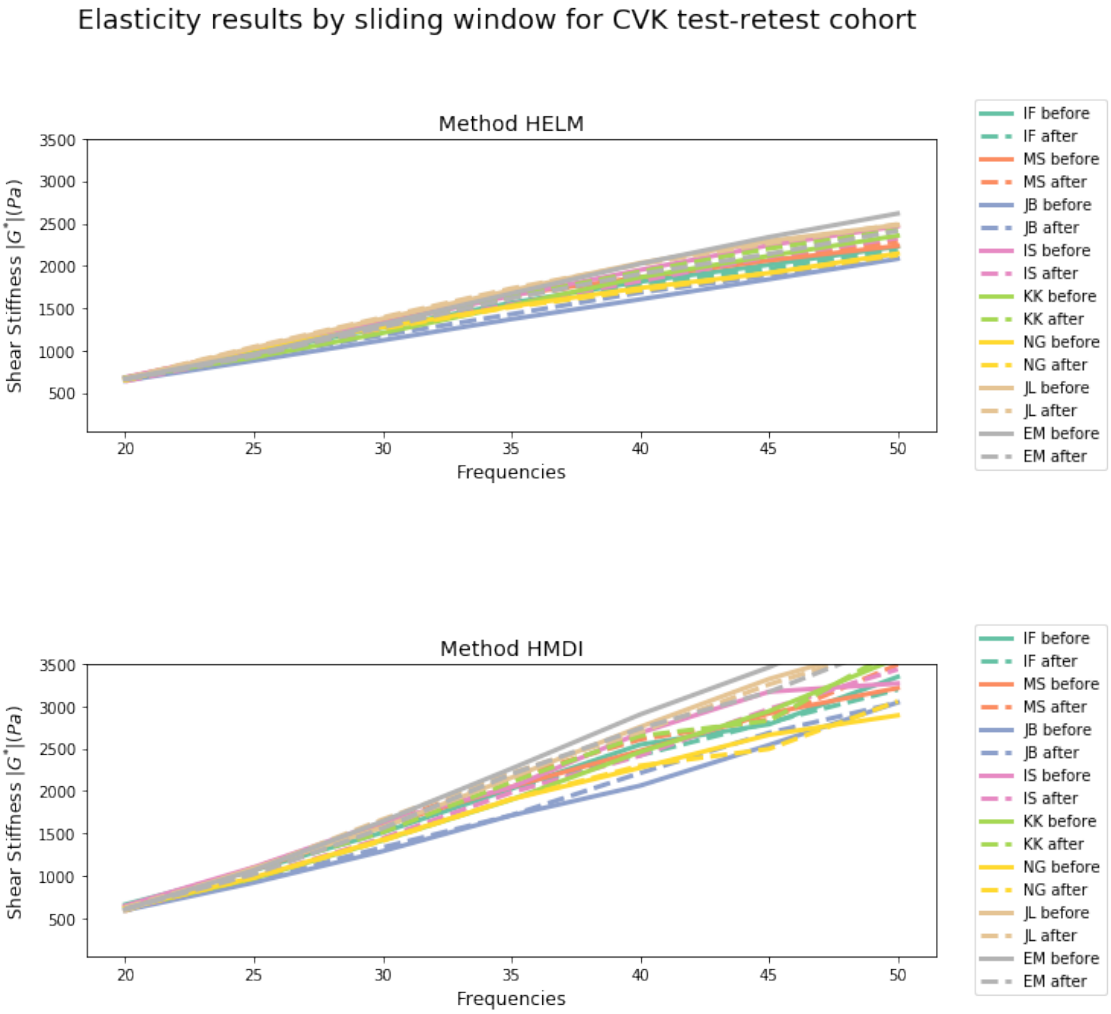
3 main observations:

- 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 zigzagging. 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 accquisition 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.)

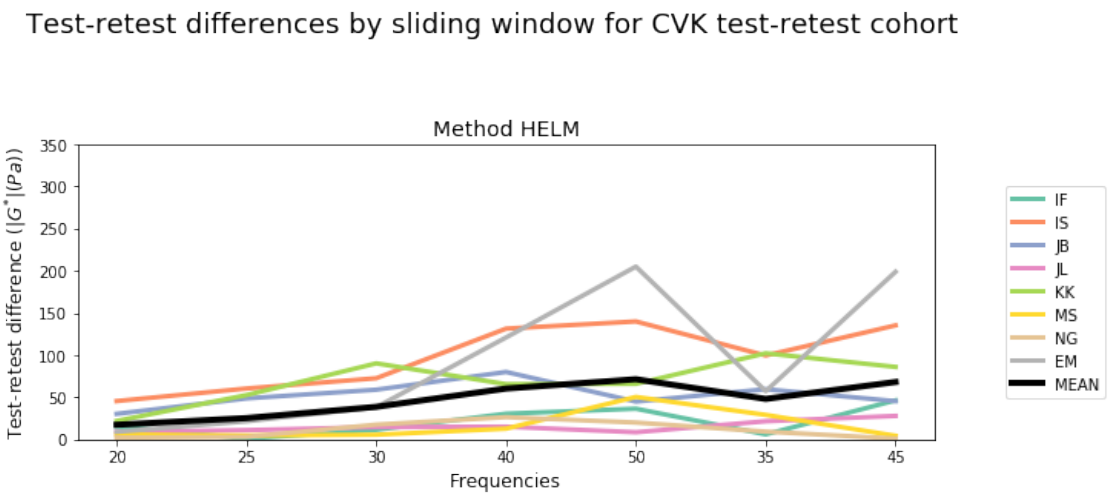
2 Wideband individual frequencies

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:

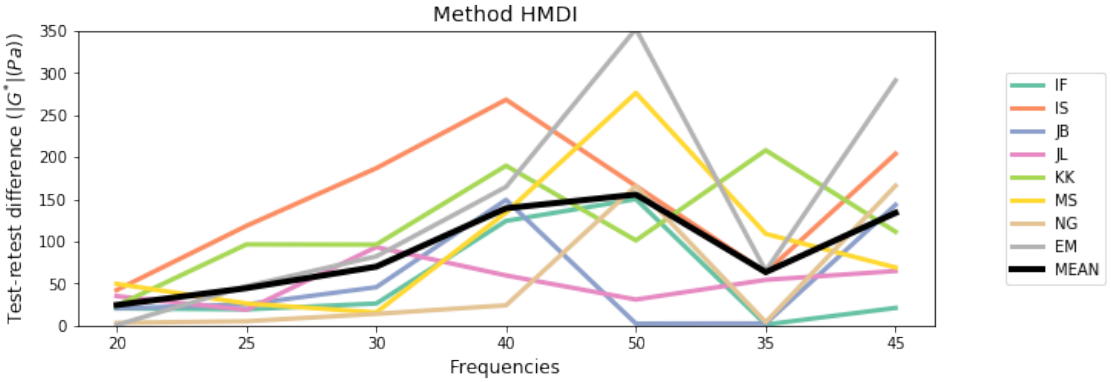


Now for the key question: is the test-retest reliability more dependent on frequency or on lateness in the cycle? Let's first look at test-retest reliability in the order slices were acquired:

```
Mean line:  [17.51105   25.4372125 38.751025  60.4159625 71.4518625 48.1387625
 68.187625 ]
```

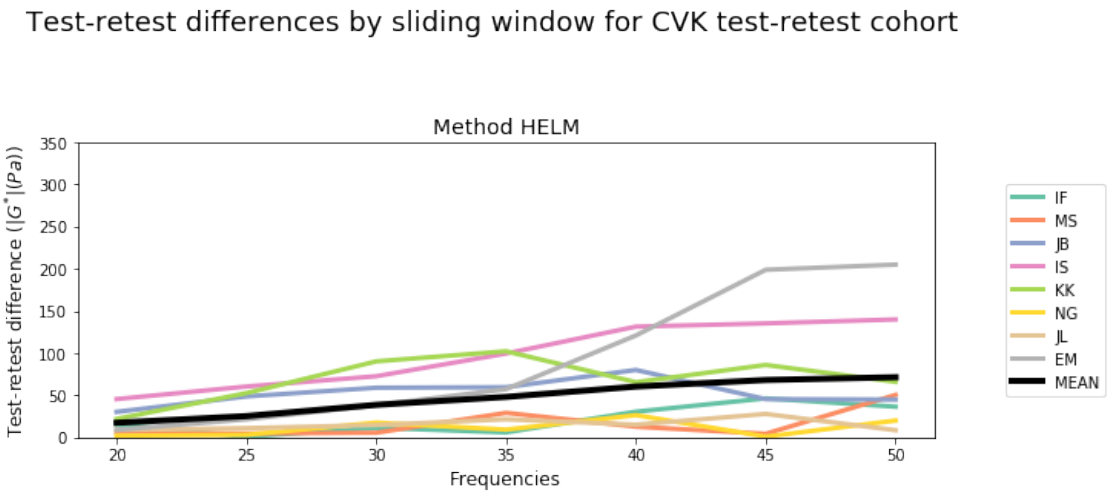



```
Mean line:  [ 24.4556625  44.5880875  70.0370375 139.3377375 155.6334375  63.5135375
 133.8909   ]
```

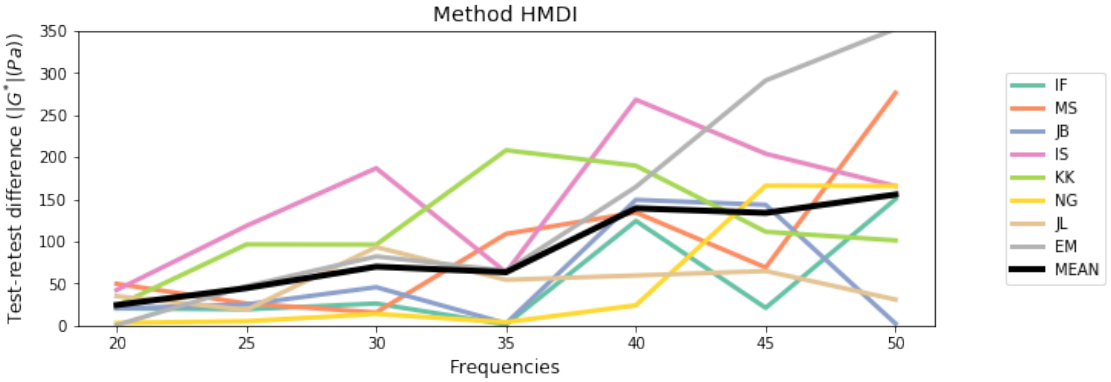


It is looking like a clear spike at 50 Hz despite it being acquired earlier. Now lets look in frequency order:

```
Mean line:  [17.51105    25.4372125 38.751025  48.1387625 60.4159625 68.187625
 71.4518625]
```




```
Mean line:  [ 24.4556625  44.5880875  70.0370375  63.5135375 139.3377375 133.8909
 155.6334375]
```



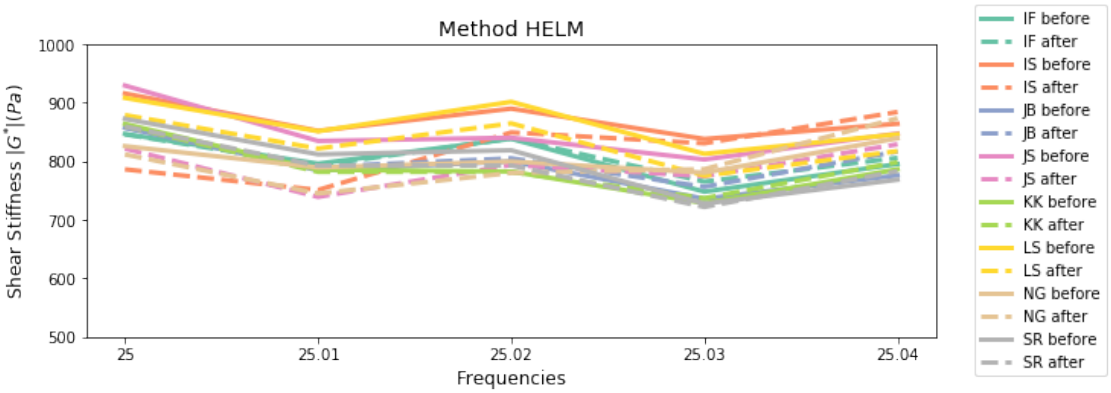
The mean line in the HELM acquisitions is now essentially linear. The wideband data thus strongly suggest that test-retest accuracy is dependent on *frequency* rather than *placement* in the acquisition sequence.

3 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.

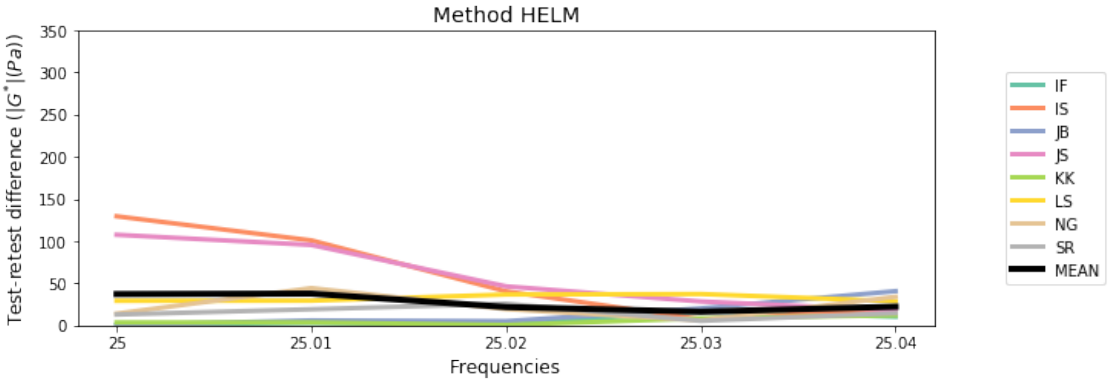
In this case there is no meaning to the sliding windows or to the sorting of frequencies, so we just look at the data set in the acquired order, and stick with HELM:

Elasticity results by sliding window for CVK test-retest cohort




```
Mean line:  [37.272625  37.8173875 21.9534    16.3853    22.3956875]
```

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.

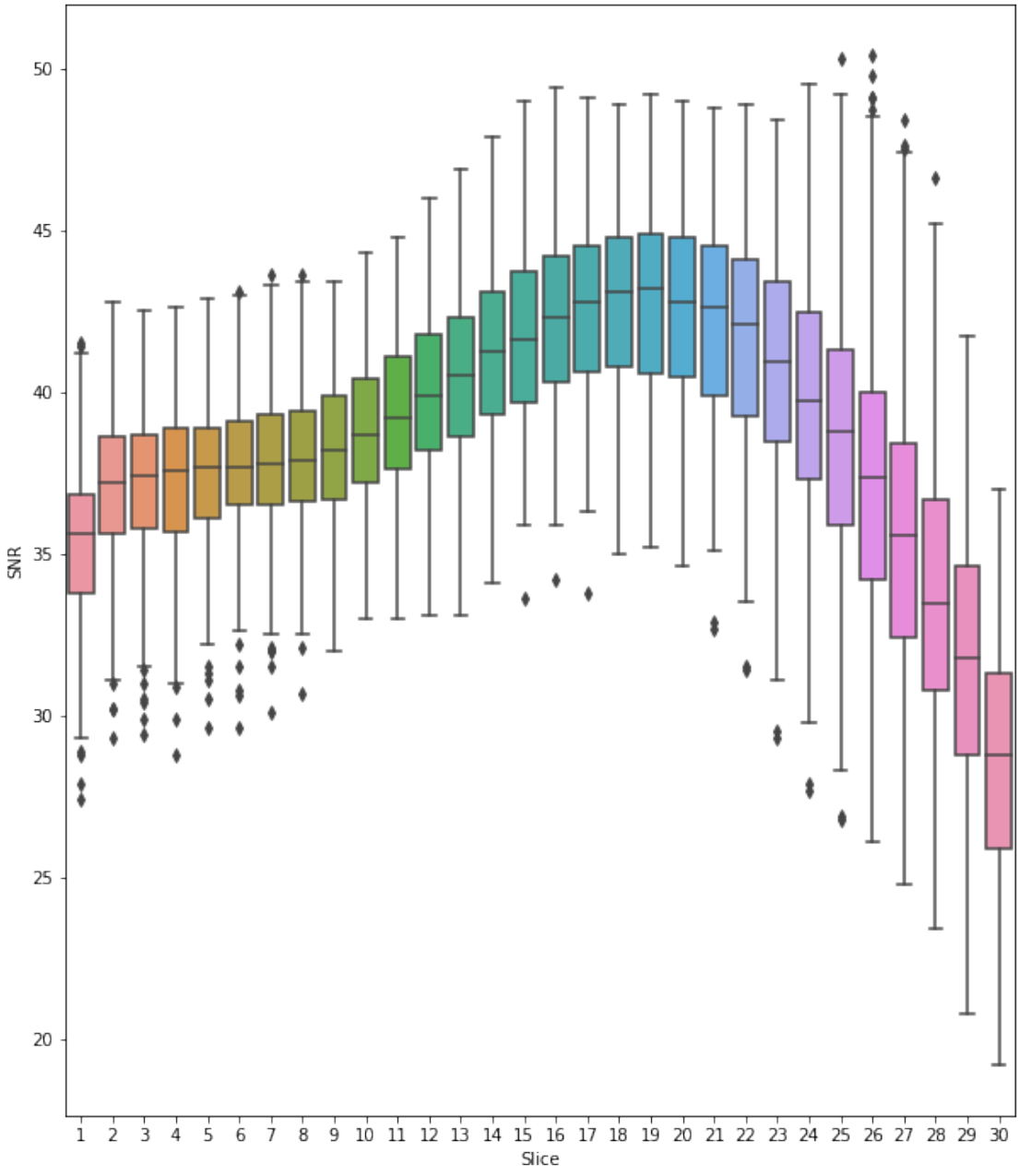
4 Slicewise SNR for the acquisitions

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.

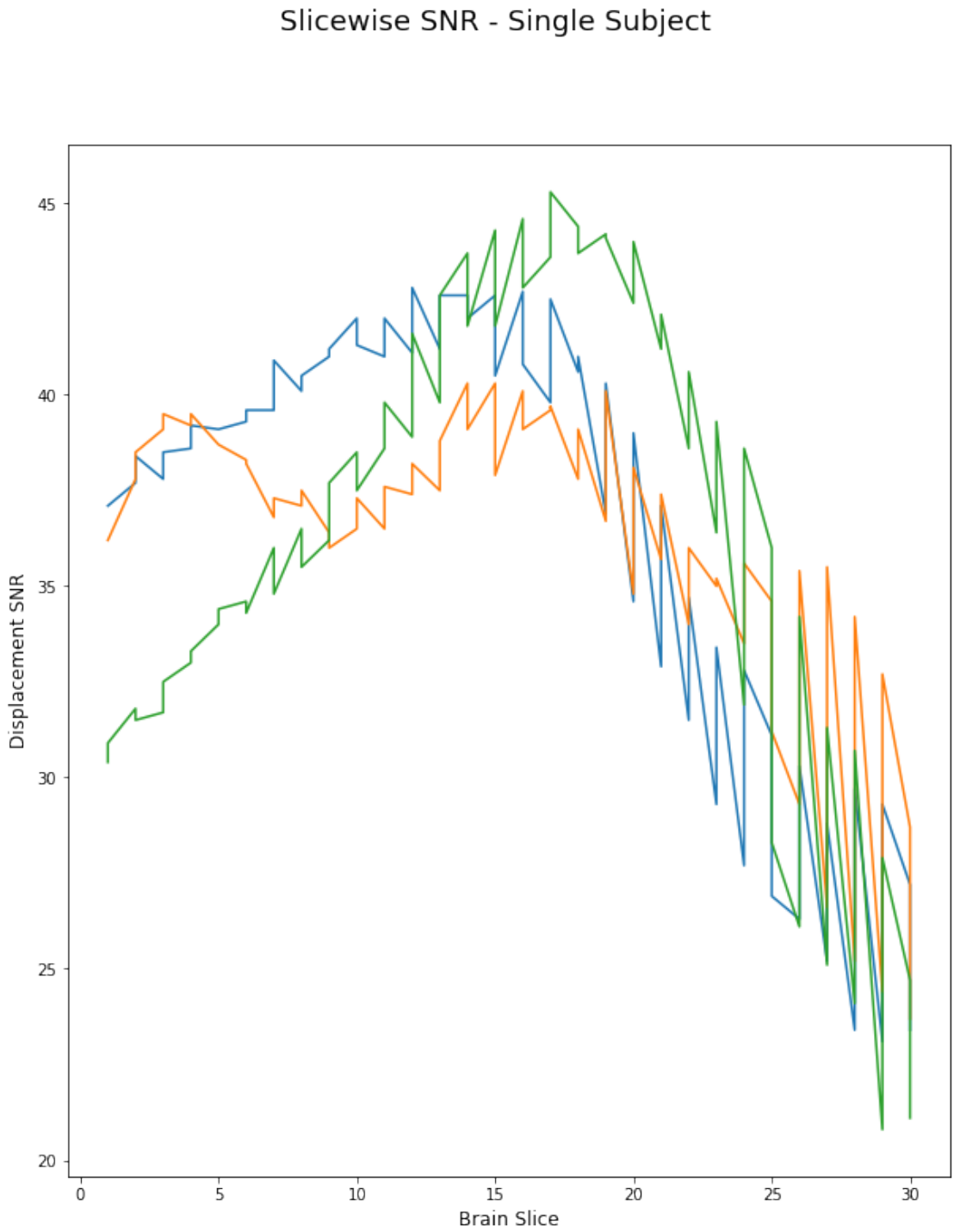
Below is a boxplot, showing SNR by slice for all acquisitions and frequencies lumped together. There appears to be a smooth overall trend, however with substantial errorbars.

```
Index(['Slice', 'SNR', 'Subj_ID', 'Subj', 'Freq', 'Comp'], dtype='object')
```

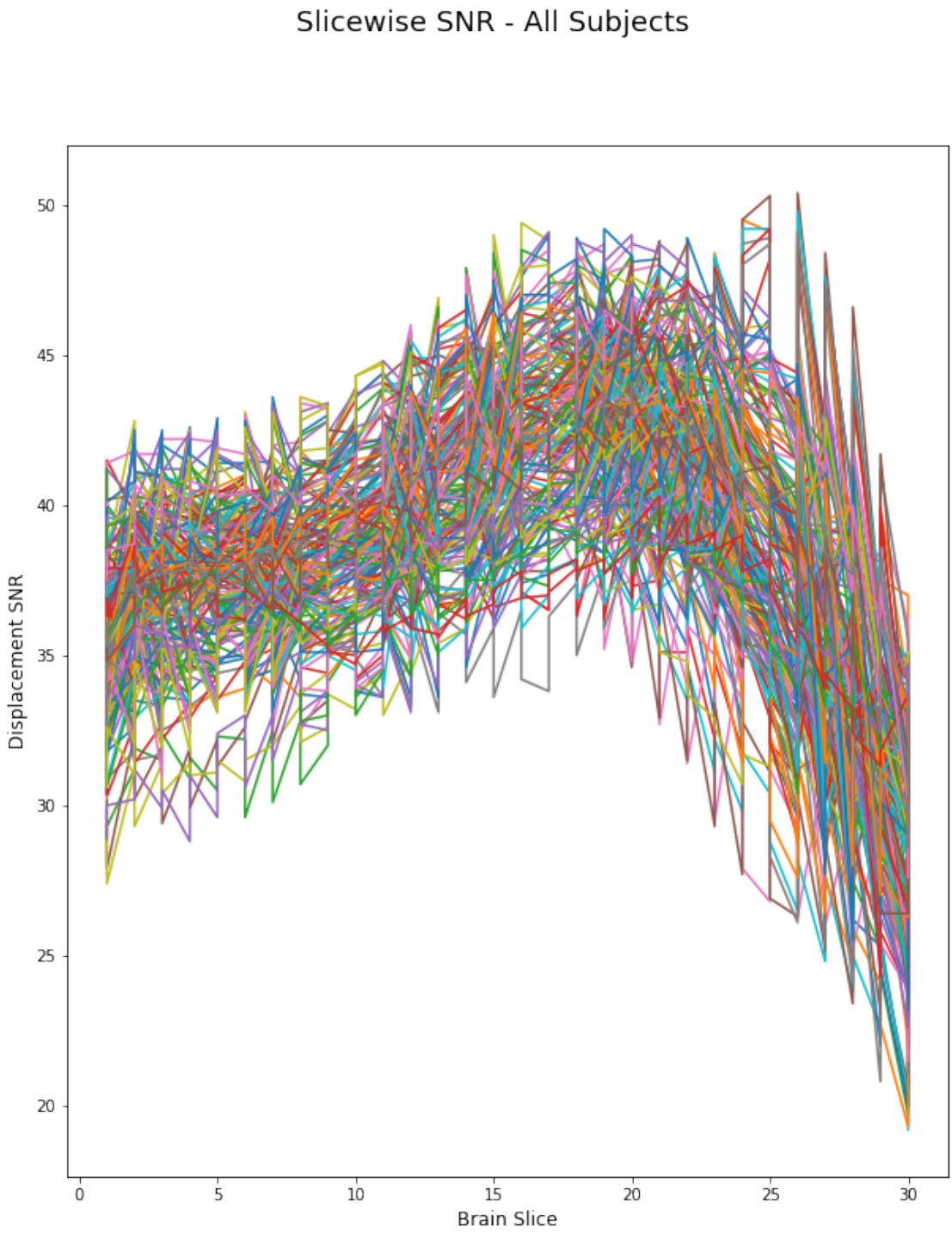
```
Out[1045]: <matplotlib.axes._subplots.AxesSubplot at 0x7f475b5796d8>
```



So we next look at all the lines plotted to get a sense of the individual trends. Lines are colored by subject and stroke is by frequency, so see if any dependencies on either factor can be seen. First plot is for a single subject, second is for all subjects.



```
<matplotlib.figure.Figure at 0x7f472e17a198>
```



```
<matplotlib.figure.Figure at 0x7f4726932e48>
```

So, the pattern appears to be consistently followed and the bar chart is sufficient to show the data. The higher bars at the higher slice numbers reflect increasing instability in SNR calculations, this is probably due to the decreasing amount of cortex voxels available for the calculation, but the pattern appears to still be there.