

Product Biomass Samples IR Batch Spectroscopy

[Tags >](#)

[REPLACER](#)

Sample Production

- **rehydrate** a crumb of **product biomass** w/ 10x (by mass) water
- thick CaF₂ window on bottom
- steel spacer ring
- pipette 2µl rehydrated biomass into the middle
- close sandwich with thin CaF₂ window on top

Measuring

Copropagation, 40x Cassegrain

Auto-Background before 1st measurement

Averaging: 3

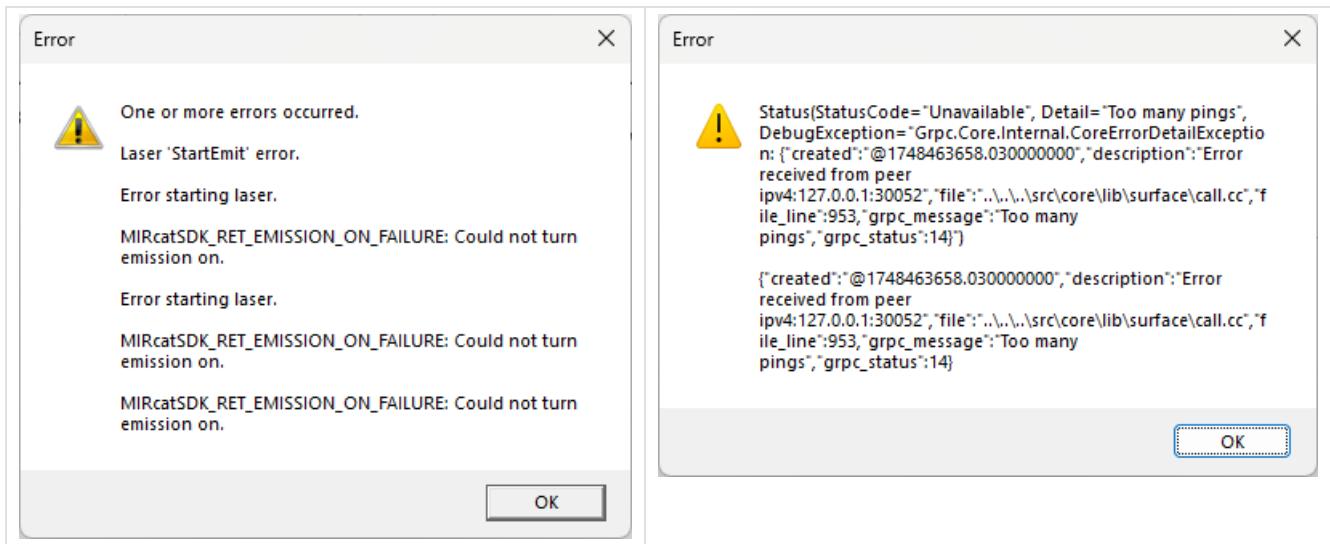
Recorded Channels were

0. **OPTIR**: The amplitude of the PT signal, corrected for the background determined by the mIRage.

1. **Phase**: Phase of the PT signal
2. **X**: The real part (or in-phase component) of the PT signal
3. **Y**: The imaginary part (or quarter-phase delayed component) of the PT signal

Sample	Labelled	Focus [µm]	Detector Gain
1	0,1476	2660.7	10x
2	illegible (cap blue, paper sticker: 0.01.25)	2484.5	50x
3	30.01.25 A. platensis 0,15 g	2649.9	50x
4	0,1653 g	2592.0	50x
5	98,6 mg	2657.2	50x
6	03.02.25 A. platensis 0,19 g	2569.8	50x
7	A. platensis	2641.9	50x
8	A. platensis 0,20 g 26.01.25	2578.8	50x
9	A. platensis 0,23 g		
10			
11			
12			

After production of the ninth sample, a crash of the setup was noticed:



The crash was unfortunately determined to be irrecoverable. No further measurements could be made.

Results

Inspection

For each sample, autofluorescence micrographs were recorded to help select regions for spectroscopy.

Sample 1: "0,1476"	Sample 2: "0.01.25"	Sample 3: "30.01.25"	Sample 4: "0,1653 g"
Sample 5: "98,6 mg"	Sample 6: "0,19 g"	Sample 7: "A. platensis"	Sample 8: "26.01.25"

Spectra

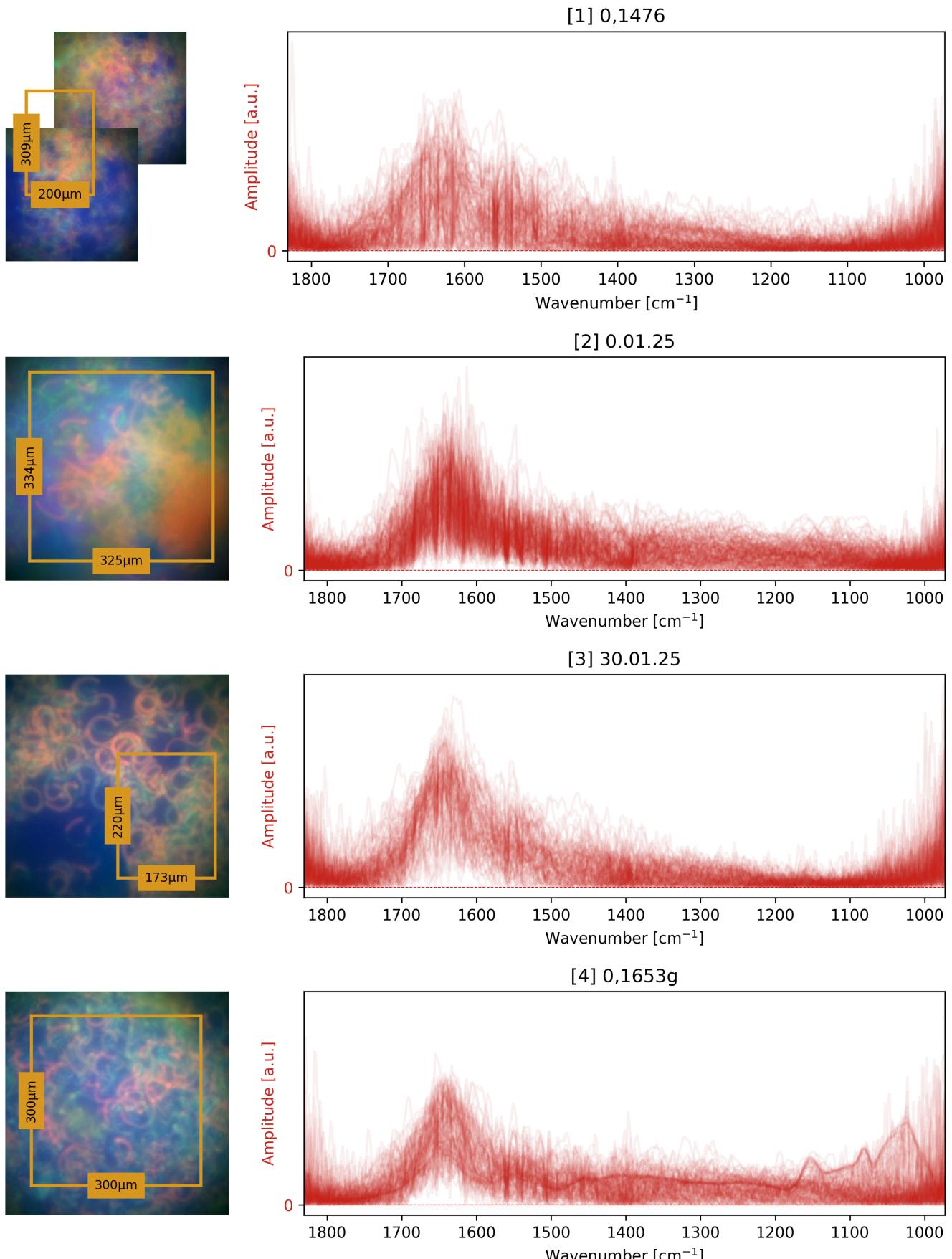
Definitions

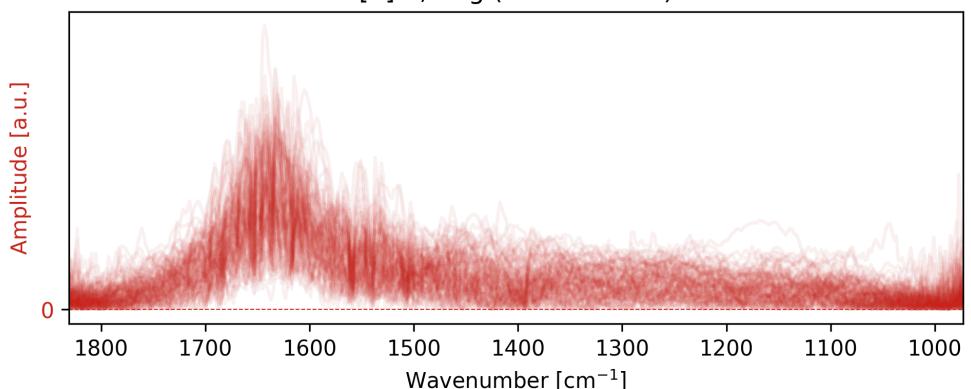
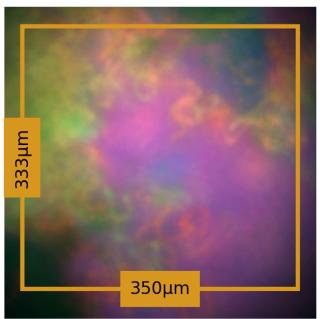
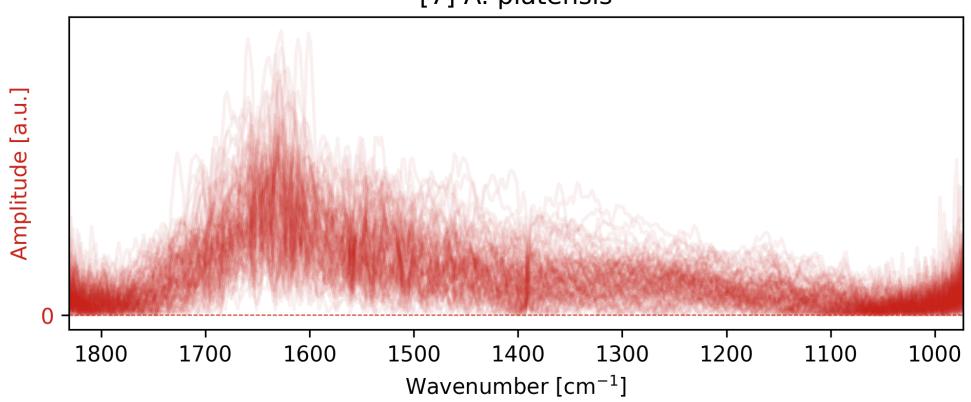
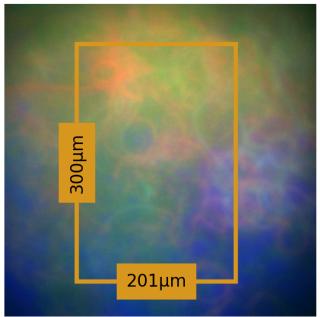
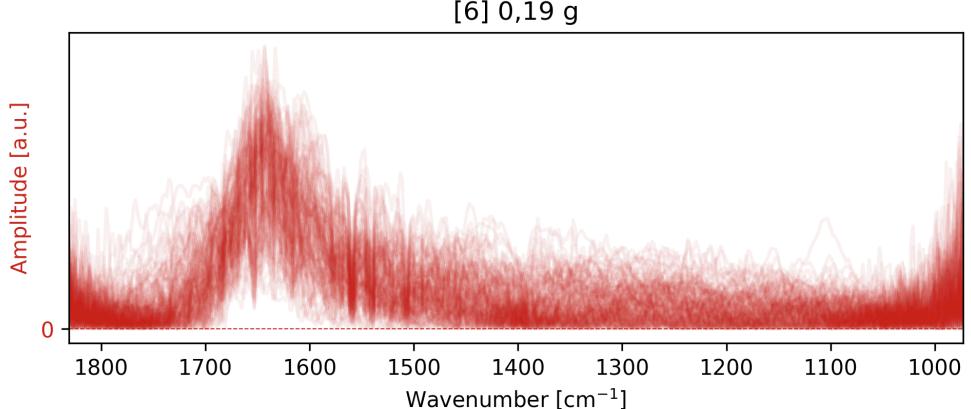
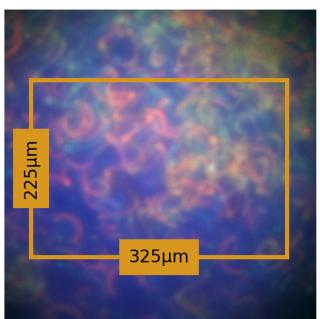
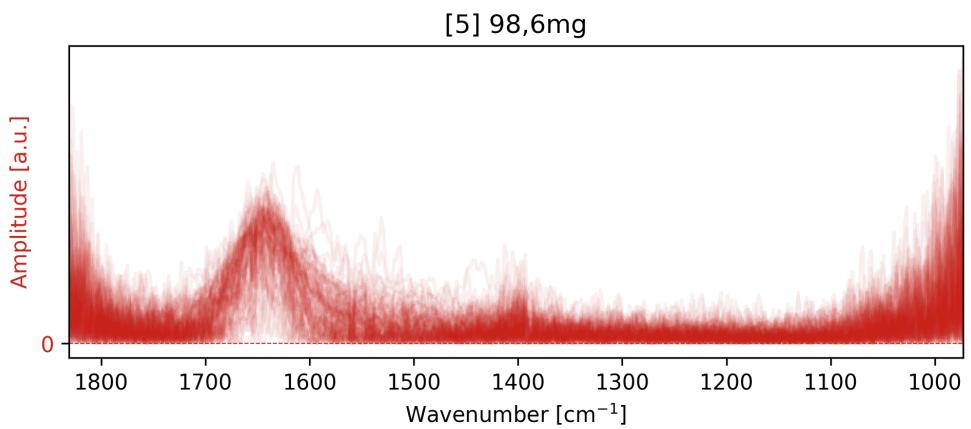
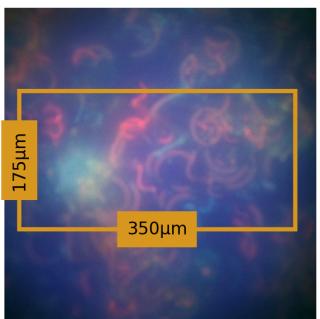
We assume that the "true" photothermal signal is encoded by $X + iY$. We thus calculate the **amplitude** of the signal as $\sqrt{X^2 + Y^2}$ and the **phase** as $\text{atan2}(Y, X)$, instead of taking the phase as recorded by the software. We do this, because the phase of the average signal is not the same as the average phase of the signal:

$$\text{atan2}(\langle Y_i \rangle, \langle X_i \rangle) \neq \langle \text{atan2}(Y_i, X_i) \rangle$$

Individual Spectra

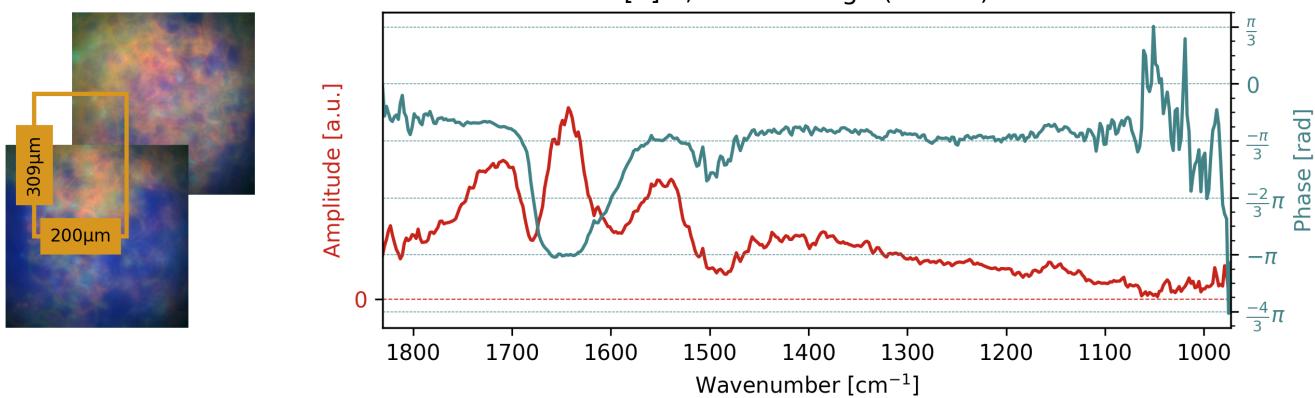
...were very noisy. Shown below are the L^2 -normalized amplitude spectra from each sample. For each sample, an autofluorescence composite micrograph is given. The indicated areas are those, wherein spectra were recorded.



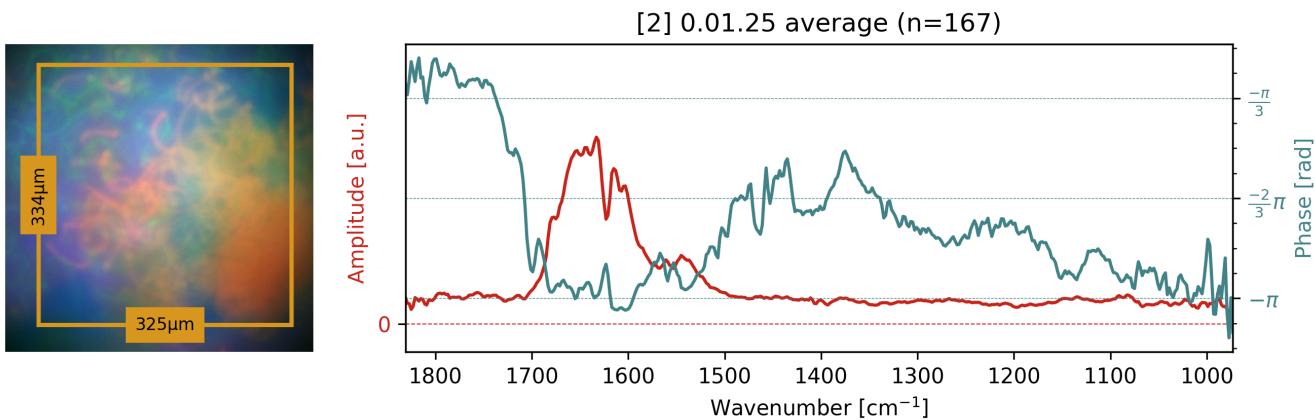


Average Spectra

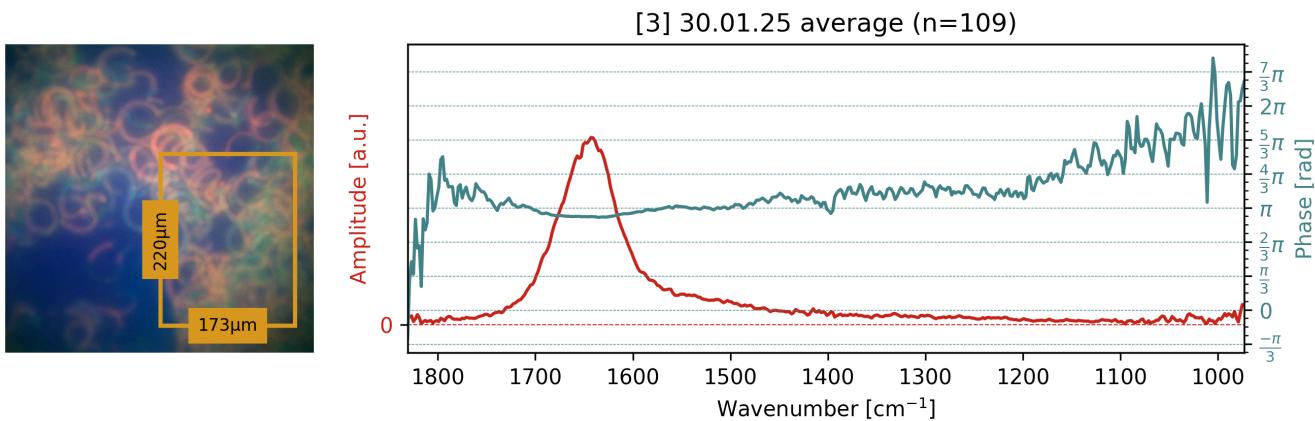
Here, average spectra are shown with amplitude and phase.



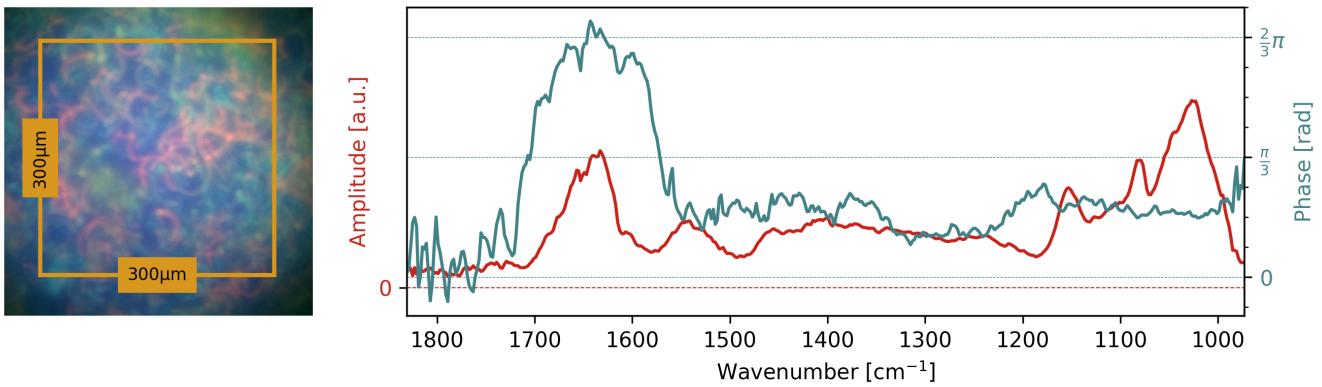
[1] Note how the main peak has a different phase from the secondary peaks. This might hint at the main peak corresponding to heating of a large volume (water background), taking a longer time for the temperature profile to relax and the secondary peaks corresponding to actual biomolecules, making up much less volume and therefore relaxing more quickly.



[2] Here, we similarly see a lagging phase in the region around 1650cm⁻¹.

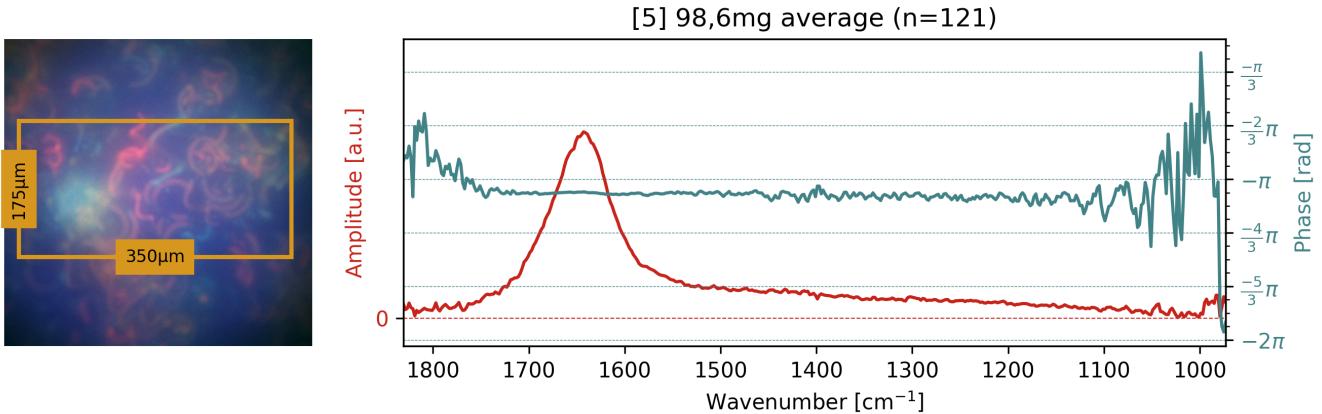


[3] Here, the phase seems to be invariant. The amplitude looks very smooth with only one visible peak and no secondaries. It doesn't look like any biomass is detected here, for some reason.

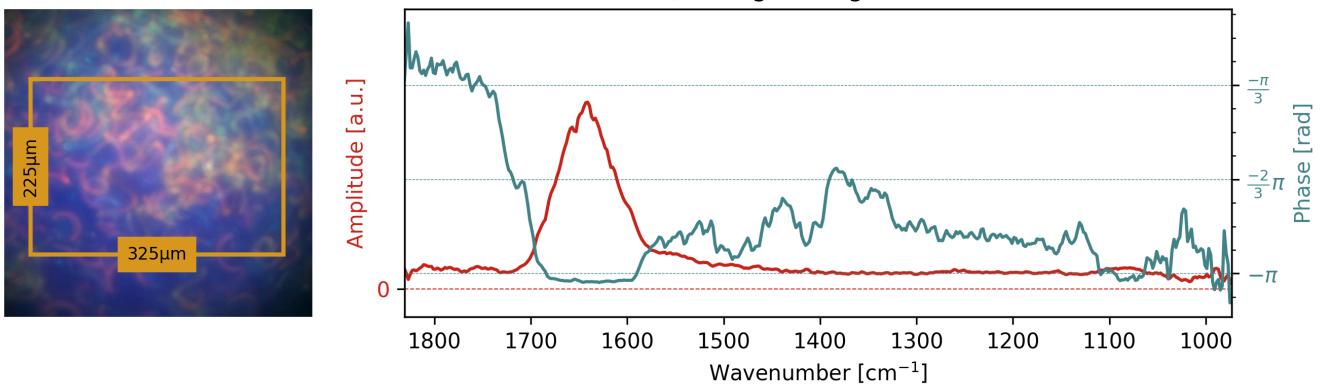


[4] Here, the phase doesn't lag around the water absorption but instead seems to advance. Reason unknown. Many additional peaks are visible in the amplitude of the signal. Their locations are consistent with usual bio-signatures, e.g.

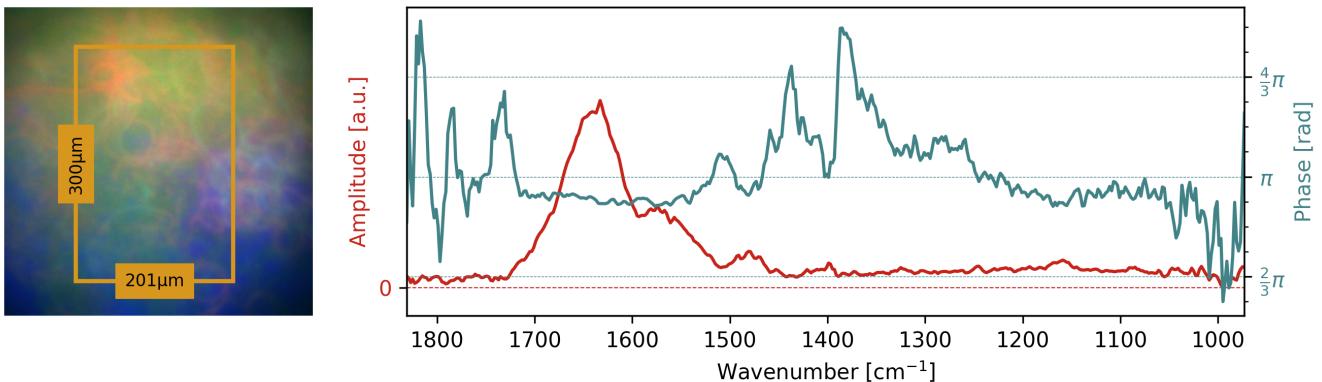
$\sim 1550 \text{ cm}^{-1}$	$\sim 1460 \text{ cm}^{-1}$	$\sim 1150 \text{ cm}^{-1}$	$\sim 1080 \text{ cm}^{-1}$	$\sim 1020 \text{ cm}^{-1}$
Amide II band	CH ₂ scissor		PO ₄ stretch	



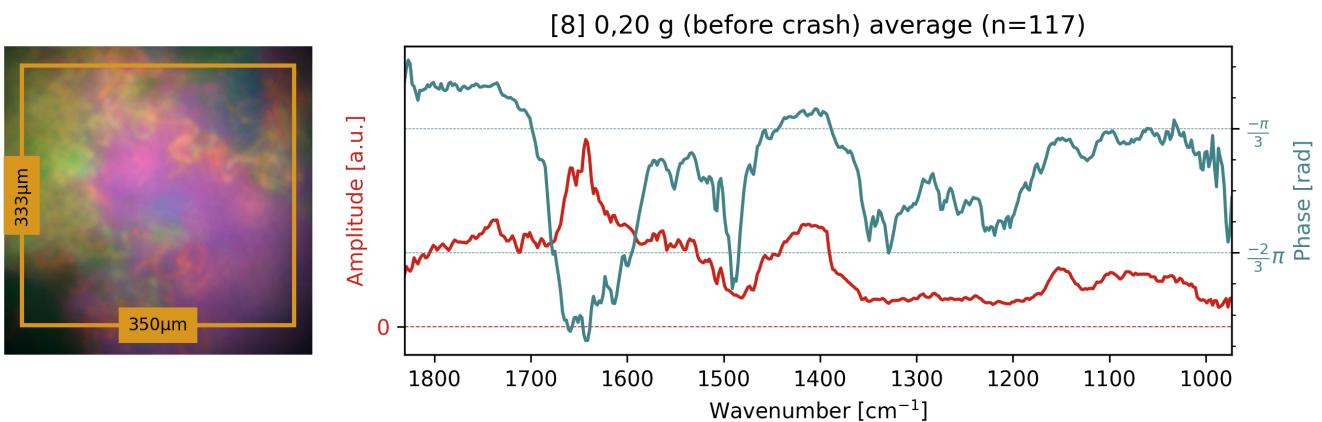
[5] as in [3], the phase is remarkably constant and no secondary peaks are visible.



[6] Secondary peaks are almost invisible. However, the phase clearly lags around the water absorption range.



[7] A peak in the high 1500s appears more like a shoulder to the side of the main peak at 1650 cm^{-1} . Smaller peaks are discernible in the amplitude at lower wavenumbers, and coincide with clear local phase lags.



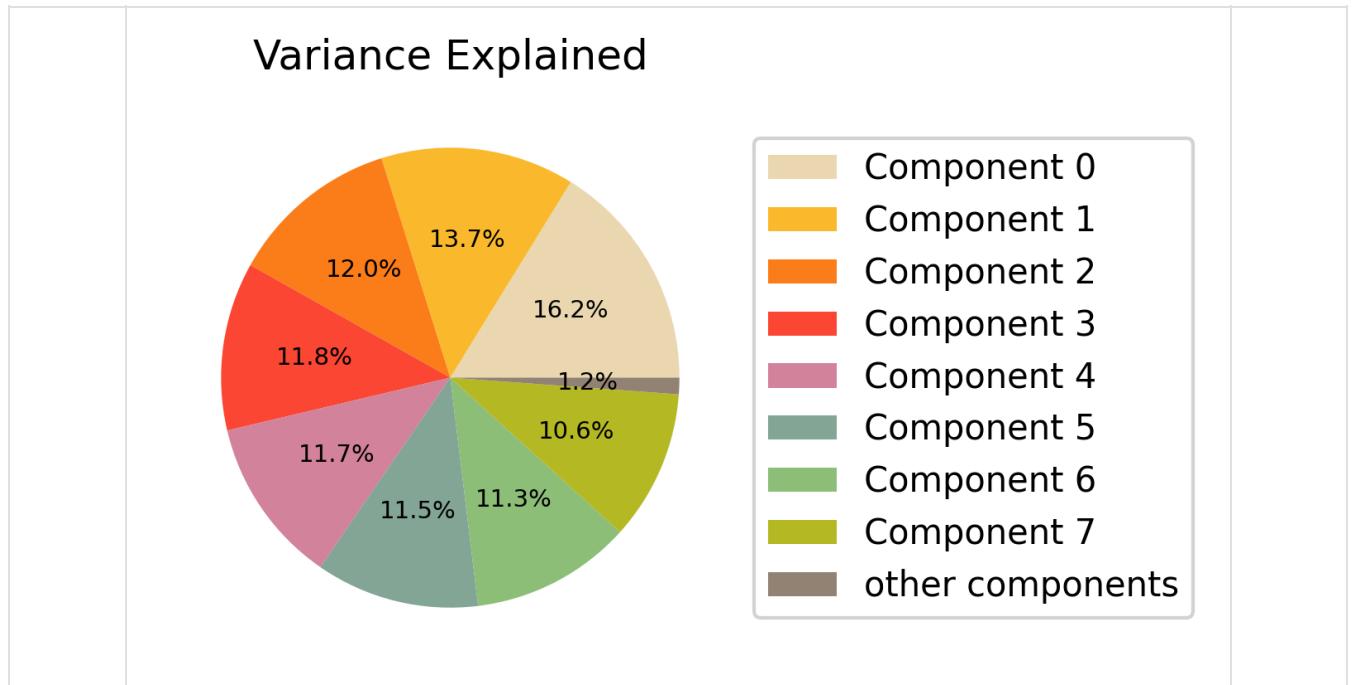
[8] Multiple peaks with irregular shapes are discernible in the amplitude. Many features (moth peaks and valleys) of the amplitude spectrum coincide with features in the phase spectrum.

Decomposition of Spectra

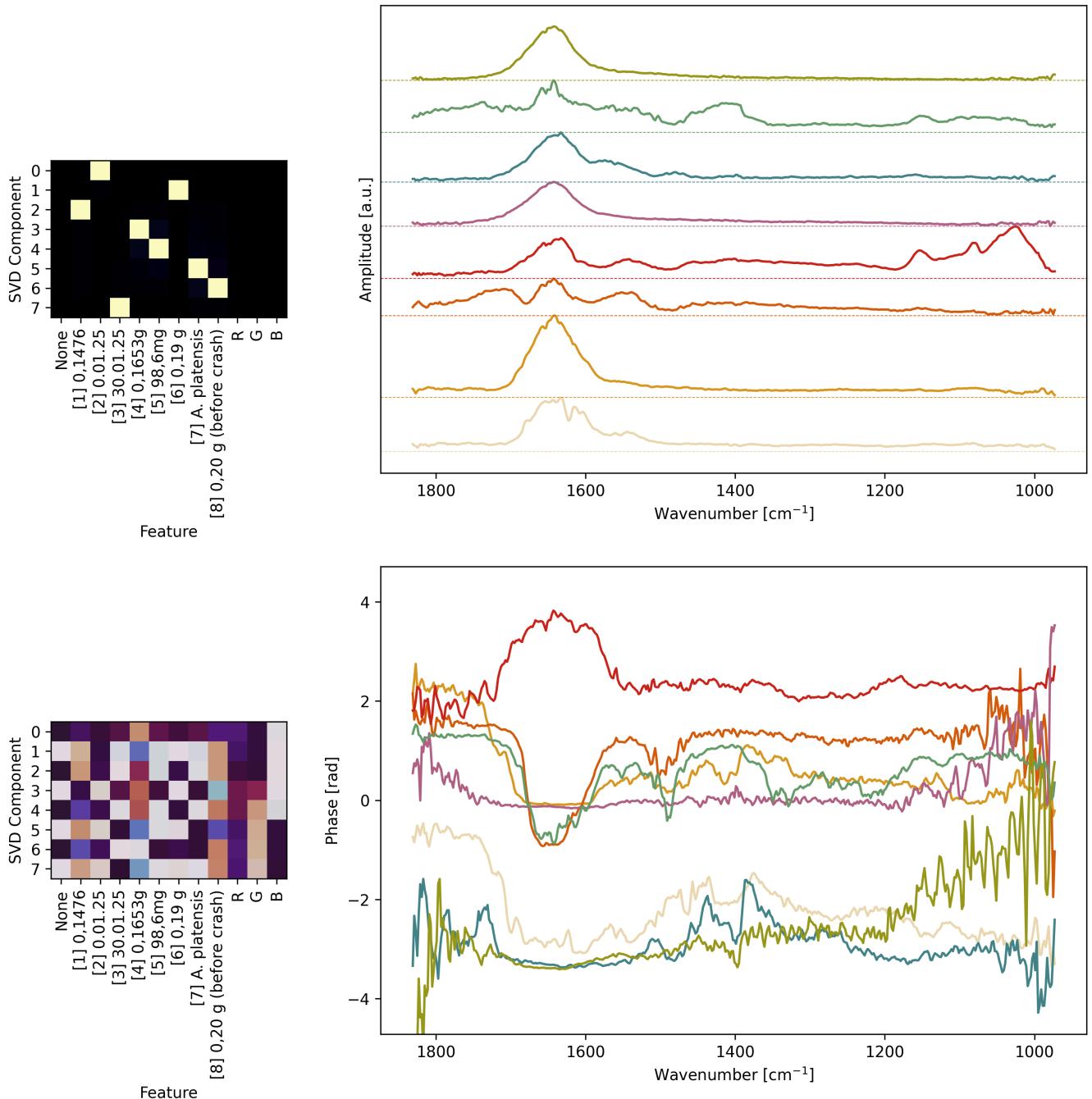
As usual, we cannot extract comprehensible information directly from the complete hyperspectral datasets and therefore must resort to some manner of dimension reduction. In this, it appears prudent to analyse the complex-valued spectra and to not disregard the phase.

Singular Value Decomposition

I performed PCA (or SVD, because `sklearn.decompose.PCA` doesn't support complex-valued input) The explained variance per SVD component looked like so:



This means that the first eight components were significant and the rest could be disregarded, essentially. Inspecting the components gave the following:

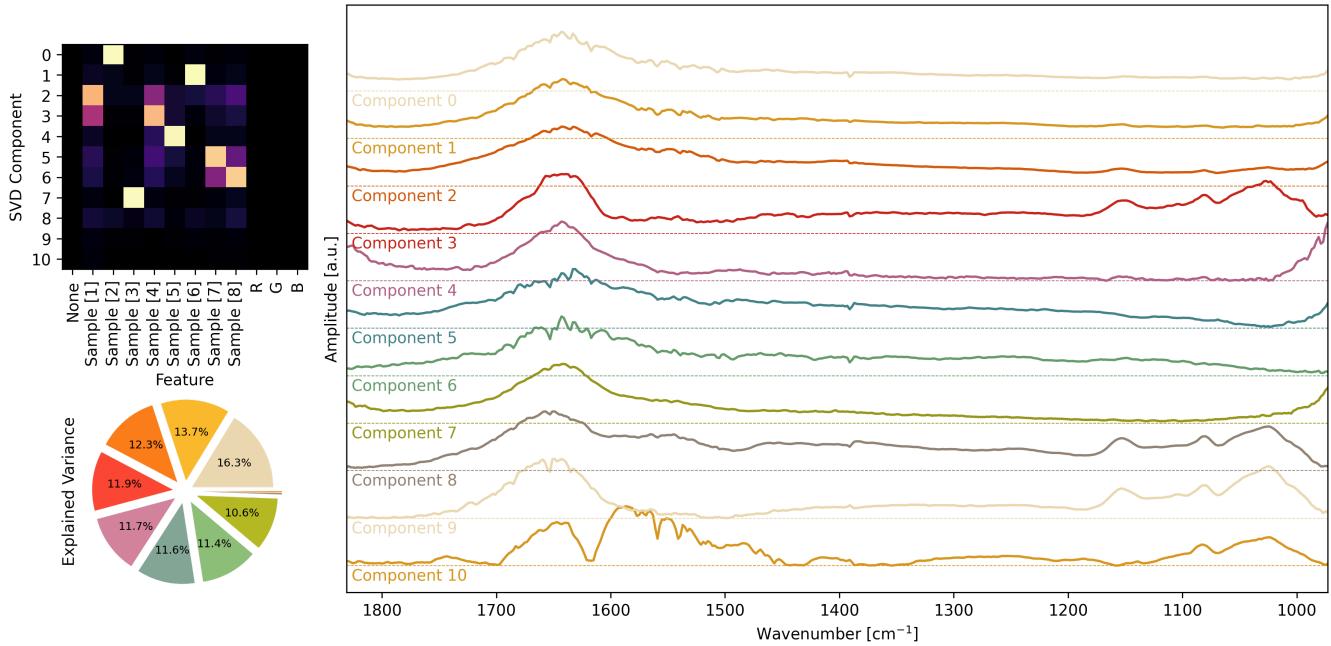


- **top left:** magnitude of the mapping between SVD components and features (sample indices)
- **top right:** magnitude of the SVD components vs wavenumbers
- **bottom left:** phase of the mapping between SVD components and features (sample indices)
- **bottom right:** phase of the SVD components vs wavenumbers

The interpretation is as straight-forward as it is disappointing:

The eight relevant SVD components were nothing but the average spectra of the eight samples.

So, counter to the earlier idea, I attempted the decomposition disregarding the phase. When running the **SVD only on the amplitude** of the signal, we unfortunately get similar results:



- **top left:** magnitude of the mapping between SVD components and features (sample indices)
- **bottom left:** explained variance per SVD component. only 0.6% of the total variance of the dataset are explained by components beyond 7.
- **right:** SVD components vs wavenumbers

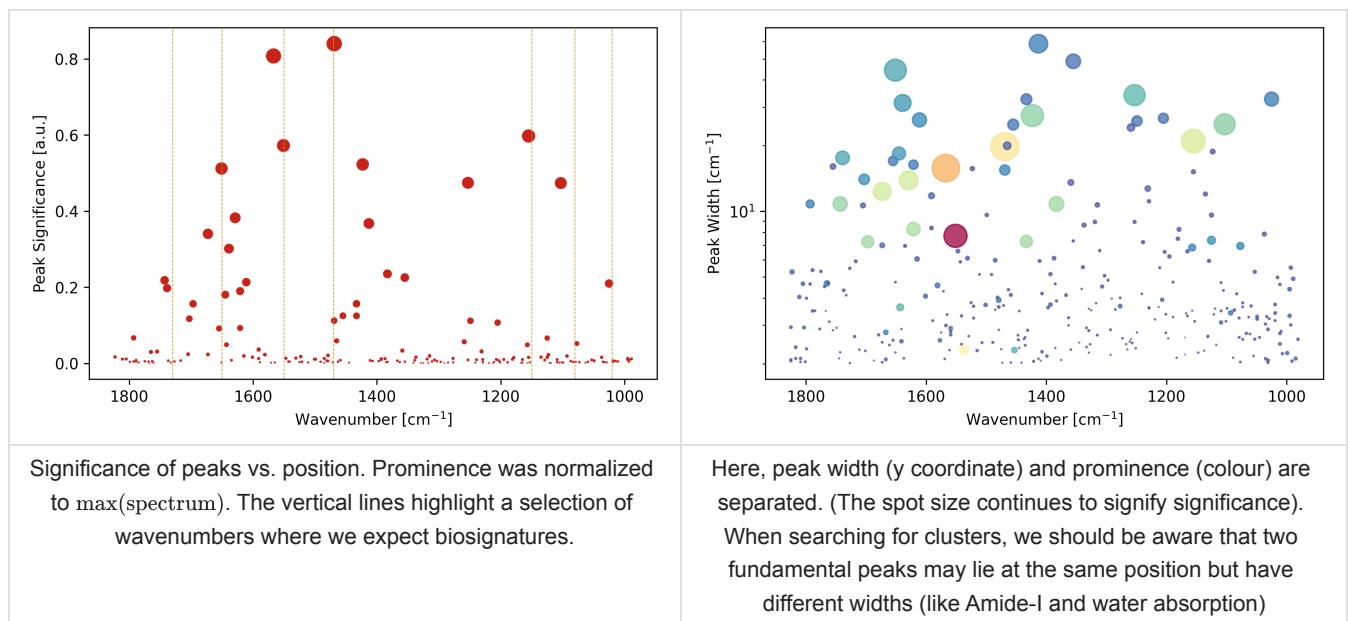
However, there is some overlap between measurements now. Still, it ought to be very clear that we should attempt to model the individual peaks and try decomposing into those.

Attempts to Isolate Spectral Peaks

Let's accumulate information about all spectral peaks using `scipy.signal.find_peaks`. Obviously important is the position (in terms of wavenumbers) of each peak, but also, we require a measure for total absorption that a peak is associated with. Given the values of "prominence" and "width" that the `find_peaks` algorithm computes, we define a peak's

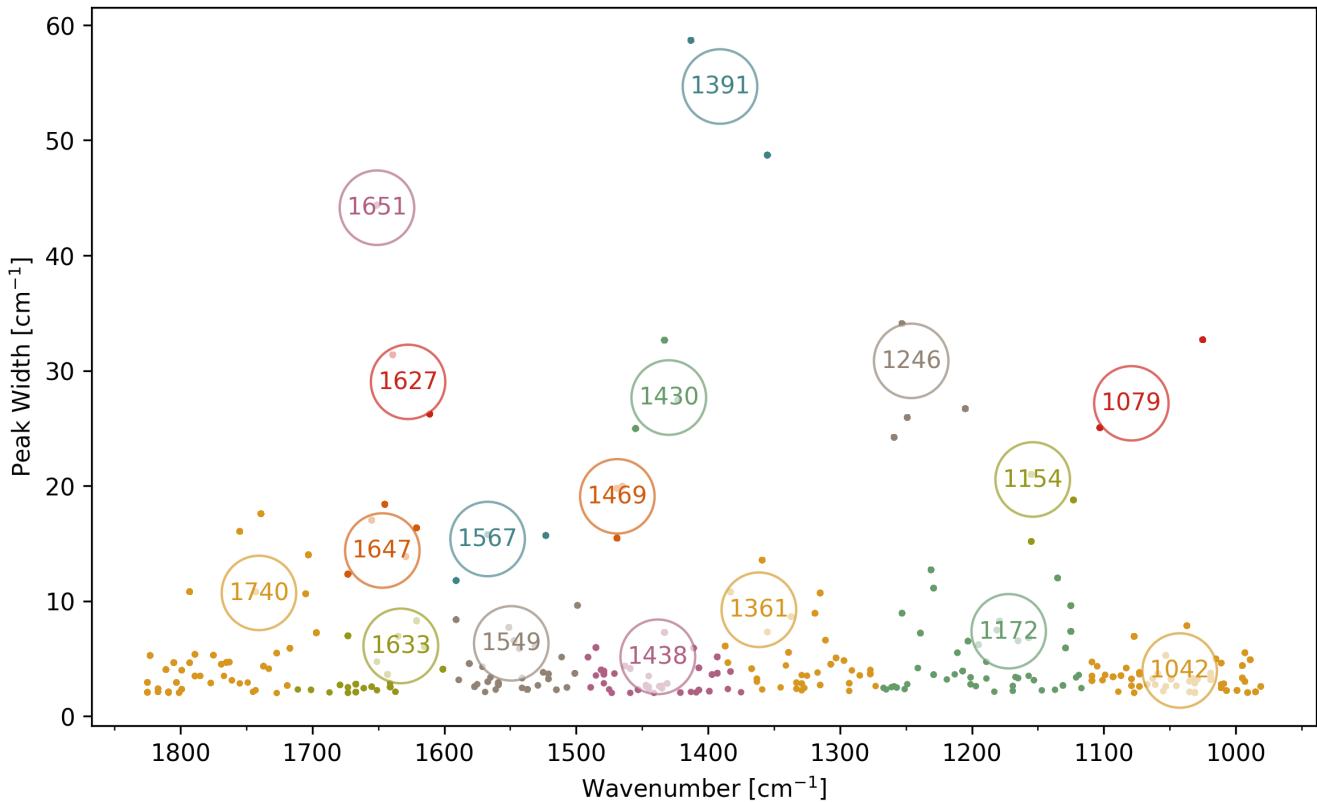
$$\text{significance} := \text{prominence} \cdot \text{width}.$$

Then, we can visualize all peaks like so:



Next, we require a way to sort these peaks into clusters.

I implemented weighted k-means clustering, tagging peaks by wavenumber and width, normalising width to 10x w.r.t. the wavelength dimension and duplicating datapoints by weight (width x prominence).

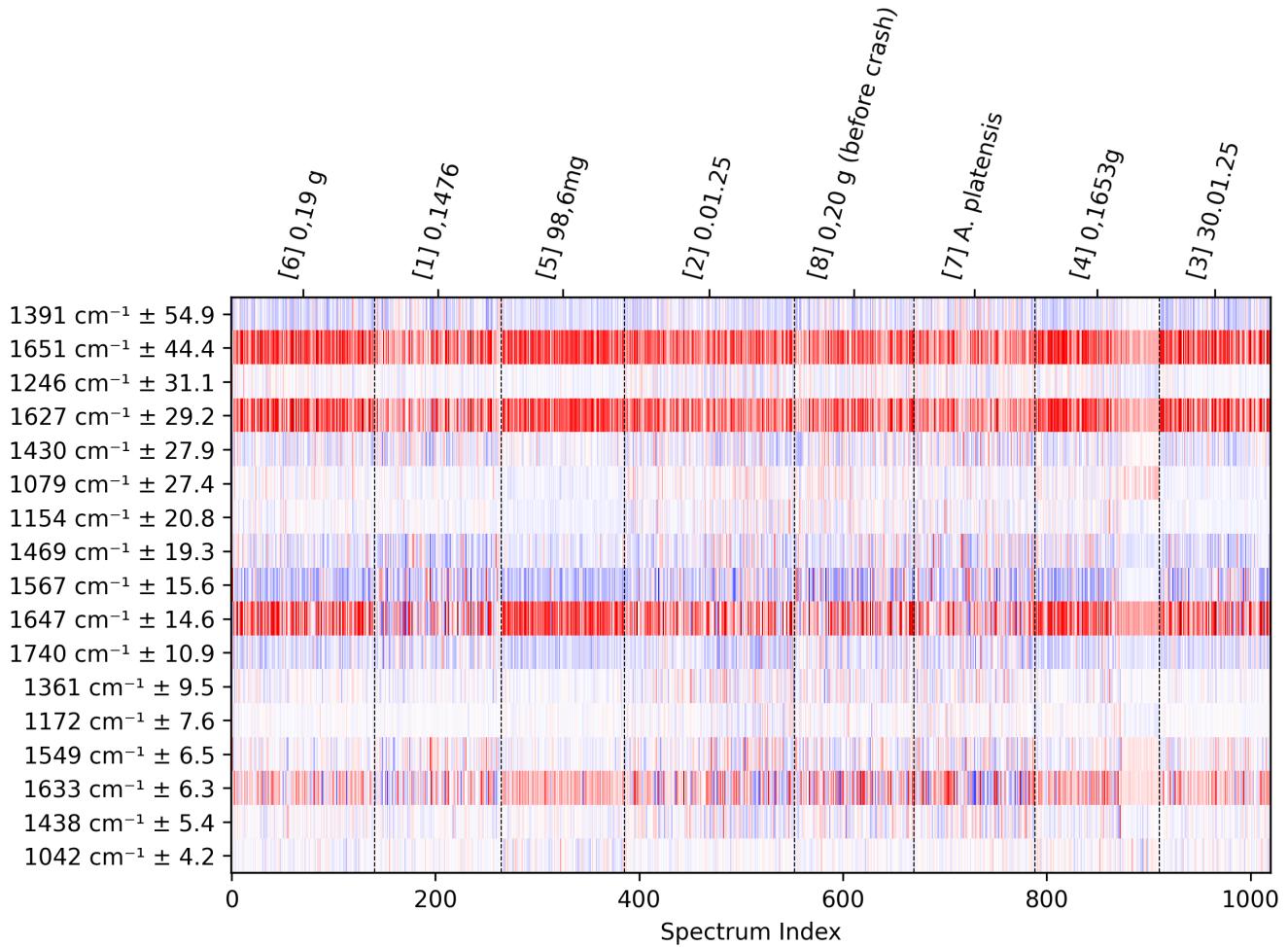


The following clusters of peaks are identified:

- 1391 cm^{-1} (width: 54.9) : ...
- 1651 cm^{-1} (width: 44.4) : **Amide I / Water**
- 1246 cm^{-1} (width: 31.1) : ...
- 1627 cm^{-1} (width: 29.2) : **Amide I / Water**
- 1430 cm^{-1} (width: 27.9) : ...
- 1079 cm^{-1} (width: 27.4) : ...
- 1154 cm^{-1} (width: 20.8) : ...
- 1469 cm^{-1} (width: 19.3) : ...
- 1567 cm^{-1} (width: 15.6) : **Amide II**
- 1647 cm^{-1} (width: 14.6) : **Amide I / Water**
- 1740 cm^{-1} (width: 10.9) : **C=O stretch**
- 1361 cm^{-1} (width: 9.5) : ...
- 1172 cm^{-1} (width: 7.6) : ...
- 1549 cm^{-1} (width: 6.5) : **Amide II**
- 1633 cm^{-1} (width: 6.3) : **Amide I / Water**
- 1438 cm^{-1} (width: 5.4) : ...
- 1042 cm^{-1} (width: 4.2) : ...

Note: Position and width were given to the k-means algorithm without any transformation applied. Significance was used as a pseudo-weight, i.e. multiplicity of points. To achieve more "relevant" clusters, one could construct a transformation that virtually amplifies the distance between wider peaks. Then again, it should actually be more likely that a pair of peaks with a given distance in position should belong to the same cluster, if they are themselves wider.

As a measure of how well any of these peaks coincides with an actual spectrum, I calculated the scalar product of the spectrum with certain basis functions corresponding to each peak cluster's center. We get...



Red means positive, blue means negative.

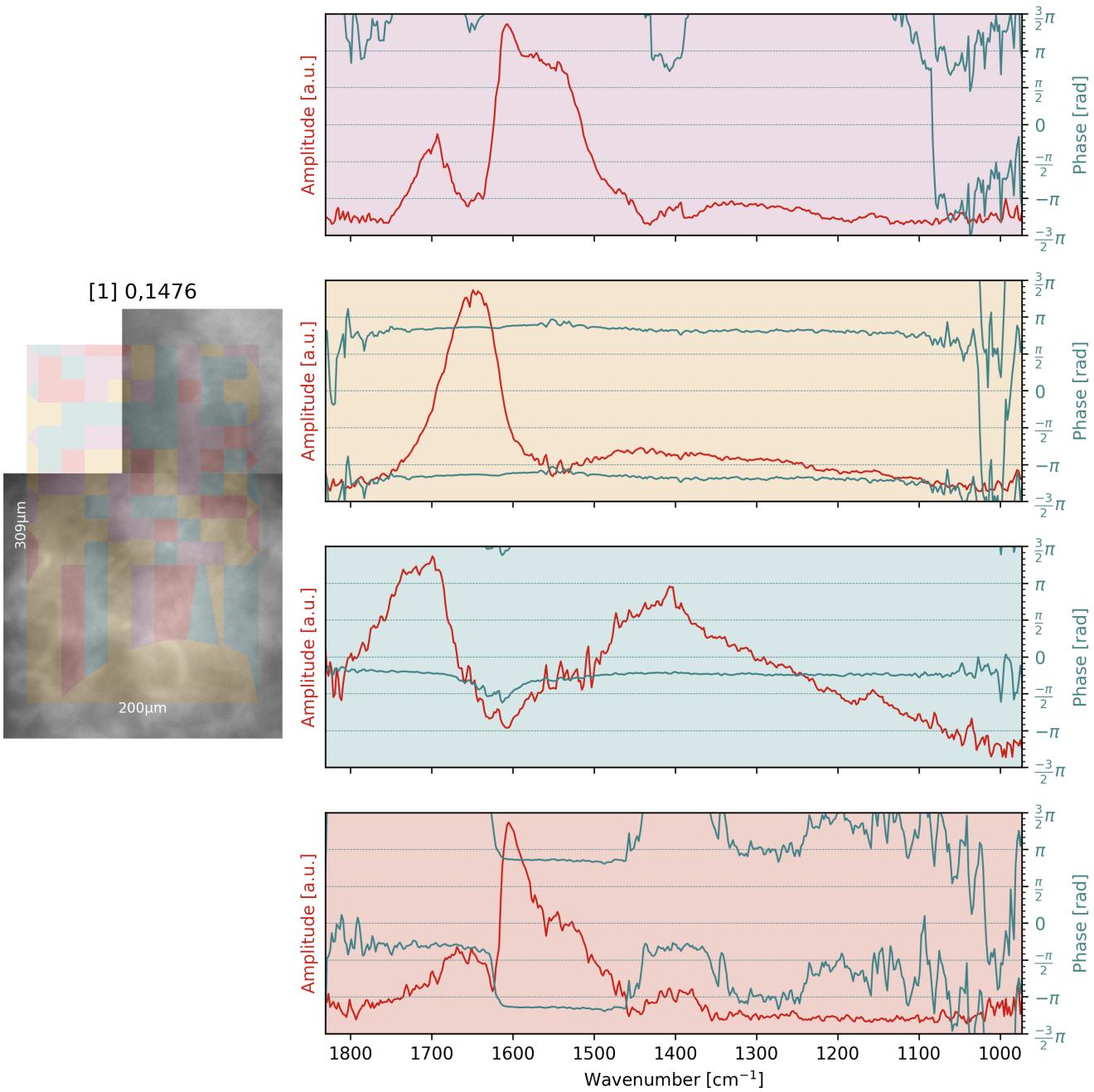
Unfortunately, decomposing spectra into a basis of these peaks did not uncover any correlation between samples either.

Hyperspectral Image Segmentation

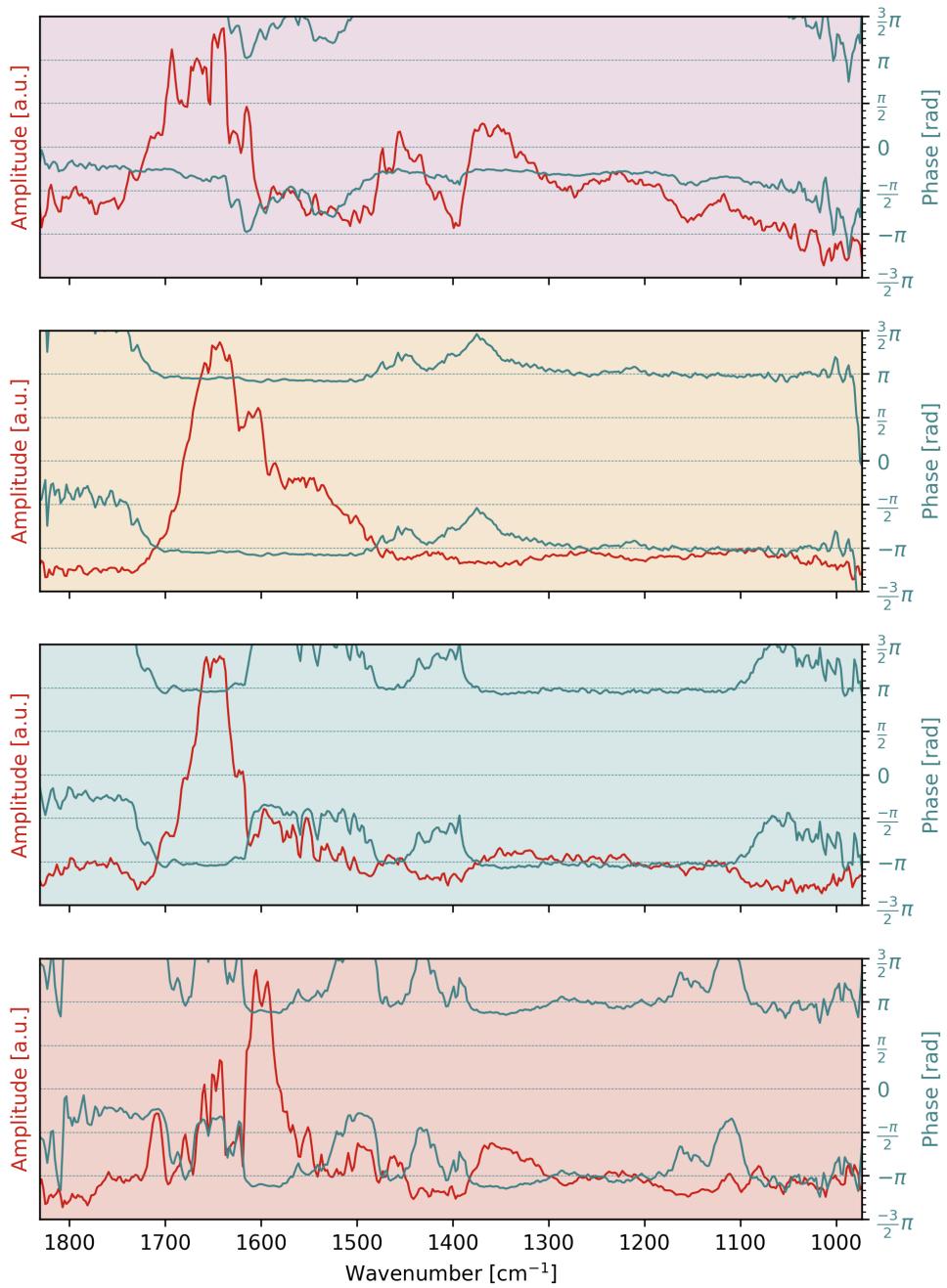
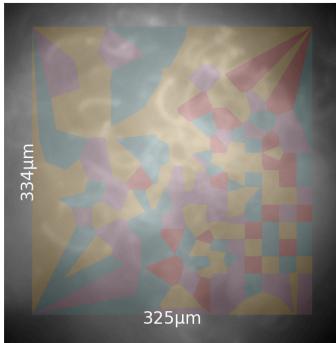
Instead of trying to model and track hypothetical spectral features through all measurements, let's attempt to classify spectra and see if these classifications are spatially correlated.

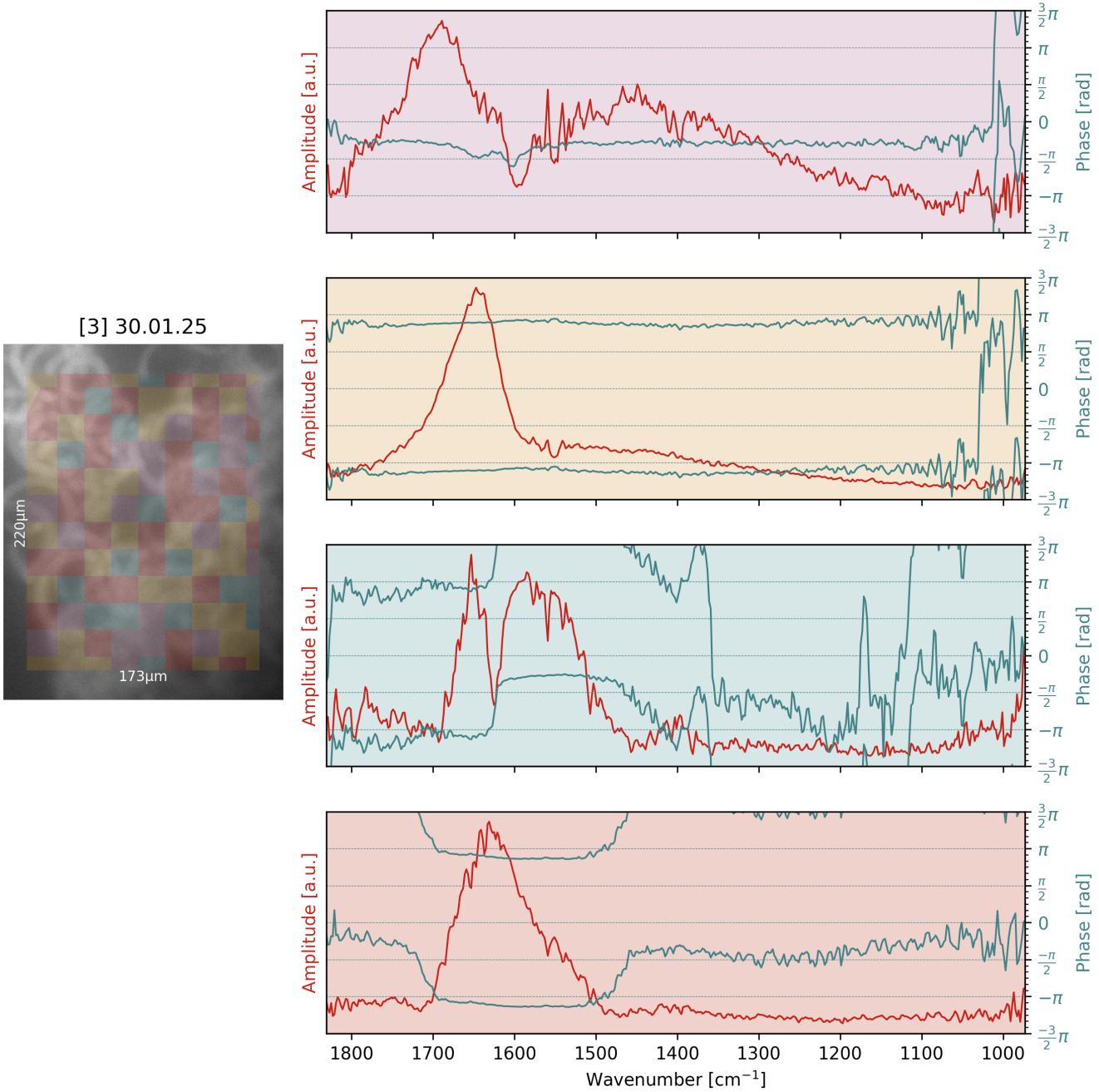
I calculated k-means clusters of the spectra in each sample (separately). I prescribed the number of clusters to be 4. We consider the X and Y channels.

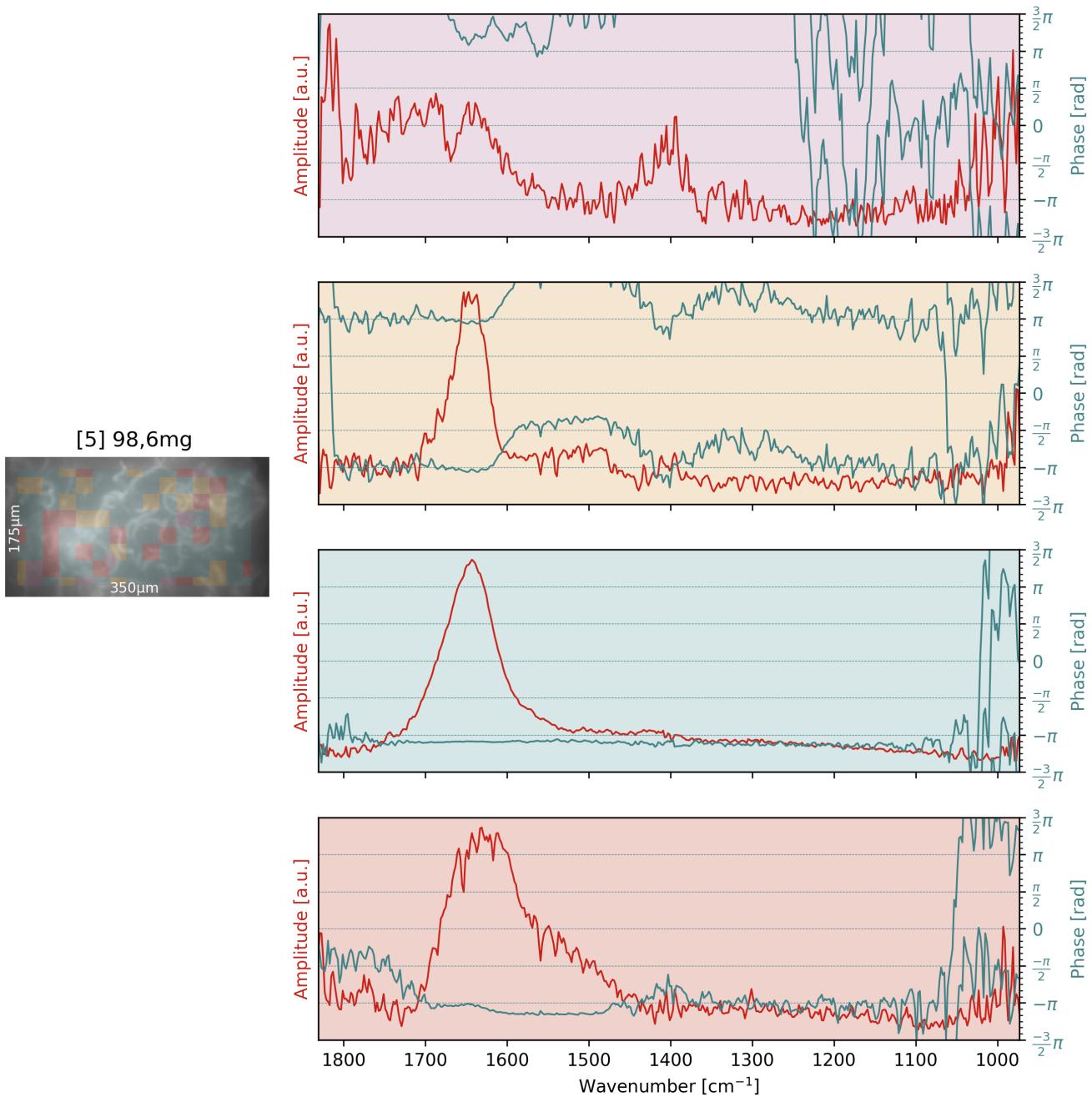
Each region of the image is then labelled with the cluster that its spatially closest spectrum belongs to. In the following, labelled images are presented on the left and representative spectra of the 4 clusters (the cluster centers) are shown on the right.

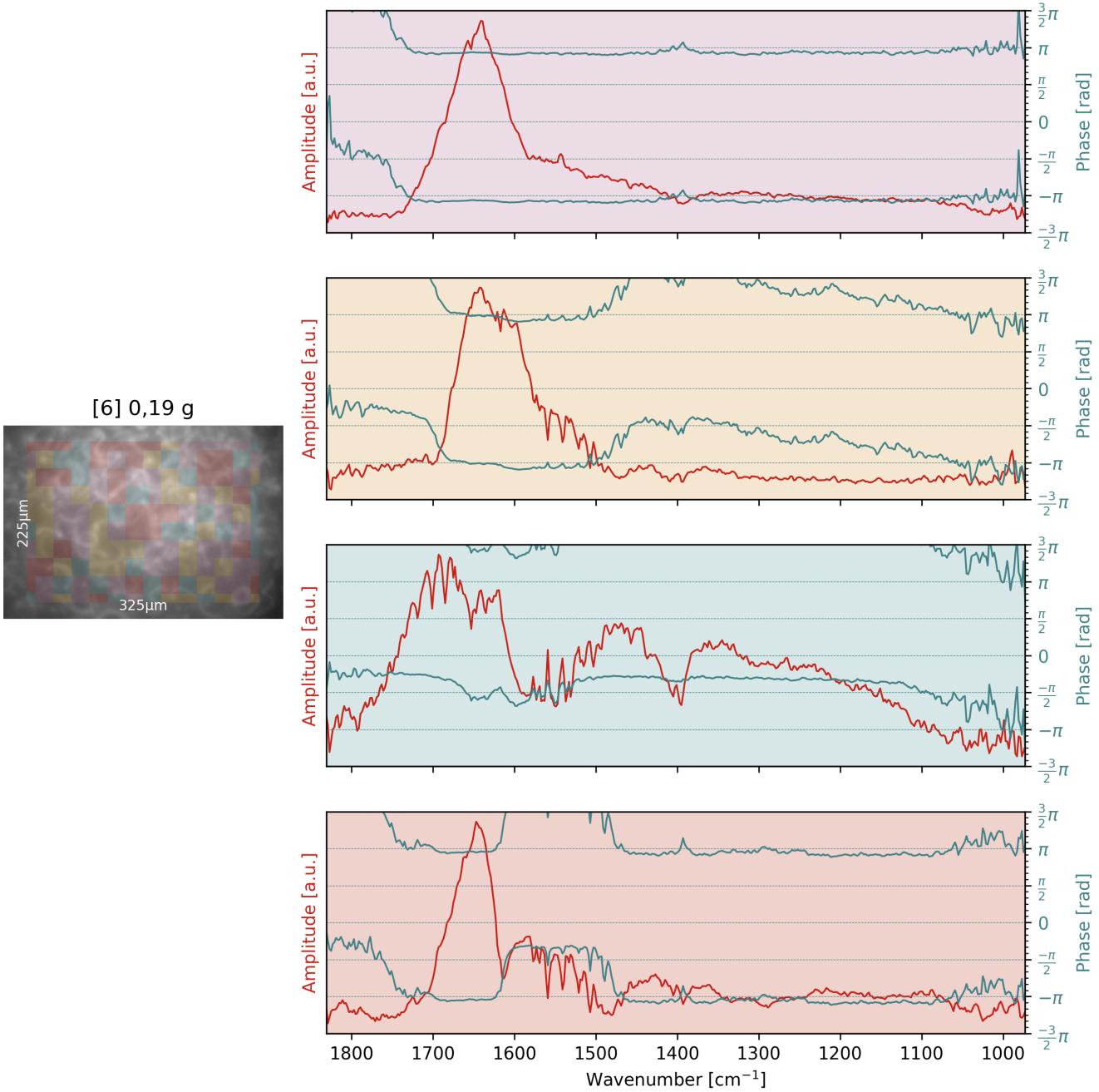


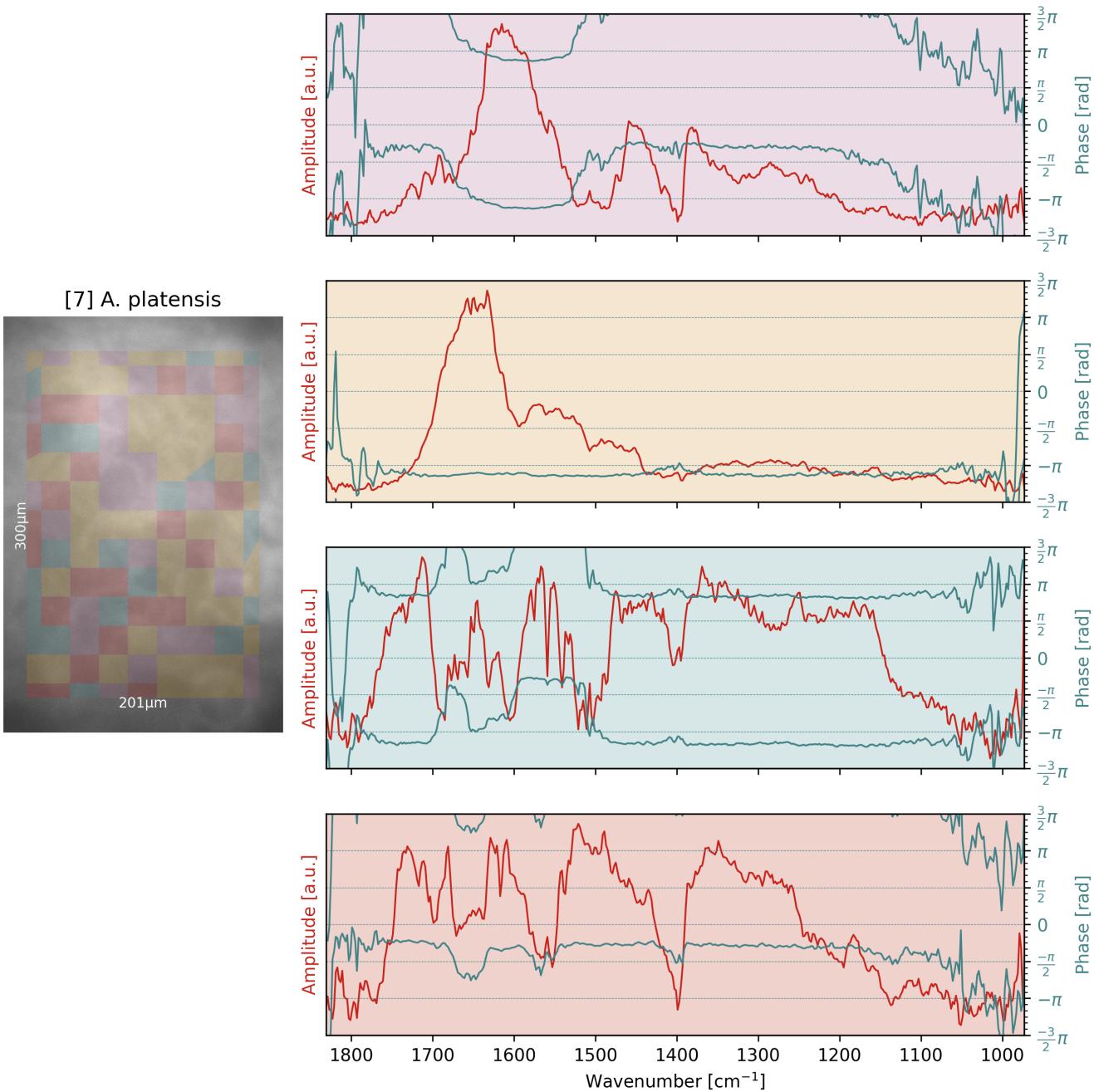
[2] 0.01.25

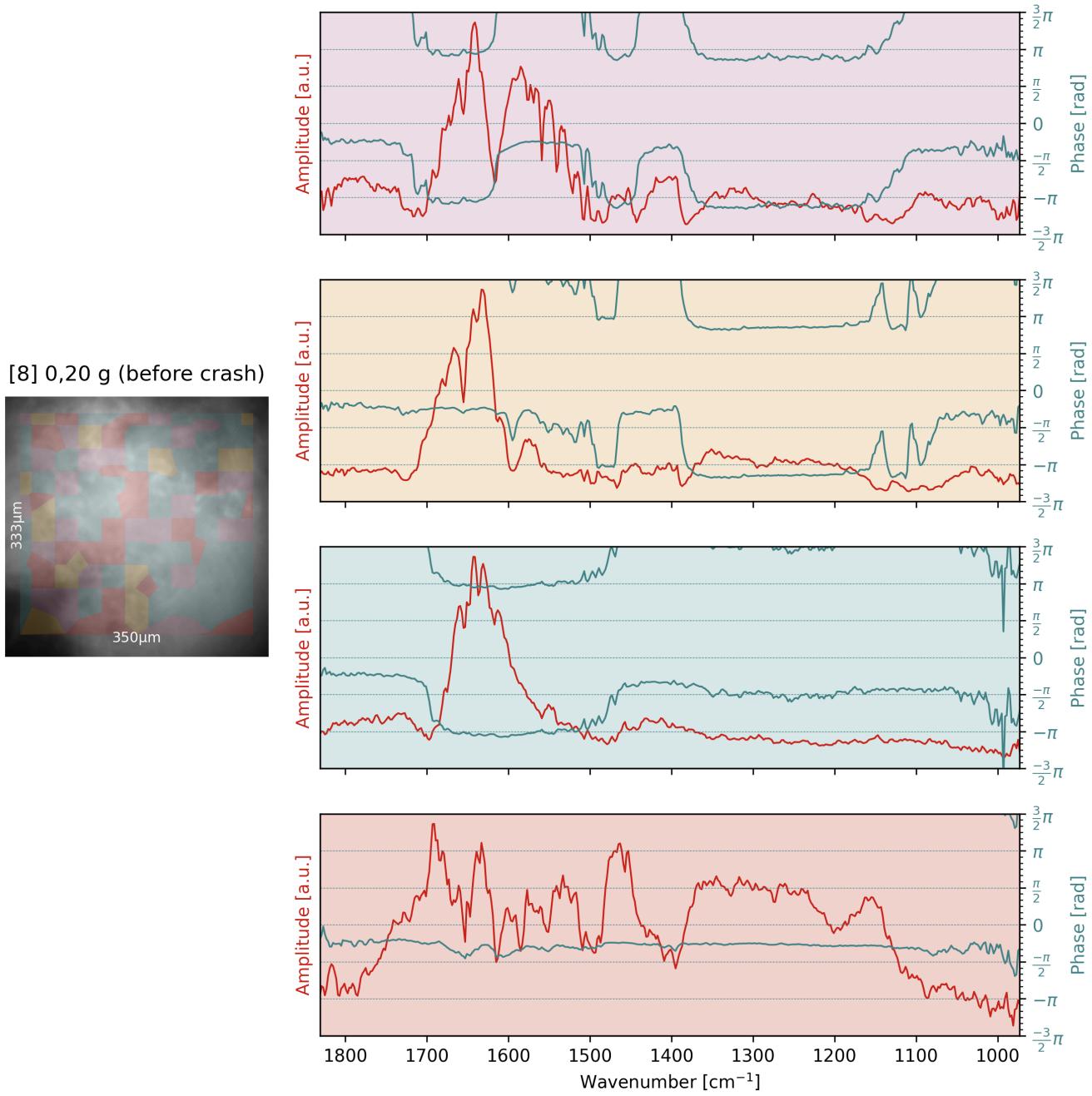




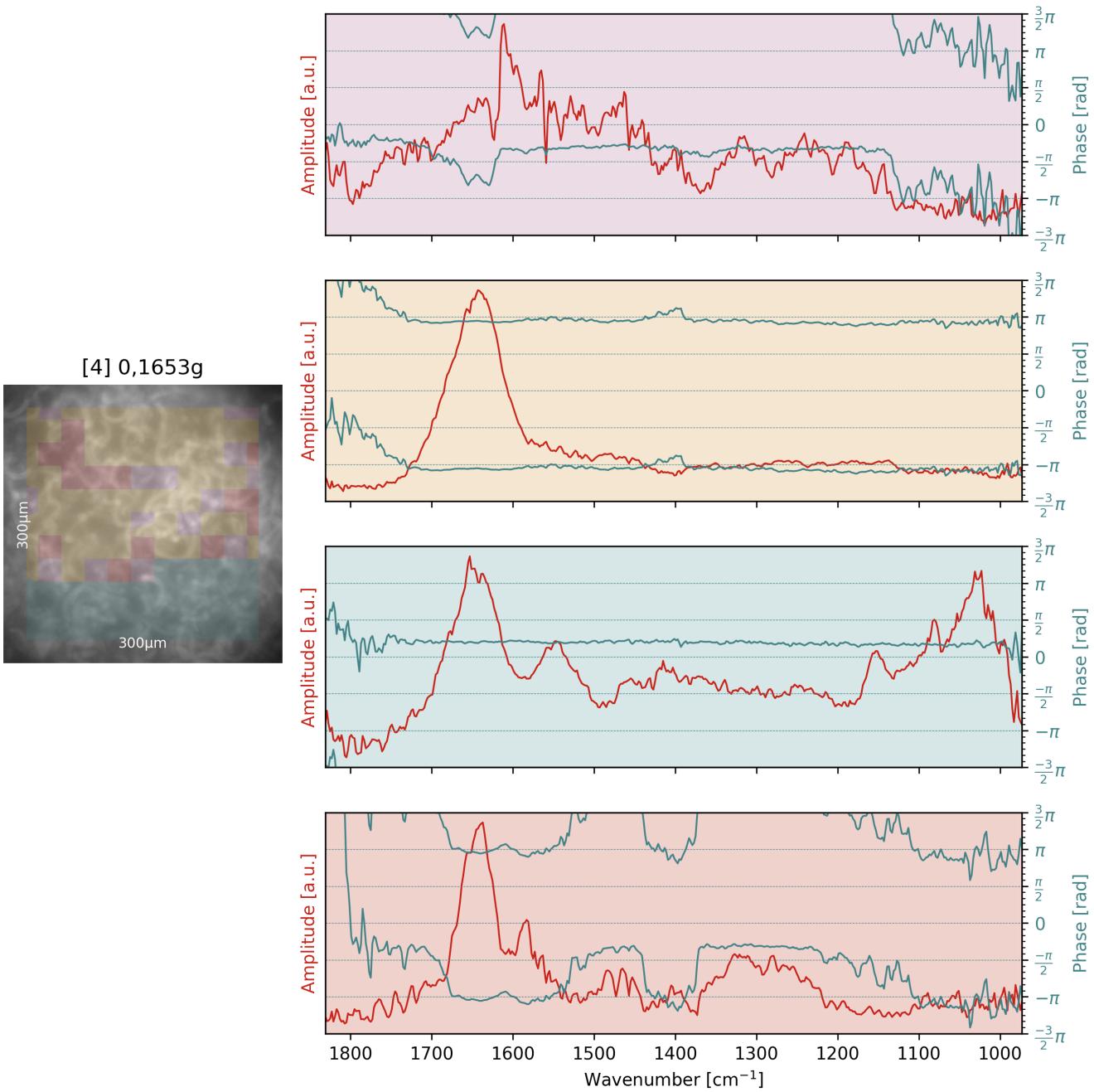








In most of these, there seems to be no apparent spatio-spectral correlation.



Here, however, a clear spatial correlation is apparent. The upper two thirds of the image is labelled mostly as cluster 3 (yellow) with some clusters 1 (red) and 4 (purple) in between. Meanwhile, (something like) the lower third is exclusively labelled as cluster 2 (blue).

Both clusters 2 and 3 have, at their centres, comparatively low-noise spectra. Note that only cluster 2 shows the secondary peaks in the region between 1000 and 1200 cm^{-1} and that its Amide-II peak is the most prominent one and the one placed most closely to 1550 cm^{-1} .

One might surmise that in the blue-labelled region, and, to a lesser extent, in the red-labelled region, both proteins and lipids are being detected, while in the yellow-coloured region, only water is being detected.