Detection of erroneous scans

# Finding invalid scans

In Opus

* Start Quant 2
* Click <Spectra> tab
	+ Click <Add Spectra>
		- Navigate to folder containing spectra
		- <Ctrl A> to select all spectra, followed by <Open>
		- Repeat to add scans from additional folders.
	+ Click <Set Sample Numbers>
		- In “Set Sample Numbers” dialog box:
			* Check “Set sample numbers according to file names”
			* Click <Set>
			* Click <Exit>
	+ CHECK the number of scans loaded equals the total number of scans in the relevant folders. If there are fewer scans showing in OPUS, it is probably because one or more of the scan files is invalid and cannot be loaded.
* If there is an invalid scan file it will probably be smaller than the others (e.g. 2 kb instead of 27 kb). Use File Explorer to find which scan(s) are invalid:
	+ Go to the relevant folder(s)
	+ Click the <View> tab
	+ Select <Details> (so that the file size is displayed)
	+ Click the heading of the column labelled “Size” – this will sort the files in order of size. Look for any files that are NOT 27 kb. Move them to a separate folder named ‘MIR\_Spectra\_Errors’.
	+ Rescan these samples using OPUS. When finished make sure the filename is EXACTLY the same as the file removed (including the “.0” or “.1” etc)

# Finding outlier scans

Next look for outlier scans. These have valid files but were not scanned properly or have unusual properties relative to the other samples. If they were not scanned properly, they will require rescanning.

* In Opus load the spectra into Quant 2 as above, then:
* Click <Parameters> tab
	+ Under “Preprocessing in calibration regions” select “No spectral data preprocessing”
	+ Check “Mean centering”
	+ Under “Calibration regions” enter from 4000 to 600
	+ Check “PCA”
	+ Under “Factors” enter 10
	+ Click <Factorize>
		- Wait for principal component analysis to run
	+ Click “Show Scores”
* Opus initially shows PC1 on the x-axis and PC2 on the y-axis.
	+ - You can zoom in on potential outliers by clicking near the point, holding down the left button and drawing a rectangle around the point. This will let you see if there are several points overlapping each other.
		- To zoom out, right click on the graph.
		- To identify a point, play the cursor over the point and a box will appear with the details, including the file name.
	+ Look for outliers, which might look like this:



* + - The sample circled in red is clearly an outlier. It is a single replicate scan and something probably went wrong during scanning. This sample should be re-scanned.
		- The cluster of four points circled in blue are the four replicates of the same sample. This could be because:
			* The sample was contaminated
			* The sample has unusual properties compared to the other samples.
		- Check for outliers using the other PCs by changing which PCs are displayed on each axis using the “x Axis” and “y Axis” boxes at the top.
	+ Make a note of all ouliers
	+ Click “Exit”
* Now remove the outliers and re-run the PCA as follows:
	+ Click <Spectra> Tab
		- Scroll down the list of samples to find the outliers
		- Click in the ‘Data Set’ column for the outlier samples and select ‘Excluded’
	+ Click the <Parameters> Tab and re-run the PCA
		- Under “Preprocessing in calibration regions” select “No spectral data preprocessing”
		- Check “Mean centering”
		- Under “Calibration regions” enter from 4000 to 600
		- Check “PCA”
		- Under “Factors” enter 10
		- Click <Factorize>
			* Wait for principal component analysis to run
		- Click “Show Scores”



* With the outlier removed, the PC plot should look something like this.
* However, if there are still outliers, the steps above can be repeated to remove them.
* When finished, exit “Setup Quant 2 Method” by clicking the X in the top right corner.

# Examining outlier scans

* In Opus click <File><Load file>
* For single outlier scans (i.e. when only 1 replicate is an outlier):
	+ Browse to the outlier file, but select all 4 replicates including the outlier.
	+ The spectra should look something like this, with the outlier spectrum showing in blue (with arrow) and the other 3 normal replicates showing in pink. There is clearly something wrong with the blue spectra and it needs to be re-scanned.



* Where all the replicates appear to be outliers:
	+ Browse to the outlier files, but select all the spectra in the same profile
	+ The spectra may look like this, with the outlier spectra in red and the other samples from the same profile in blue. The spectra look different to others, but it appears the actual scanning was OK. This may mean the sample is unusual or contaminated. The sample can be re-scanned, but it may or may not make a difference. If they are still outliers it may be difficult to estimate some soil properties for these samples.

