

Converting_OPUS

Andrew Sila

7/30/2019

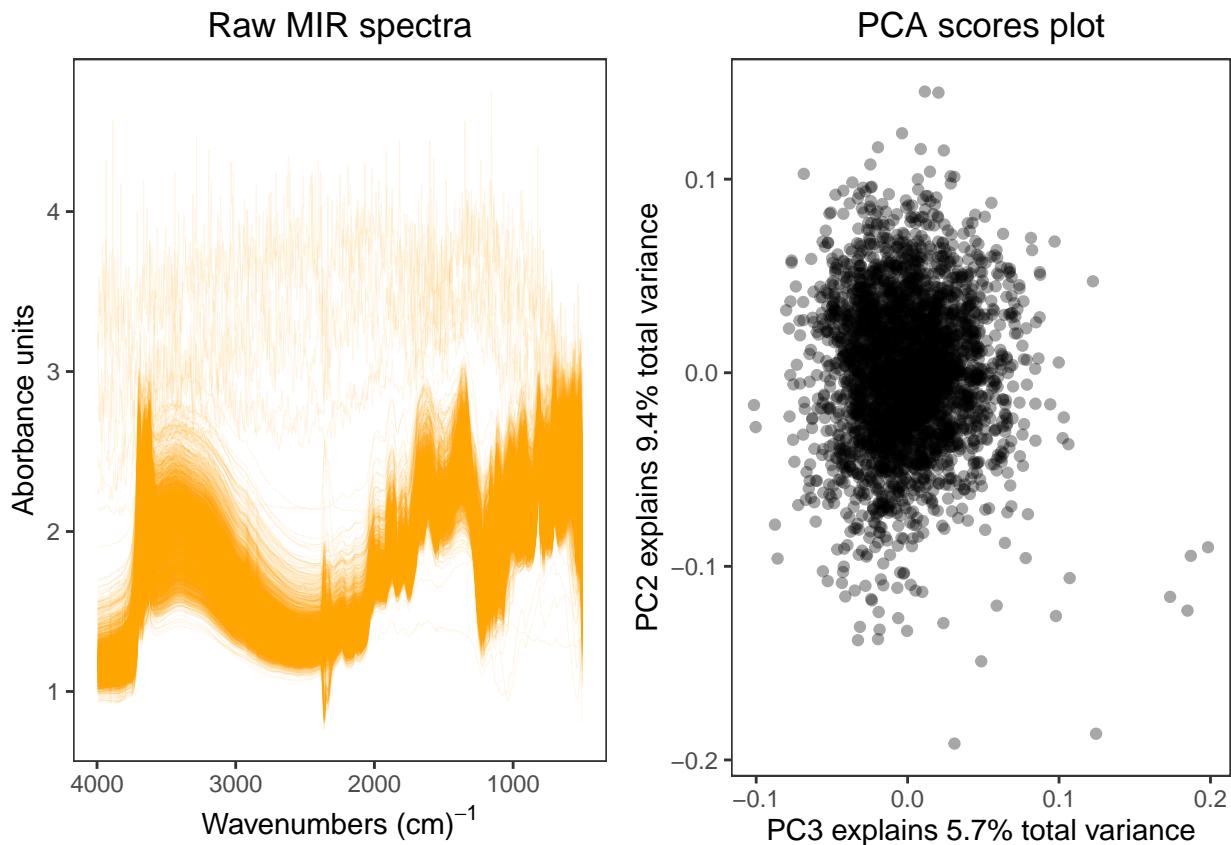
Convert ZnSe OPUS files

The first step is to convert all the OPUS binary files into a format compatible with other programs like microsoft Excel, R, Python, etc.

After conversion a data table is produced which contains spectrum scanning details and the actual intensities. A copy of the table will be saved as csv/feather into a path provided.

A visual check of the converted spectra is produced in form of spectral signatures and a PCA scores plot using preprocessed spectrum with Savitzky and Golay filter method to check for spectral clustering and screen for outliers (samples lying far from the rest or with spectrum signature shapes different from the rest).

Visualize raw spectra



Screening for outliers

There are noticeable outliers in the scores plot for points with PC3 scores above 0.1 and others with PC2 scores below -0.18.

Get sample ids for these samples:

```
##          SSN Scan_date Max_Absorbance
## 1: GH-BA-JAMS-fg2TGN7v 08/05/2017 2.728551149368286
## 2: GH-BA-JAMS-RtHq1NYB 08/05/2017 2.6795525550842285
## 3: GH-BA-DORE-5gBgJRIG 11/04/2017 4.588751792907715
## 4: GhaSIS_NR_STK_BD_Top-nmh8P2nM 20/09/2018 1.9483386278152466
## 5: GH-BA-JAMS-a3uuQL4R 08/05/2017 2.792893409729004
## 6: GhaSIS_NR_KUM_Sub-rPhph9QG 08/08/2018 4.250682830810547
## 7: GhaSIS_NR_KUM_Top-MQ8BR49U 06/08/2018 4.171093463897705
## 8: GhaSIS_NR_KUM_Top-dyJEwC6q 08/08/2018 4.170435905456543
## 9: GhaSIS_NR_STK_Sub-g0stzwLr 03/10/2018 2.801908254623413
## 10: GhaSIS_NR_KUM_Top-h0Qp8HAn 08/08/2018 3.6005265712738037
## 11: GH_BA_WECH-UB02sxN9 06/04/2017 4.795644283294678
```