Discrimination of olive oils origin based on FTIR Spectroscopy data by OPLS

Evgeniia Martynova (s1038931)

Lili Mészáros (s1015790)

Radboud University

# Data. Spectrum is obtained by Fourier transform infrared spectroscopy

- FTIR spectroscopy is fast and no complex samples pre-processing is needed

- 2 measurements of each sample made within 1-24 days interval

- 60 samples of olive oils from 4 countries obtained for the original study [1]

| Group designation	Country of origin	No. of samples
1	Greece	10
2	Italy	17
3	Portugal	8
4	Spain	25
	total:	60

[1] Henri S. Tapp, Marianne Defernez, E. Katherine Kemsley, FTIR Spectroscopy and Multivariate Analysis Can Distinguish
the Geographic Origin of Extra Virgin Olive Oils
J. Agric. Food Chem. 2003

Radboud University

Preprocessing. Raw data

- **Mean-centering** – will be used to make wavelength values more comparable

- **Autoscaling** – should not be used for the FTIR data, because we would loose important information about peaks in spectrum

- **Normalization/SNV-scaling** – we are not sure if it is useful for these data
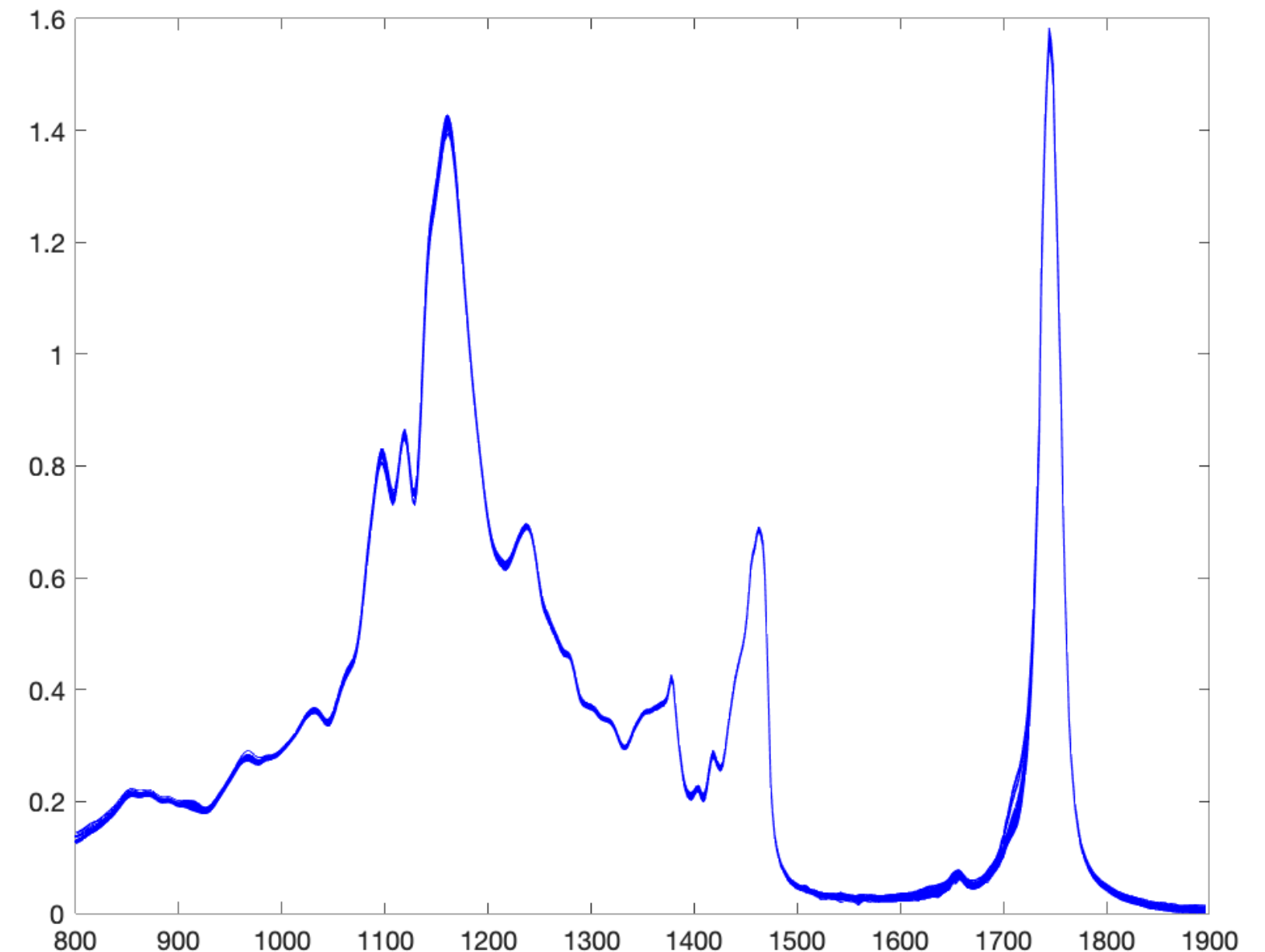
Images of raw and mean centered spectrum. Greece and Italy
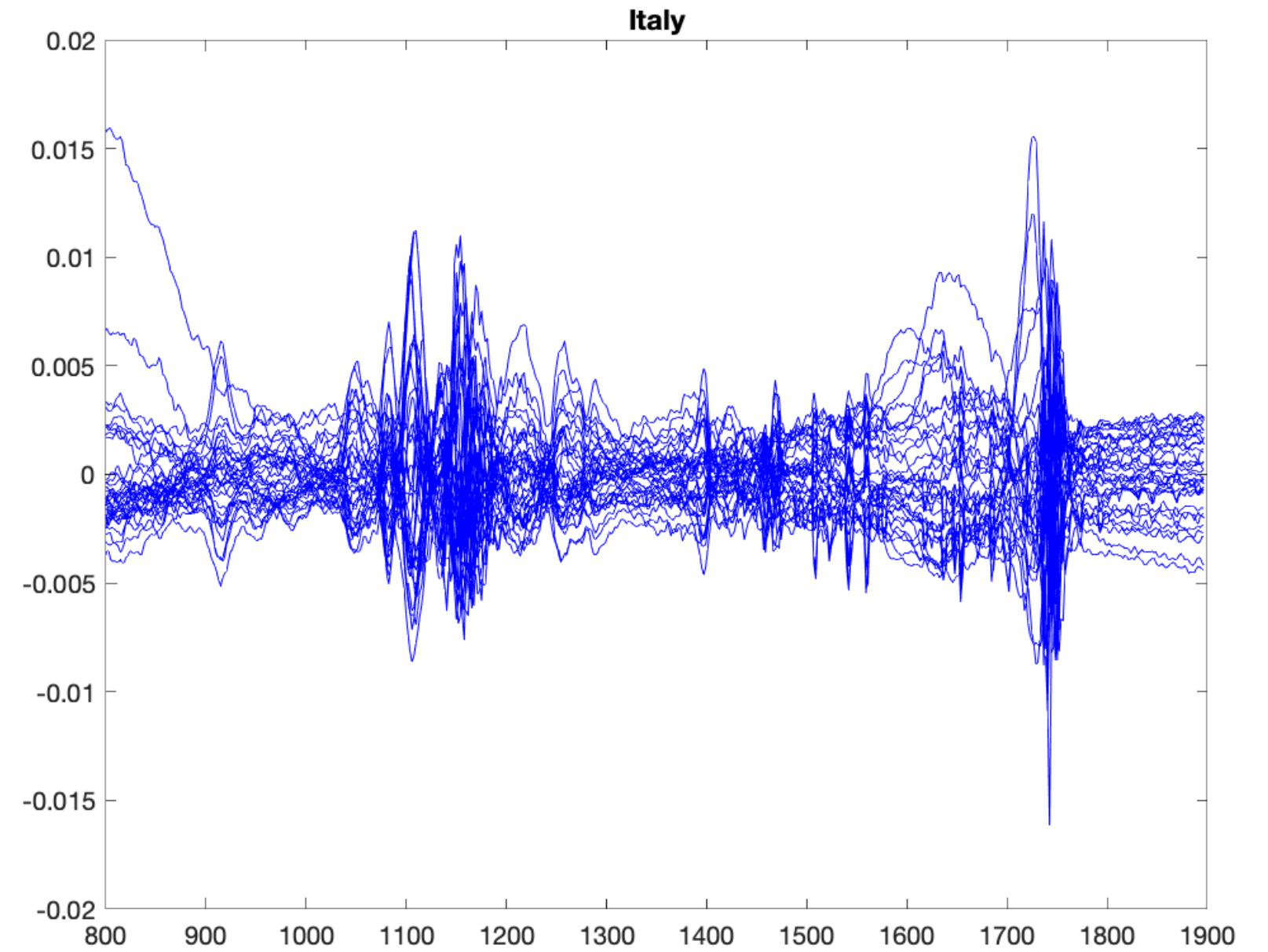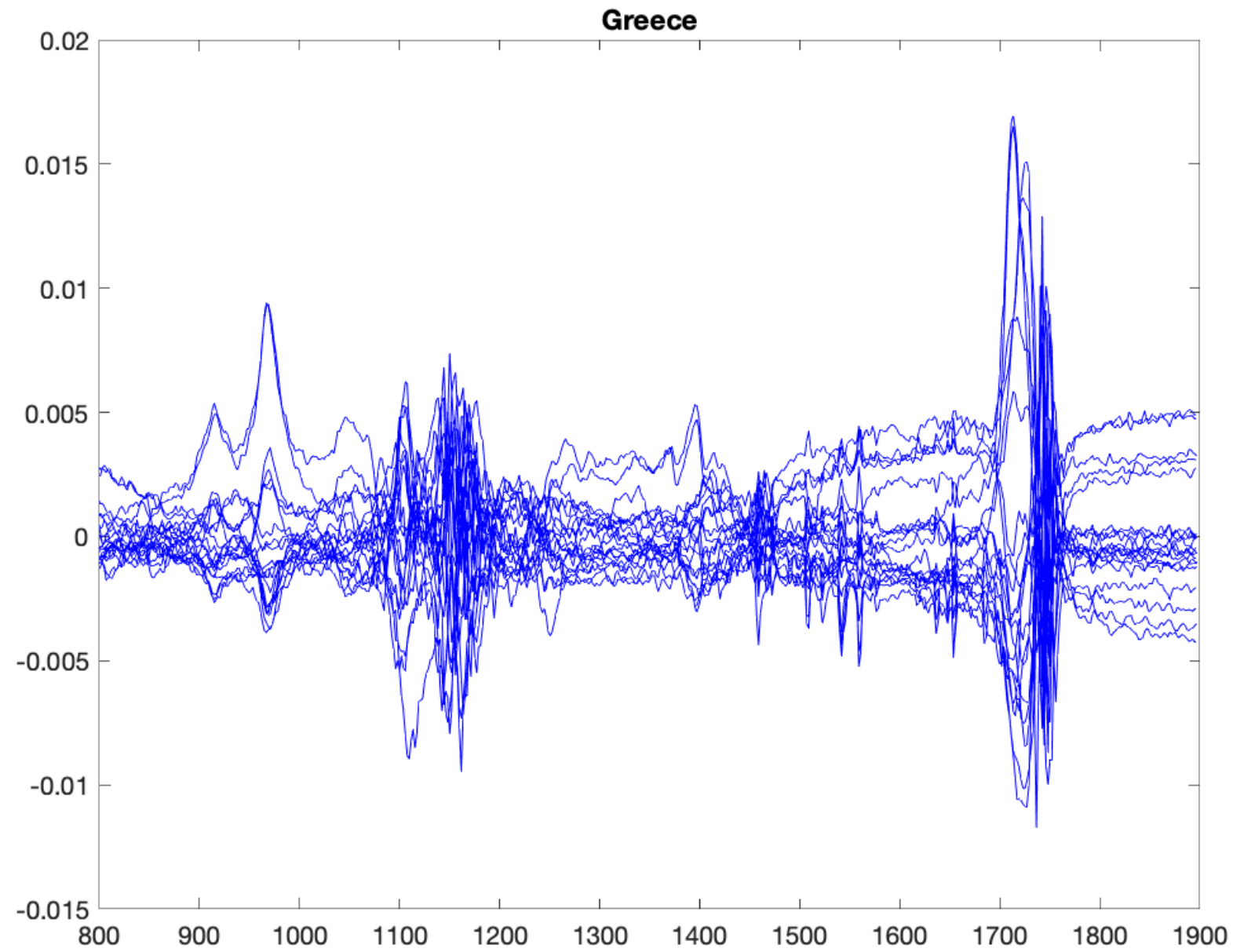
Images of raw and mean centered spectrum. Portugal and Spain
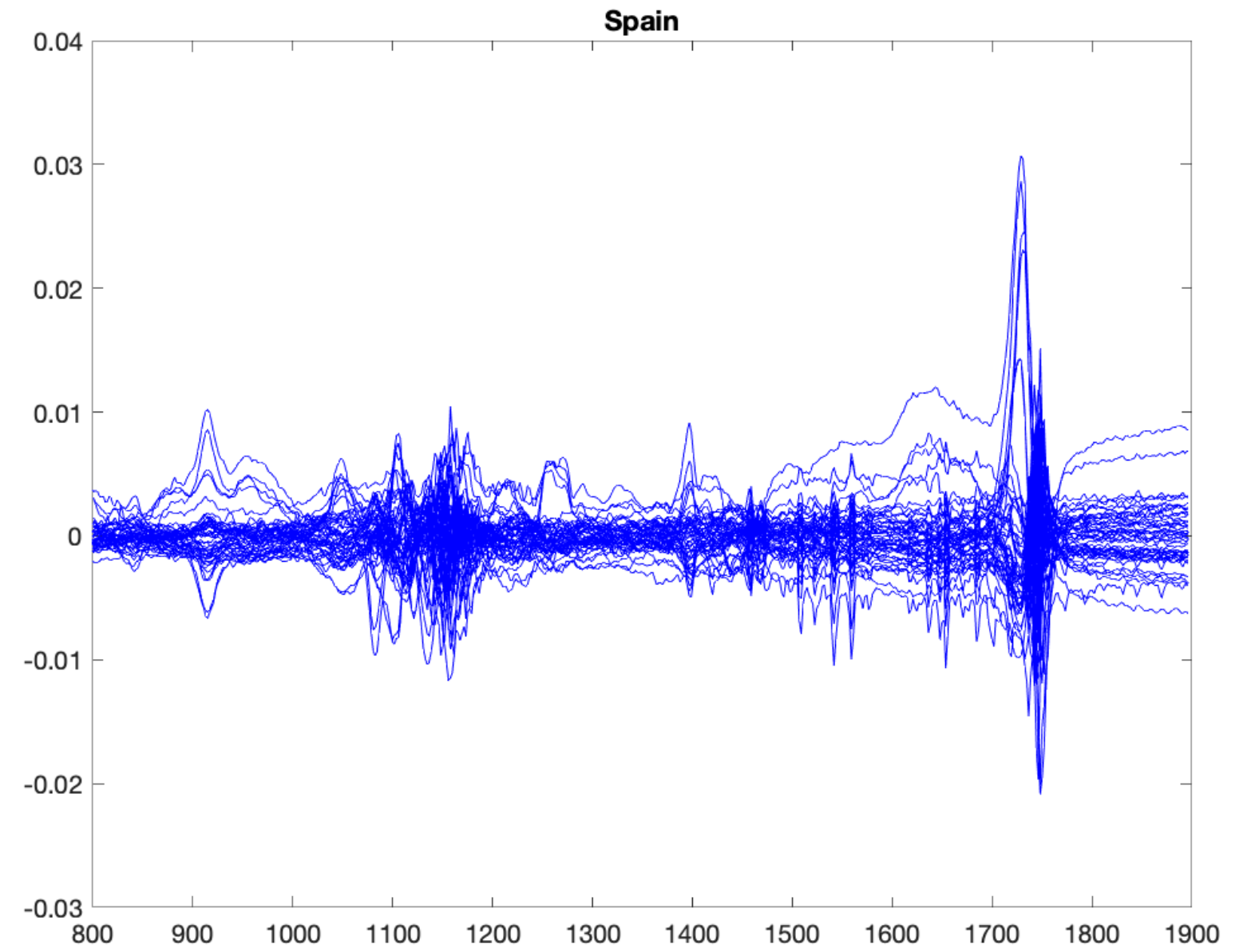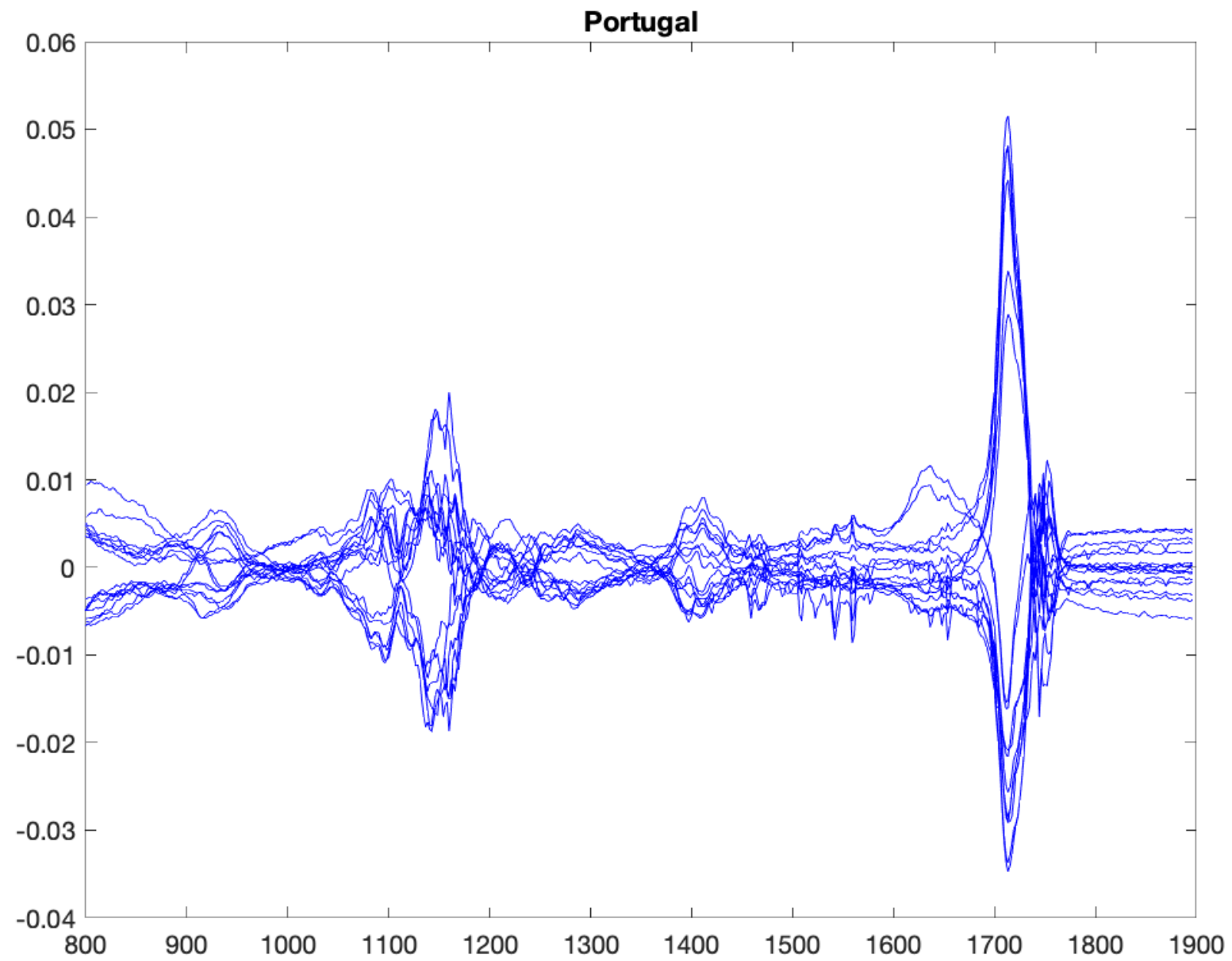
Double cross validation

- **25 %** of the data set will be used as independent **test set**

- The data is splitted into validation and test set **randomly, but**

- Since we have 2 measurements for each sample we ensure that both go to the same set to avoid sharing information between validation and test set

- We ensure that samples from different countries are evenly distributed in validation and test set

- We use the analogue of LOO cross-validation (with the correction of duplicates) for model tuning