Problem 3: Electrical load curves

The file `load.txt` contains the hourly measurements in gigawatts of the power withdrawn from the electricity grid - also called the *load* - in 2023, in Italy. Each row represents one of the 365 days of the year and each column one of the 25 hourly-spaced time instants 00:00, 1:00, ..., 24:00 (`h0` to `h24`). Additionally, the information on the type of day (`Working day` or `Holiday`) is provided through the variable `daytype`. We consider a functional data analysis approach where the hourly observations are regarded as discrete measurements of smooth functions defined on the domain $[0, 24]$.

a) Smooth the data using a Fourier basis with 13 basis functions. Provide a plot of the smoothed observations.

b) Perform a functional principal component analysis on the smoothed observations. Which proportion of the total variance is explained by the second principal component? Considering dimensionality reduction, which number of principal components would you retain? Justify your answer.

c) Provide an interpretation of the first two principal components. Support your answer with a plot.

d) Is the representation given by the first two principal components satisfying for distinguishing the working days from the holidays? Support your answer with a plot.

e) Could you find a 1-dimensional representation of the data allowing for a better discrimination than the one given by the first principal component?

Upload your results here: https://forms.office.com/e/ukD1AiS4v0