Functional Data Analysis (M-level)
Take Home Exam: Total 55 marks
Due: ~~March 17~~ **by 5pm**
April 1

General Instructions

1. Your final answers should be submitted as a **PDF** file, labelled with your name, which version of data you are using with the relevant figures, tables etc. included. The filename should be of the format Surname_Firstname_studentnumber.pdf. There is no word limit on the assignment, but please try to be concise in your answers, including only results which are relevant to the specific questions you are answering.

2. No computer code is provided for the test: but you are allowed to look up any of your notes and use/adapt any code discussed in class or write your own. However, your full **R script**, in working condition, must also be submitted as part of the test, in a separate text file. The filename should be of the format Surname_Firstname_studentnumber.R.

3. **Important: you are not allowed to discuss any part of the assessment with your classmates or anyone else** Your submitted work should be entirely your own-any evidence of collaboration will be heavily penalised, in accordance with University policy.

---

1. **Analysis of Enhance Vegetation index data**

   [You can use a maximum of four pages including graphs and 11pt Arial font to answer this question.] The data consists of remote sensing measurement of greenness of trees of 200 locations over a year at 8 day interval. See column names for exact day of the year. Download the appropriate data from the course website for performing the below mentioned tasks.

   ```
   ID number ending with 1 or 6            version1
   ID number ending with 2 or 7            version2
   ID number ending with 3 or 8            version3
   ID number ending with 4 or 9            version4
   ID number ending with 5 or 0            version5
   ```

   (a) Provide one composite visualisation of the raw data as connected lines over the year for every single observation, using the `matplot` function. Make sure the x-axis reveals the actual time scale.   [3 marks]

(b) Use a saturated B-spline basis, spanning over the whole year to fit the data using a standard roughness penalty and choose the penalty para mater using cross validation. Comment on the smoothness of the curves and provide graphs of GCV and the final smoothed data. [8 for code and graphs and 2 for comment]

(c) Adjust your code to use a harmonic acceleration penalty with a period of 1 year. Choose an appropriate penalty para mater using cross validation. Comment on the smoothness of the curves and provide graphs of GCV and the final smoothed data.
[8 for code and graphs and 2 for comment]

(d) Based on the appropriate harmonic acceleration penalty fit calculate and plot the graphs of the first and second and derivatives of the curves. [2 marks for code, 3 for graphs]

(e) Conduct a un-penalized principal components analysis of these data. How many components do you need to recover 80% of the variation? Do the components appear satisfactory?
[6 for code and graphs and 4 for comment]

(f) Try a smoothed PCA analysis from the raw data. Choose the smoothing parameter by cross-validation. Plot the cross-validation curve. Plot the new smoothed principal components. Does this appear to be more satisfactory than the unsmoothed version?
[8 for code and graphs and 4 for comments]

(g) Provide a interpretation for the smoothed principal components. [2 marks]

(h) Using the `inprod` function from the `fda` library numerically show that the un-penalized principal components are orthogonal, but the penalized principal components might not be. You can just use the first 4 principal components in each case. Ignore any "convergence" warning while executing the `inprod` function. [5 marks]