CE885 Assignment 1

1 Introduction

This assignment requires you to analyse time series data. It is for you to learn how to use basic Matlab functionality to handle and analyse data. Much of what you will need to know to complete this assignment will have been covered in lectures and laboratory classes, but you will also need to solve some new problems.

Please note that the Assignment 2 will be a continuation of Assignemt 1. In Assignment 2, you will apply advanced analysis and processing methods to the same data. Keeping this in mind may help you in making decisions for Assignment 1.

2 Specific requirements

You will have to complete five steps:

1. Find a suitable data source
2. Import the data into Matlab
3. Clean the data
4. Analyse the data
5. Produce a report

The details of each step are given below.

2.1 Finding a suitable data source

You can choose any time series data for your analysis as long as you think that it is suitable for the analysis expected in Assignmet 1 and Assignment 2. The data also has to satisfy the following conditions:

- The data should be publicly available.
- The data has to be in the form of time series (or you should be able to convert it into time series).
- You will most likely need several related time series spanning the same time period.
- The amount of data should be sufficient for advanced analysis. Usually, you would be expected to have at least a few thousand values per time series although there can be exceptions.

Below are a few examples of suitable data sources, but you can pick any other sources you find interesting.

- The UK power generation data available from the GridWatch website (http://www.gridwatch.templar.co.uk/download.php). For example, the demand, wind and hydro time series are likely to have some correlation. You can also study the distributions of these variables. In Assignment 2, you might be able to learn the relation between all three of these variables and analyse periodicity of the data.

- Stock price data. You can download the data from Yahoo Finance (https://finance.yahoo.com), searching by the code of a stock, selecting the Historical Data tab, choosing the maximum time period, clicking Apply and then Download data. For example, you could consider the TSCO.L (Tesco), BRNT.L (Brent Crude) and RR.L (Rolls Royce) stocks. You may expect some correlation between them due to variation in the British economy. You may need to trim the datasets so that they have identical timespans. In Assignment 2 you will, for example, be able to train and evaluate a predictor of the next value in the time series. Stock price data is also likely to have some periodicity.
- Currency exchange rate data. You can find this data at the website of the European Central Bank:
  https://www.ecb.europa.eu/stats/exchange/eurofxref/html/index.en.html
  The specific file you should download is near the end of the page, under the heading 'Downloads'. Ignore the first list of files (current rates) and look further down to the small heading 'Time Series'. You will need the Zipped CSV file (eurofxref-hist.zip), because you need to read this data into Matlab (the PDF and XML formats are not convenient for this). The analysis of the currency exchange data can be similar to that of the stock price data.
- Sound file. It is easy to extract the track data. If the file contains two tracks (stereo sound), the corresponding variables are likely to be closely related. Interesting analysis includes studying the volume of the sound as it changes over time. In Assignment 2, you may want to study the lag between the tracks, sound specrum and its changes over time, music rithm, etc. You can also train a machine learning system to distinguish different types of sounds.

While you can use any of the above examples, you are encouraged to be creative in choosing the data. You are also strongly encouraged to discuss your choice of data with the lecturer.

2.2 Importing the data

The original format of your data depends on the source you select. You are required to find an appropriate method to import the relevant data into Matlab for processing. As a result of this step, you will have raw data imported into Matlab for further processing.

2.3 Cleaning the data

You may have to convert the raw data into time series, or might have to trim the time series so that they have identical time periods. It is also common in real-world datasets that some data values in are missing or are unrealistically low or high, or NaN. It is a part of the assignment to derive an automated approach to prepare and clean the data to exclude such discrepancies. You may also want to normalise the time series.

2.4 Analysis of the data

Raise some meaningful questions about your data and use appropriate tools to answer them. These tools may include plotting the time series, plotting histograms, plotting correlations, measuring correlations, performing statistical analysis, smoothing the data down, etc. Be creative and feel free to discuss your ideas with the lecturer!

2.5 Report

Your main submission for this assignment will be a written report. The report must be sensibly organised and clear to read.

The report should include the following sections:

1. **Data**. In this section, describe the source of your data, the data itself and your motivation in selecting this data (e.g., what kind of analysis are you willing to conduct in this and the second assignments). Include clear instruction for obtaining the source data.
2. **Import**. In this section, describe how you imported the data into Matlab. Include code snippets for the key pieces of code. Discuss the challenges you faced when importing the data.
3. **Data cleaning**. In this section, include a discussion of what preparation and cleaning might be needed for your data, and what methods did you derive for that. You may want to include one or several plots giving examples of issues that you observed and how they were corrected. If you did not observe any issues in your data, this section will only include a set of tests that demonstrate that.
4. **Data analysis**. Start this section with the research question(s) that you want to answer, and then give details of the analysis. Include figures where appropriate.
5. **Conclusions**. Discuss the strengths and weaknesses of your study. What would you do differently next time? Propose further analysis that could be undertaken in Assignment 2.

3 Submission

You need to submit the report and the Matlab scripts/functions that you developed for this assignment. All the code, including data import and cleaning, has to be included in this form.

4 Grading guidelines

This assignment will receive a single percentage mark, according to the following descriptors, which are provided for guidance:

Grade	Descriptor
80% or greater	The report and the work undertaken exceed expectations of the distinction level.
Distinction (70-79%)	A well presented and clear report explaining all your decisions, methods and conclusions. The wording of the report is concise and precise, with appropriate use of mathematical language. The figures are of high quaility and graphs include axes names, units and caption. There will be evidence of original or creative thinking (not necessarily mathematically rigorous) in the method going beyond the material covered in class, but supported by references or clear explanations of what has been done.
Merit (60-69%)	A well presented and clear report explaining all your decisions, methods and conclusions, possibly with some details missing. The report is mostly easy to follow. The figures are of high quality. The method may be based on concepts that were covered in class, or readily found in the literature or online, supported by references.

Pass (50-59%)	An adequate report and method, but the method may be incomplete (for example the data may have been imported and cleaned but no significant analysis done). There may be some lack of clarity about how the data was analysed, or what was the motivation for choosing this data. The report may have some shortcomings, including grammar, use of language or quality of figures.

5 Academic offences

The work you submit for this assignment must be your own work. Any aspect of the submission which uses the work of others (students or not students) must be formally acknowledged and/or referenced. It is acceptable for you to discuss this assignment with other students, and even to cooperate on small parts of it, but **any ideas or help that you obtain from someone else must be acknowledged in your written submission**. Using ideas that you find in books, or in articles or other sources is encouraged, provided they are referenced, but you must explain them, and explain why you chose these methods.