

# Signal Processing: End of Module Open Assignment

## Preamble

There are a set of four time series in the following site:

<http://www.robots.ox.ac.uk/~sjrob/AIMS/Assignment/PredData/>

## The data

The four data sets represent *real* data - our goal is ambitious, and will stretch in some data streams the ability of *any* model to perform good forecasting and modelling. Please note that you are welcome to use any code you have already written, such as Gaussian Processes as well as third-party code, e.g. for optimisation and inference routines, or other components of your models.

There are four data sets here, the aim of this is to get experience with some real data of differing provenance. Do remember that it's not just the final performance of any algorithm that's of importance here - getting some experience with what works and what doesn't, in the details of implementation (the tricks and the pitfalls) is all part of this.

**Finance:** The data contained in `finPredProb.mat` consists of a single stream of data, `ttr`, taken once per second for 4 hours from a foreign exchange rate, as shown in Figure 1.



Figure 1: Some financial data.

**The goal** - is to produce a sequential model to provide the most accurate prediction into the future and quantify the prediction performance. Don't be alarmed if there is very little predictability.<sup>1</sup>

---

<sup>1</sup>If you find otherwise, please start a company immediately!

**Mackey-Glass chaotic system:** This system is defined by the chaotic differential feedback system of the following form

$$\frac{dx}{dt} = \beta \frac{x_{t-\tau}}{1 + x_{t-\tau}^n} - \gamma x, \quad \gamma, \beta, \tau, n > 0$$

For different values of the parameters in this model we obtain some wildly chaotic behaviours.

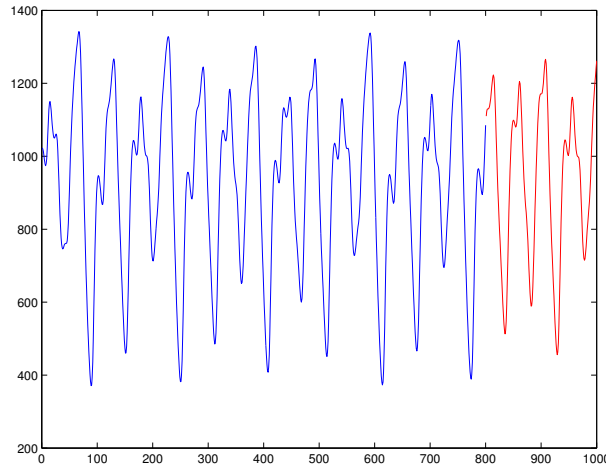


Figure 2: The Mackey-Glass sequence. you can train on the blue and use this to forecast the red.

**The goal -** is to use the 800 samples of training data (in  $t_{tr}$ ) to make accurate model forecasts about the next 200 points (in  $t_{te}$ ). Figure 2 shows the full sequence, with training in blue and the out of sample test set in red.

**Sunspot data:** This data set is somewhat of a classic data set, having been used in many analyses looking at detecting trends and cycles at multiple timescales. The data is monthly.

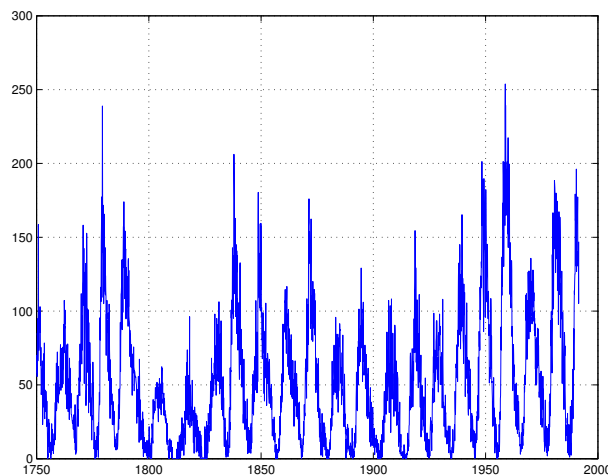


Figure 3: Sunspot data.

**The goal -** is to use the historic data up to develop a forecast for the year ahead. You can use all the data up to 1900 as a training set if needed, but then create a rolling regression to forecast 12 samples ahead each time.

**CO<sub>2</sub> data:** This is a recent data set of the atmospheric levels on a monthly basis.

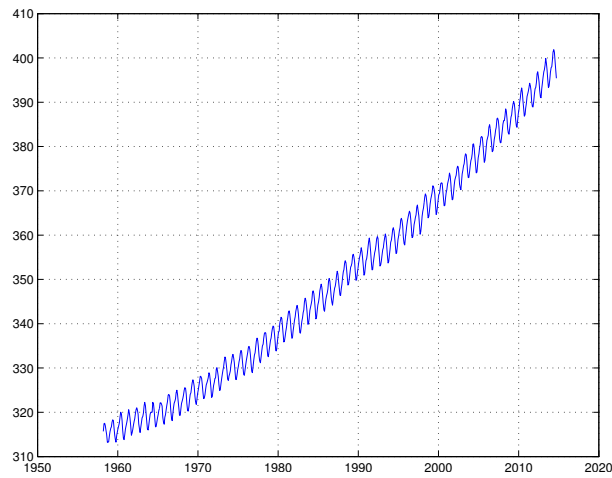


Figure 4: CO<sub>2</sub> data.

**The goal -** go on, frighten yourself (or not) by forecasting what levels will be between now and 2050.

## Submission

Please package up all code and a report (max 10 pages, typically 4-5), detailing the approaches taken and performances reached, into an archive file (e.g. zip file) and email to [sjrob@robots.ox.ac.uk](mailto:sjrob@robots.ox.ac.uk) by 6pm Monday 9th November 2015.