Background

- The MATH3802 and MATH5802 modules are assessed by an examination (80%) and a practical (20%). This is the practical, worth 20% of your final module mark.
- Comments on this sheet which are only relevant to MATH5802M students are indicated by *italic font*.
- Reports must be clearly marked with your name, student ID, and module and should be no more than six sides of A4 paper in length *(eight sides for MATH5802M)*.
- I will be available to give advice as follows:
  - in the Fourman O & P Clusters on Monday, 27th February from 3-5pm for MATH3802 students
  - *and in the Cohen B Cluster on Tuesday 28th February from 10am-12noon for MATH5802M students*.
- You can ask me for help. You can share ideas with other students, but the work you hand in must be your own.
- You must hand in your solutions by **5pm on Wednesday, 22nd March.** Late work will be penalised at the rate of 5% of the available marks per calendar day.

Data

The data are monthly water levels (in feet, measured on the first day of each month) in a small pond in rural Hampshire.

For this practical you will need to obtain the data set $X = (X_1, \ldots, X_n)$ from the file "`pond.RData`" which you can download from VLE. Once you have the file "`pond.RData`" saved in the folder where you are running R, use the command `load("pond.RData")` to read the data into R.

[Left click on "pond.RData" on the VLE Learning Resources page. When asked "Do you want to ...", click on the arrow to the right of "Save", and then "Save as". Then save exactly as it is (ie. "pond.RData") in your M-drive ("My Documents"). the data is called `X`.]

The task

The task of this practical is to analyse the pond level data, using the following list of points as a guide. Your analysis should consider the points below, but your report should *not* be laid out as "answer 1", "answer 2", ... Instead, structure your report in whichever way you prefer to make it easy to read.

- Plot the data X and examine any prominent features. Comment on your findings. Use linear regression to remove any linear trend or seasonal effects which you believe to be present in the data. Denote the residuals after removing the trend and/or seasonal effects by Y.

- Inspect the process Y and comment on whether an MA or AR process might be suitable to explain any structure present.
- For each of $p = 1, 2, 3$, using the Yule-Walker equations, fit an AR(p) model to the time series Y. You can either use the R function `ar` or solve the Yule-Walker equations "by hand".
- For each $p = 1, 2, 3$, consider the residuals of the AR(p) model. Plot a correlogram of the residuals and comment on how well the three models fit the data. Choose the model with the best fit; refer to these as Z.
- *[MATH5802M only] For each of X, Y and Z, plot periodograms. Comment on your results.*
- Summarise your results, but combining the above steps into a complete model for the original time series X.

Guidance on the report

You should take some care with the presentation of your results. This includes using a clear structure and layout, careful explanation of your results and how you obtained them, meaningful plots with appropriate labels, *etc*. You should start your report with a short (1 paragraph) summary of your findings, written in a style suitable for a non-statistician.

Note that large amounts of R output (other than plots) are *not* needed. Use R output (other than plots) sparingly or not at all. If you do include R output, make sure it is in a fixed-width font `like this` — eg. Courier in Word or verbatim in LaTeX.

Do not repeat large sections of theory from the notes — I already know what is in them! Use your limited page count to describe your analysis and conclusions.