TIME SERIES ANALYSIS

STAT 4390/5390 – OCEAN 4210/5210

Lecture: MWF 230-330 in Chase 319

Instructor: Michael Dowd
Office: Chase 116
Phone: 494-1048
e-mail: mdowd@mathstat.dal.ca
Office Hours: MW 1:00-2:30pm

Course Website:

`www.mathstat.dal.ca/~mdowd/stat4390/stat4390.html`

Copies of lecture notes, assignments, computer code, and data sets will be made available on course website. Announcements will also be posted here.

University Policies

Students with Disabilities:

Students with disabilities should register as quickly as possible at Student Accessibility Services if they want to receive academic (classroom/exam) accomodations. To do so please phone 494-2836, email *disabilities@dal.ca*, or drop in at the Killam G28.

`http://www.dal.ca/~services/ssd.html`

Intellectual Honesty

`http://plagiarism.dal.ca/policies/index.html`

Method of Evaluation:

- Assignments 20%
- Project 30%
- Midterm Exam 15%
- Final Exam 35%

(Note: no supplemental examination)

Grading Scheme:

A+	93 and above
A	85 up to but not including 93
A-	80 up to but not including 85
B+	75 up to but not including 80
B	70 up to but not including 75
B-	65 up to but not including 70
C+	60 up to but not including 65
C	55 up to but not including 60
C-	50 up to but not including 55
D	47 up to but not including 50
F	below 47

Formal Prerequisites:

1. STAT 3340 - Regression Analysis

2. STAT 3460 - Intermediate Statistical Theory

Skills Required:

- Basic concepts in Statistics including: probability, expectation, and inference

- Familiarity with calculus, matrix algebra, complex numbers

Course Text:

The text for the course is:

Spectral Analysis and Time Series
by M.B. Priestley.

It is available in Dal Bookstore.

Course Notes:

Course notes will be provided on the website. These notes are under development. Consider them as a draft and please report any errors and feel free to make suggestions. The purpose of the notes is to supplement, not replace, Priestley's book. I will assume that you have a basic familiarity with most of the material covered in sections 2.1-2.9, and 2.11-2.13 of Priestley's book. If you have problems with this material see me.

Software:

Assignments and projects will require some use of software to carry out time series analyses. I will emphasize the use of the software package MATLAB, but you may also use other analysis software for your work (e.g. Splus).

Course Goals:

This class deals with the analysis of systems characterized by dependence structure, such as variables recorded over time (but also spatial data). The emphasis is on both theory and application.

1. Develop an understanding of the theory underlying time series in the time and frequency domain.

2. Provide an understanding of how time analysis procedures work, as well as the ability to apply methods to real data sets and to interpret the results.

3. Provide some basic experience in using modern software packages for time series analysis.

Course Outline

- Introduction to time series analysis
- Stationary Random Processes
- Classical analysis in the Time Domain
- Analysis in the Frequency Domain
- State Space Models

Course Evalulation Details

Assignments:

There will be regular assignments worth 20% of the total mark. Late assignments will be given a zero mark unless a medical excuse is presented. Assignments will require both analytic work, as well as the use of computing software.

Project:

There will be a project worth 30% of the total mark. The project will require written report, as well as an oral presentation. The project will be based on an analysis of time series data, or the detailed treatment of a more theoretical topic. Further details will be provided during the course.

Exams:

The midterm will take place during regularly scheduled class time. The final will take place during the Dalhousie exam period as scheduled by the registrar. Further details for these exams will be announced in class

TIME SERIES ANALYSIS
STAT 4390/5390 and OCEAN 4210/5210
Final Exam – 2-5 PM , April 12, 2006

This final exam is worth 35% of your final grade. The time available is 3 hours. You may use two formula sheets (double sided) and you may use a calculator. Answer questions in the exam books provided. All questions are worth equal value. Graduate students must answer 10 of the 12 questions, and Undergraduates must answer 8 of the 12 questions.

1. Suppose you given two time series. Outline the goals of a time series analysis of these data, and the steps that you would take for both the time and frequency domain based analysis.

2. Consider the univariate model $X_t = 0.8X_{t-1} + \epsilon_t$. Let the forcing be given as $\epsilon_t = \cos(\omega t)$. Determine the transfer function and express it in terms of the gain (ratio of the input forcing and output response for a linear system, as a function of frequency).

3. Given the process: $X_n = aX_{n-1} + \epsilon_n$ where $|a| < 1$ and ϵ_n is a zero-mean, purely random process with variance σ_ϵ^2. Derive a general expression for X_n in terms of the ϵ_n. Use this result to establish the covariance between the X_n and X_{n+k}. For large n, show how these only depend on their relative separation k.

4. Consider the univariate state space model

$$x_t = ax_{t-1} + e_t^m, \qquad z_t = x_t + e_t^o$$

where $\{e_t^m\}$ and $\{e_t^o\}$ are independent zero-mean Gaussian purely random processes with variances σ_m^2 and σ_o^2 respectively. Consider that $R_{XX}(k)$ to be known for the AR(1) model.

   (a) Derive and sketch the auto-covariance $R_{ZZ}(k)$. Comment on the role of σ_o^2, σ_m^2, and a in determining the shape of the $R_{ZZ}(k)$.

   (b) Derive and sketch the power spectrum $h_{ZZ}(\omega)$. Comment on the role of σ_o^2, σ_m^2 and a in determining the shape of the power spectrum.

5. Given a time series, clearly explain the two approaches that can be taken to obtain the power spectrum of the series. Outline the advantages of disadvantages of each procedure.

6. Given the univariate state space model

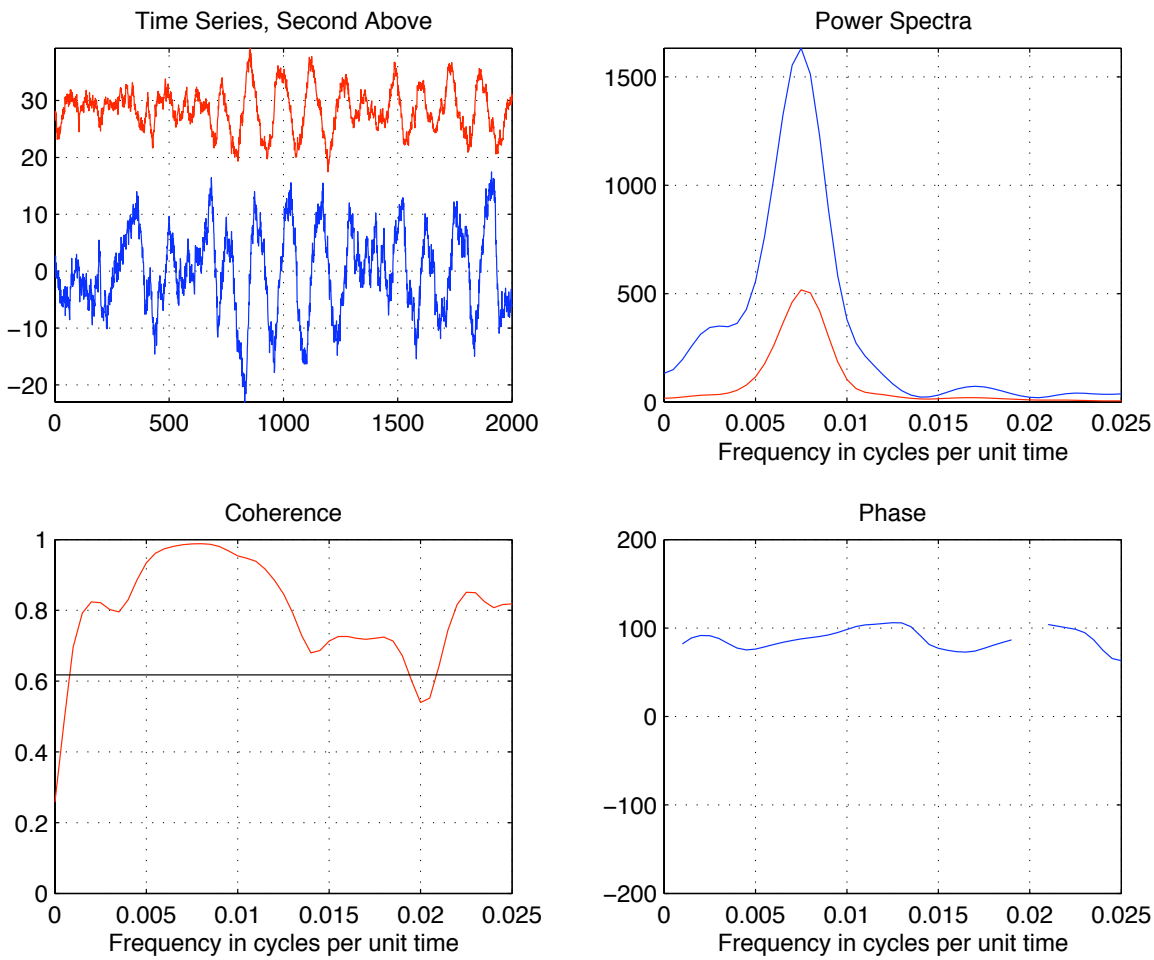
$$x_t = ax_{t-1} + e_t^m, \qquad z_t = x_t + e_t^o$$

where $\{e_t^m\}$ and $\{e_t^o\}$ are independent zero-mean Gaussian purely random processes with variances σ_m^2 and σ_o^2 respectively. Consider the situation for time τ where a predictive state $\bar{x}$ is available along with its variance σ_f^2. Suppose also that a measurement z is available. Set up and solve the regression to show that the filter state can be expressed as

$$\hat{x} = \left(1 + \frac{\sigma_o^2}{\sigma_f^2}\right)\bar{x} + \left(1 + \frac{\sigma_f^2}{\sigma_o^2}\right)z$$

Interpret this equation and consider the limits where $\sigma_o^2 \to \infty$, and where $\sigma_f^2 \to \infty$.

7. The figure below shows two time series (where time is measured in hours), as well as results from a cross-spectral analysis. Using these results, explain the properties of these time series.



8. Consider the AR(2) process: $X_n + a_1 X_{n-1} + a_2 X_{n-2} = \epsilon_n$. Write this as a multivariate AR(1) and find explicit expressions for the eigenvalues of A in terms of a_1 and a_2. Under what condition on a_1 and a_2 will the eigenvalues be complex? What do the existence of real or complex eigenvalues imply for this model?

9. Starting with the definition of the power spectral density, $h(\omega)$, in terms of the auto-covariance function $R(k)$, i.e.

$$h(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R(k) e^{-i\omega k}.$$

Express $R(k)$ in terms of $h(\omega)$. What happens for the case $k = 0$, and what does it mean?

10. Suppose you are given an observation set $(z_1, z_2, \ldots, z_N)$ and the multivariate nonlinear and non-Gaussian state space model

$$x_t = d(x_{t-1}) + e_t^m, \qquad z_t = h(x_t) + e_t^o$$

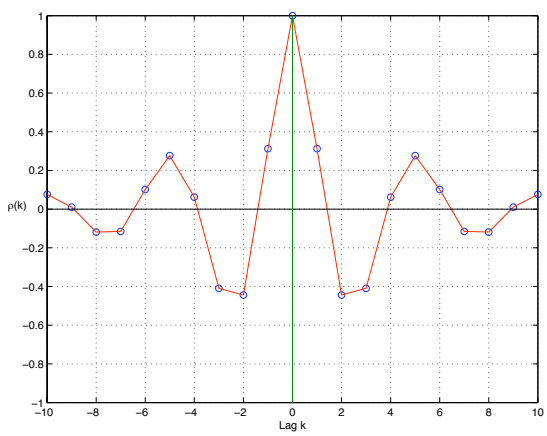
where $\{e_t^m\}$ and $\{e_t^o\}$ are independent zero-mean purely random processes. Outline the procedure used to carry out a single stage transition for the filtering problem (i.e. that is the transition of the system from time N to time $N+1$ during which a new measurement, z_{N+1}, becomes available).

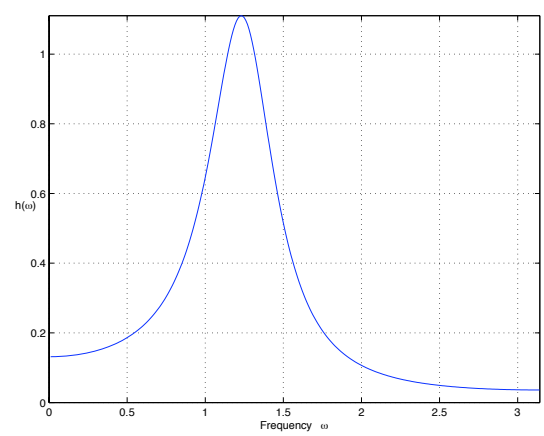
11. Consider the linear and Gaussian state space model

$$x_t = Dx_{t-1} + e_t^m, \qquad z_t = Hx_t + e_t^o$$

where x_t is $n \times 1$ and z_t is $p \times 1$. Suppose that we have an estimate for the filter density at time $t-1$ as $p(x_{t-1}|Z_{t-1}) = N(\hat{x}_{t-1}, P_{t-1})$. (i) Derive the mean and the covariance matrix for the forecast density $\bar{x}_t$, and (ii) Setup the regression to solve for the filter density, $\hat{x}_t$, and derive its variance-covariance matrix P_t.

12. The following figures shows the theoretical autocorrelation function (left panel) and power spectral density (right panel) of an AR(2) process:





(a) Estimate the variance of the process.

(b) Find the dominant period of the process using (i) auto-correlation function and (ii) the power spectrum

(c) Sketch a typical realization from this process being sure to highlight and comment on the main features.

(d) Explain the role of the auto-regressive parameters in setting the shape of the auto-correlation function.

(e) Explain the role of the auto-regressive parameters in setting the shape of the power spectrum