NCSU Department of Civil, Construction and Environmental Engineering

CE 796A Stochastic Methods in Water and Environmental Engineering – Spring 08

Section 001 : 01:30- 02:45 TH; Mann Hall, Room 425

Instructor : Dr. Sankar Arumugam (sankar_arumugam@ncsu.edu),
314A Mann Hall; (919)515-7700

Prerequisite : ST 515 or approval by the Instructor.

Topic 1: Predictor Identification in Environmental Problems

Introduction to Probability and Statistics – Probability Distribution Function (PDF) – Joint PDF, Moments, Conditional density – Transformation of PDFs – Random Variable Generation – Covariance, Correlation, - Partial Correlation – Mutual Information

Topic 2: Dimension Reduction in water and environmental data analysis

Issues in stochastic modeling with correlated environmental variables – Methods to identify dominant components – Simple averaging – Principal component analysis – Canonical correlation analysis

Topic 3: Periodicity Identification and Estimation with Environmental Variables

Theory of Expansion – Periodic and Non-periodic Functions – Fourier Series– Fourier Integrals and Fourier Transforms – Spectral Density Estimation –Coherence Spectrum – Multi-taper method

Topic 4: Model Development, Verification, Validation and Model Averaging

Model Development – Conditional density estimation – Multivariate regression – Nonparametric approaches – Kernel density estimation – Resampling - Ensemble Generation – Model Verification and Validation methods

Recommended Text: There is no single text book that can cover all these topics. I will provide photocopies from various textbooks. Some of them are listed below.

References :

1. Statistical Methods in Water Resources by D.R.Helsel and R.M.Hirsch (Available online for free: http://pubs.usgs.gov/twri/twri4a3/html/pdf_new.html)
2. Probability, Random Processes and Estimation Theory for Engineers by H.Stark and J.W.Woods, Prentice Hall.
3. Multivariate Analysis by W.Dillon and M.Goldstein, John Wiley & Sons.
4. Spectral Analysis and Time Series, M.B. Priestley, Academic Press.
5. Density Estimation for Statistics and Data Analysis, by B.W. Silverman, Chapman and Hall.

Course Purpose:

The primary aim of the course is to pursue a systematic approach in developing stochastic models that estimate the conditional distribution of the predictand, which could be any water and environmental attributes (e.g., precipitation, streamflow, concentration in a pipe distribution network) of interest. Students are encouraged to bring both problems and data from their own research area to work to learn different techniques.

Topic 1 provides an overview of methodologies for identifying predictors, influencing the predictand, from a given set of environmental data that varies both spatially and temporally. Naturally, these environmental variables exhibit significant correlation, both in space and time, which needs to be reduced to fewer components that represent most of the variance in the data. Topic 2 addresses this issue by emphasizing dimension reduction techniques based on principal component analysis and canonical correlation analysis.

Topic 3 provides an overview of methods that estimate the periodicities in environmental variables under spectral domain. Using these techniques, periodicities in predictors and predictands will be estimated along with their coherence in spectral domain. Topic 4 utilizes all the analysis discussed so far in developing stochastic environmental models using parametric and nonparametric techniques.

Course Objectives: By the end of the course, you should be able to:

a. Perform data analyses and identify predictors that influence the Water and Environmental (WE) decision variables of interest.
b. Apply diagnostic analyses to understand the relationships and dependence structure between predictors and the WE decision variables.
c. Identify the dominant components exhibited by the environmental variables using principal component analysis and canonical correlation analysis.
d. Estimate the periodicities and spectral coherence (between predictors and predictands) in the spectral domain for environmental variables of interest.
e. Develop stochastic models, using parametric and nonparametric modeling techniques, by employing principal components of predictors and predictands.

Grading:

Four Problem Sets	-	10%
Mid-term	-	10%
Final	-	10%
Project Presentations (Two)	-	10%
Project I	-	30% (Topics 1 and 2)
Project II	-	30% (Topics 3 and 4)

Problem Sets: Four problem sets will be given to consolidate the fundamentals on each topic. Typically, two weeks will be given to solve each problem set. Solutions for the problem sets will be provided once they are graded.

Tests: Two take-home tests, mid-term and final, will be assigned during the course. Typically, take-home tests will be assigned on Thursday and it will be due on the following Tuesday. Students are allowed to use any information they wish from anywhere, but they are strictly not allowed to discuss the material with anyone other than the Instructor.

Projects: Projects constitute the major portion of grading accounting 60%. Two projects will be assigned during the course and each student should work independently on both of them. A short presentation (15 minutes) needs to be given by each student on both projects. Project report should be submitted on the exact due date. It is ***strongly recommended*** that student select project topic on their research problem or area of research interest. Each student need to discuss and agree with the instructor on the project topic before the end of January.

Academic Integrity: Students should refer to the University policy on academic integrity found in the Code of Student Conduct. The policy can also be obtained at: http://www.ncsu.edu/policies/student_services/student_discipline/POL11.35.1.php Policies and procedures detailed in the above website will be strictly enforced in the class. It is the responsibility of the students to read it and follow those procedures in the class.

Office of Disability Services: This class will enforce all the guidelines related to services for students with disabilities. More information can be found at http://www.ncsu.edu/dss/. Students in need of service are requested to get permission from the Office of Disability Services located at 1900 Student Health Center, (919) 515-7653.

Tentative Schedule

Week	Topic
01/10-01/17	Review of Probability and Statistics • PDF, CDF, Moments • Joint PDF, Conditional Density • Transformation of PDFs • Generation of random variables from a given distribution
01/22-01/31	Predictor Identification in Environmental Problems • Estimation and Sampling Distributions • Correlation, Partial Correlation, Rank Correlation and Kendall's Tau • Nonlinear Measures: Mutual Information, Entropy
02/05-01/22	Dimension Reduction of Environmental Variables • Correlated predictors and predictands • Issues in stochastic modeling with correlated environmental variables • Multivariate Techniques – PCA and CCA
02/26-02/28 02/29	Project I Presentation Report Due before Spring break
03/01-03/08	Spring Break
03/06	Mid-Term Exam Assigned (Topics 1 and 2); Due on 03/11
03/04-03/28	Low frequency and High Frequency Variability in Environmental Variables • Theory of expansion, Periodic and Non-periodic functions • Fourier Integrals and Transforms • Spectral density estimation • Spectral density estimation of environmental variables • Coherence spectrum between predictand and predictors
04/01-04/18	Stochastic Modeling for Water and Environmental Problems • Estimation of conditional distribution of predictands • Mutivariate Regression, Transformation to Normality • Nonparametric methods – Kernel Density estimates, Resampling methods • Model verification measures – Bias, Mean square error. • Model validation – split sampling and leave-one out cross validation
04/22-04/24	Project II Presentation Project II report due on 05/06 – Last Day of final exam
04/24	Final Exam Assigned; Due on 04/29