41c A Graduate Elective in Statistical Methods for Mathematical Model Building
Gary E. Blau, Selen Aydogan, and George Applequist
Statistical Modeling and Quality Control has long been an integral part of the undergraduate chemical engineering curriculum at Purdue. This course is preparatory to the senior lab course. In many cases, Purdue Co-op and interns are sought out for process improvement assignments because of their skills in applying statistical methods. However, the course is only one semester in duration making it virtually impossible to cover topics in nonlinear statistical methods which are essential to modeling chemical reaction and other engineering systems.

Although most of the needs of the undergraduate students were being met, the graduate students, both foreign and domestic, had little exposure to statistical methods. An attempt was made to use a compressed short course to introduce them to linear experimental design and good laboratory practices. This attempt was spectacularly unsuccessful. They were unable to grasp the concepts in the time allotted and simple performed design and analysis by rote. It was decided instead to develop a full semester graduate course which would cover these basic statistical concepts and then introduce the tools and philosophy necessary for effective nonlinear model building.

The first eight weeks of the semester covered all the material in the undergraduate course including linear regression and design of experiments. Applied Statistical and Probability for Engineers by Montgomery and Runger was used as the text.[1]. A significant departure from the text was exposure to the use and abuse of subjective probabilities and Bayesian Methods. This discussion was woven throughout the lectures. After the eight weeks, there was general discontent among the grad students because of the "all the work required for an elective". Following this introduction, the topics of nonlinear parameter estimation, design of experiments for discriminating rival nonlinear models and design of experiments for improving parameter estimates were presented. Rather that taking the conventional approach to this topic, a Bayesian one was used in which Likelihood functions were used to generate joint posterior probability distributions for different models and their parameters from Subjective Prior Probability Distributions [2]. Monte Carlo Analysis were used for this purpose. However, classical nonlinear parameter estimation and nonlinear design of experiments concepts were also presented [3].

To familiarize the students to the software packages used in the class: Crystal Ball[4}, JMP [5]and Athena[6], a computer lab session was carried out in parallel with the lectures and projects. For each software package, a self-learning module was developed, each containing example problems and solutions. The main challenge in preparing the modules was to ensure that they were simple enough for everyone to understand yet broad enough to cover the basic aspects of the software package. This was particularly challenging when using Athena for nonlinear parameter estimation because of the paucity of help files.

The course was built around four projects. In each of these projects, the students were required to design and generate data from a simulated laboratory interface. The four projects were designed to teach the following concepts: (1) How to use Monte Carlo methods to generate posterior probability distributions for parameter estimates using Bayes Theorem. This provided a practical illustration of the uncertainty in parameters obtained by fitting models to data. (2) How parsimonious parameterization can be used to build a model consisting of a systems of ordinary differential equations. The students generated concentration time data and built a model describing the fate and distribution of an insecticide in an aquatic system containing fish and soil and plants. (3) How to determine the optimum operating conditions for a process using empirical models. The students sampled data from a simulated Dimethyl terephthalate transesterification plant consisting of a series of four CSTR's with 14 factors. (4) How to

Discriminate three potential Langmuir-Hinchelwood catalytic rate models for the reaction over a temperature input composition range for three reactants. Once a suitable reaction was found, D-optimality methods were required to minimize uncertainty in the parameters.

The students used a customized web site, access-controlled for identification of students, providing a form to enter independent variable values, a button to run, and a tabulation of the dependent variables (random error included). The underlying model functions, parameters, and error distribution were hidden. The values of parameters to be estimated and the selection from model alternatives were assigned to each student randomly. The web interface was written with Perl scripts which allowed closed form models, or using open-source numerical modules for matrix operations or root-finding. For acceptable web server load, preprocessing could be done offline, e.g. in Mat Lab[7], to calculate intermediate results and save them to an input file associated with each student. After project completion, the instructor had the expected values of parameters unique to each student, plus logs of input and output values for use in grading. Since each student is working on his/her own problem cheating is eliminated and interaction is encouraged.

The course also used Quickplace, a Lotus web server product, providing access-controlled readership of documents and distributed maintenance responsibility with little skill required. Course staff could rapidly post documents organized into several levels, such as lectures, labs, homework, exam solutions, etc. Students could post questions for discussion, consolidating typically redundant conversations into a centralized reference collection. Quickplace overcame the disadvantages of email: excessive broadcast of information lacking organization, inability to correct files once they are sent, size constraints, and reliability, to name a few. The differential access capability also enabled students to submit their work and receive feedback privately.

[1] D.C.Montgomery and G.C.Runger,"Applied Statistics and Probability for Engineers,Third Edition, John Wiley Inc., 2003. [2] P. M. Reilly and G. E. Blau, "The Use of Statistical Methods to Build Mathematical Models of Chemical Reaction Systems," Can. J. Chem. Eng., vol. 52, pp. 289-299, 1974. [3]P. J. Englezos and N. Kalogerakis, Applied Parameter Estimation for Chemical Engineering. New York: Marcel-Decker, Inc., 2001. [4] Decisioneering, Denver, Colorado. [5]SAS Institute, Cary, North Carolina. [6]Stewart and Associates Engineering Software Inc, Madison, Wisconsin [7]Mathworks Inc., Natik, Massachusetts