CCSU

DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Thursday, November 8
12:15 – 1:15 PM
Maria Sanford Hall, Room 204

MODELING OF FLIGHT DELAYS

RAJIV SAMBISAVAN

(Data Mining MS Thesis Presentation)

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: This thesis develops models for flight delays by major airlines at domestic destinations. Federal regulations require airlines to meet specific revenue criteria regarding their on-time arrival performance, which includes the cause of delay (if any). Causes of delay include National Aviation System (NAS) delay, Carrier delay and Late Aircraft delay. The number of records of the on-time performance dataset for a single year in recent times has exceeded five million records and poses challenges to develop comprehensive models of flight delays. This study develops such models by using sampling techniques and summarized views of the data. The large individual flight dataset was sampled to ensure that delays at all domestic destinations by all carriers are represented. Monthly performance summaries published by the BTS was the other dataset used to develop models for this study.

The CART (Classification and Regression Trees) algorithm was used to develop the models for this study. The CART algorithm is non-parametric and does not impose specific restrictions on the distribution of errors from the model. This enhances the robustness of the results. This thesis used both the classification and regression aspects of the algorithm. The classification aspect of the algorithm was used to develop the models that predicted the on-time arrival or delay (arrival delay of greater than fifteen minutes) of individual flights. The regression tree aspect of the algorithm was used to develop models that predicted the proportion of total delays and delays from specific causes.

Results of the modeling include the following. The extent of departure delays is the most important predictor of arrival delays. NAS delays are the most important predictor of departure delays. The particular airline is the most important predictor of service performance to a particular destination. There is a seasonal aspect to delays, that is, summer and winter months are associated with higher delays. Finally, delays for airlines are dominated by specific causes (NAS delays, Carrier Delays or Late Aircraft Delays). Future work should explore finding better predictors of delay, detailed analysis of interesting subsets uncovered by this study and alternative algorithms for model development.

For further information:
gotchevi@ccsu.edu 860-832-2839
http://www.math.ccsu.edu/gotchev/colloquium/