Predicting Low Strength Properties of Wood Composite Panels using Bayesian Logistic Regression

Xia Huang[1], Timothy M Young[2], Yan Zeng[3], Frank Guess[4]

[1] Research Associate, Center for Renewable Carbon, University of Tennessee, Knoxville, xhuang8@utk.edu

[2] Professor, Center for Renewable Carbon, University of Tennessee, Knoxville, tmyoung1@utk.edu

[3] Former Graduate Research Assistant, University of Tennessee, Knoxville, Department of Statistics, Operations, Mgmt. Science

[4] Professor, University of Tennessee, Knoxville, Department of Statistics, Operations, Mgmt. Science

Abstract:

Real-time process and destructive test data were collected from a wood composite manufacturer in the U.S. to develop real-time predictive models of two key strength properties (Modulus of Rupture (MOR) and Internal Bound (IB)) of a wood composite manufacturing process. Sensor malfunction and data "send/retrieval" problems lead to null fields in the company's data warehouse which resulted in information loss. Many manufacturers attempt to build accurate predictive models excluding entire records with null fields or using summary statistics such as mean or median in place of the null field. However, predictive model errors in validation may be higher in the presence of information loss. In addition, the selection of predictive modeling methods for low strength properties of wood composite panels poses another challenge to many wood composite manufactures. In this study, the expectation-maximization (EM) algorithm and multiple imputation (MI) using Markov Chain Monte Carlo (MCMC) simulation were developed for missing data imputation, and achieved more precise results, compared to mean/median substitution, simple random imputation, or the last-value-carried-forward methods. A Bayesian logistic regression based on imputed dataset was constructed to predict low strength property of wood composite panels. Based on the summary statistics of MOR and IB, the first quantile observed values were used as target wood panels with low strength. Principal component analysis (PCA) was used to remove the colinearity effects among predictors. Using ten-time random repeated sub-sampling validation, the Bayesian logistic regression, on average, had a correct classification rate of 73% for MOR, and 80% for IB.