

**WHAT BREEDERS CAN LEARN FROM NCVT
SPINNING TESTS – A NEW DATA MINING APPROACH
C.D. Delhom, J.H. Campbell, and D.P. Thibodeaux
USDA-ARS
Southern Regional Research Center
New Orleans, LA**

Abstract

The National Cotton Variety Tests generate a vast amount of fiber and yarn data each year. Data mining provides methods for searching large amounts of data for meaningful patterns. Data mining has become vital in a large number of industries to support decision-making in the face of rapidly growing quantities of data. The laboratory and machine measurement techniques used in the NCVT allow for patterns to be discovered which may be of use to breeders. Data mining techniques can aid the development of tools to provide for the rapid assessment of cotton “quality” with a minimum number of tests and observations. The NCVT provides a convenient dataset to allow for a variety of data mining techniques to be investigated for their suitability to the problem. Data mining techniques such as artificial neural networks have been used successfully to solve textile related problems. Simpler data mining techniques are investigated to allow the end user to better understand the problem solving method. Data mining techniques such as clustering methods and Boolean expressions are examined.