H.Q. FAN. ***Data mining and statistical analysis of construction equipment failure.*** *Gerontechnology 2012;11(2):320;* doi:10.4017/gt.2012.11.02.629.00 **Purpose** Construction equipment is a key resource, and contractors that own a large equipment fleet take all necessary measures to maximize equipment utilization and minimize equipment failures. Although most contractors implement scheduled maintenance programs and carry out periodic inspections and repairs on their construction equipment, it is still difficult to predict the occurrence of a specific failure of a piece of equipment in the short or long term. According to a survey in the United States, approximately 46% of the major equipment repairs was undertaken as a result of an unexpected failure. Although it is not possible to predict all failure events, a slight improvement in their prediction represents a significant saving in time and cost for a large contractor. Statistical power law models and data-mining models were compared to investigate their pros and cons in predicting critical failure events of heavy construction equipment. **Method** With large amounts of equipment failure data accumulated in a surface mining project, two different types of failure models were created for comparative analysis from a practical point of view. For selected equipment units, failure data were collected along with the relevant factors which may cause variations of equipment failure rate (or mean time to failure). In a classical approach, Power law models of equipment failure rates are fitted using RGA 7.0; while in the data-mining approach, the mean time to failure is modeled using a data-mining algorithm-decision tree induction, establishing logical, mathematical, and statistical relations between MTTF (Mean Time Between Failures) and its various factor of impact (equipment conditions, failure history, environmental conditions, etc.). Both models are used for validation tests on randomly selected time periods and compared in terms of their performance. **Results & Discussion** The two types of models were compared (*Table 1*).

Keywords: equipment management, failure analysis, data mining, decision support
Affiliation: The Hong Kong Polytechnic University, Hong Kong; *E:* bshfan@inet.polyu.edu.hk
Full paper: doi:10.4017/gt.2012.11.02.629.628

Table 1. Comparison of statistical Power law models and data mining models

Item	Statistical power law models	Data Mining
Methodology	Hypothesis-and-Testing	Model Inference From Data
Process	Model the statistical distribution based on collected failure data, determine model parameters of shape, scale and location. Check fitness of the derived models	prepare data, feed into data mining algorithm (decision tree induction), obtain, visualize and validate the data mining models
Pros	Simple model, easy to use and understand	Account for a wide variety of factors of impact on failures, apart from the elapse of time
Cons	Accurate when time is considered as the predominant factor with constant wear and tear of equipment over time: random failure with time-dependent frequency of failures.	Model is complex, higher requirements on data collection, may be inaccurate or biased when data collected are incomplete or noisy.