

Data mining: Searching for patterns

- Critical Issues

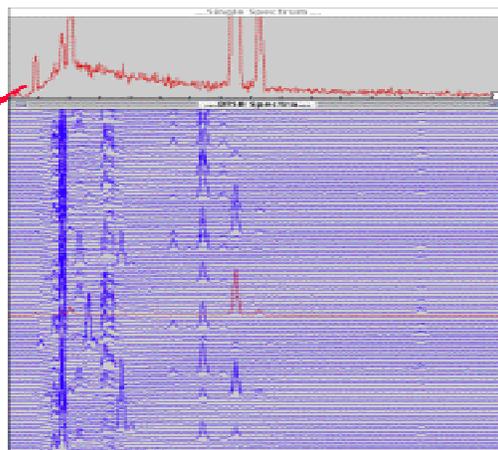
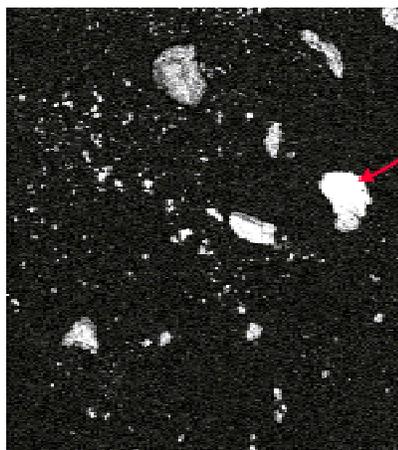
- The searching of data resulting from high-throughput experimentation often involves examination of large data sets with hundreds to thousands of features. Each feature may have one or more morphological and/or spectral images in addition to other elemental/chemical data such as x-ray spectra. We wish to classify these features using these data, where their spectra, morphology or chemical signature may not be known ahead of time.

- Research Strategy

- Our research involves development of software to automate the collection and integration of such data sets, and the development of data analysis tools for rapid visual surveying and searching. These tools are developed for Macintosh and PC platforms.

- Research Highlights

- One of the tools we are developing is the Poly Plot Package (PPP), which displays hundreds of x-ray spectra in a cascade plot, for ease of comparison. Analogous to a postage stamp display for images, the spectra can be 'clicked' to select, mark and display spectra. An individual wavelength can be clicked, and the spectra sorted by amplitude at that one wavelength. Adjacent spectra in a sorted plot can be selected and placed in a group -- these spectra will be similar if selected properly. Two or three dimensional scatter plots of amplitudes of spectra at given wavelengths (representing chemical elements) and chemical and morphological parameters also serve for comparison and selection of groups of features. All plots and images are linked so that groups of features selected in one display mode are also labeled in all other images and data sets. Together the cascaded and scatter plots, the images of the selected groups of features, and the summary statistics of each group, aid the researcher in classifying patterns in the features and finding the features of interest.



For more information ...

David Bright, Surface and Microanalysis Division