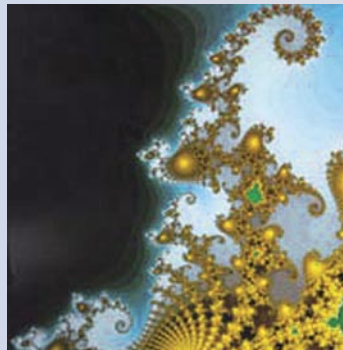
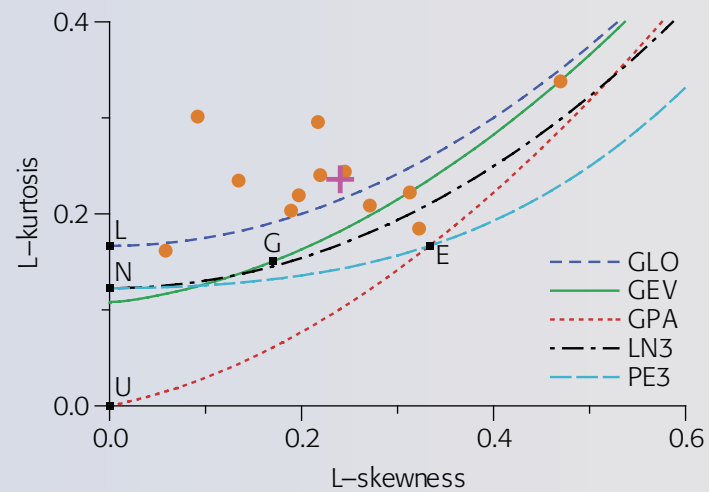




IBM Research has made fundamental contributions in statistics and data mining. For over a quarter century, an active statistics research program has explored a broad range of issues in theory and practice. The pioneering work of Benoit Mandelbrot on self-similarity (fractals) and long-range dependent statistical models has had significant impact on many scientific disciplines, including hydrology, finance, and communications network and computer system analysis. Analysis of time-dependent data and non-standard distributions is another influential area of IBM's statistics research. An example is L-moments distribution theory that led to innovative statistical methods for characterizing and estimating distributions, especially of heavy-tailed data in finance, risk management, and IT-system monitoring. Leadership in knowledge discovery and data mining (KDD) research was established in the 1990s by Rakesh Aggrawal's introduction of association rule mining. IBM's other major contributions in KDD include mining of excessive information stream throughput with lightweight data analysis techniques, high-performance mining techniques in parallel execution environments, and pioneering the area of privacy preserving data mining.



With the explosive growth of online data and IBM's expansion of offerings in services and consulting, data-based solutions are increasingly crucial. Accordingly, methodological development for business intelligence, as well as IT-system and business process monitoring, has become a focal point of statistics and KDD research at IBM. In these areas, monitoring data that has been collected over time is used to make processes more efficient, effective, predictable, and profitable. Challenging aspects include handling large time-dependent data with varied characteristics, producing accurate and practical forecasting methods, and developing analytics relevant for business decision-making. Two specific problems that IBM Research is currently addressing, for example, are customer targeting and business metric forecasting.

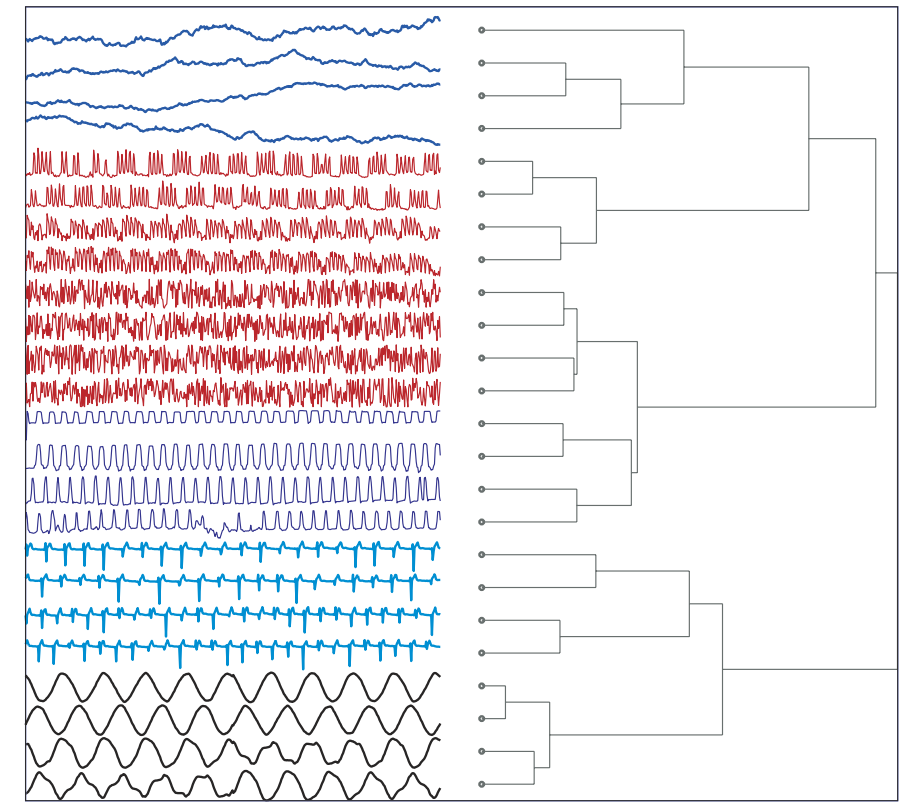


**CUSTOMER TARGETING**

In customer targeting, Research has developed tools to help identify companies with higher propensity for business engagement or with more promising profitability. The analytics in the tools are statistical/KDD methods based on models that incorporate a broad range of explanatory variables extracted from both financial time-series and corporate data, and are trained against a carefully filtered set of historical examples. The results of the analysis, such as customer scores or ranking, are delivered through a Web-based tool that also provides detailed company information to business executives in an integrated framework.

**BUSINESS METRIC FORECASTING**

Business metric forecasting methods consist of advanced modeling, prediction, and inferential analytics, and forecast various quantities important in business planning and decision-making, such as revenue, demand, cost, utilization, yield, and outcome. IBM Research introduced a set of statistical forecasting techniques appropriate for business problems. These model-based techniques incorporate various sources of relevant information, such as economic/industry/technology trends, business cycle seasonality, and opportunity pipelines. Forecasts for different horizons are issued based on the nested structure time-series model, producing accurate point and interval forecasts. The same forecasting approach has been used effectively in IT-system monitoring and planning problems.



Stream Clustering Classification

