Plant Optimization Using Chemometrics

Historical data from a process plant can be mined to find production patterns that reliably optimize operating performance around cost, productivity and safety.

If you can't measure it, you can't control it. Multivariate plant profiles give us a way to measure (and therefore control) process efficiency.

Chemometric methods increase the speed and reliability of translating process data into information that can be used to guide plant decisions. Although only recently applied to the plant as a whole, the techniques described here are well-established means of automating the interpretation of multivariate data from sensors (like NIR) on a process line.

Controlling a manufacturing process involves more than monitoring process parameters in isolation. There are interactions among pressure, temperature, flow, level and chemistry that define the overall plant performance. These interactions are further affected by factors such as the source or quality of the ingredients, position in the lifecycle of any catalytic or cleanup beds, and corrosion or fouling of reaction vessels and heat exchange units. Not only do all these factors have local impact, but in multistage processes each successive stage will be to some extent dependent on parameters in use in previous (sometimes even subsequent) zones.

The technology used in evaluating plant data as described in this overview is referred to as pattern recognition, data mining or (when applied to chemical data) chemometrics, and is based on multivariate statistics. The purpose is to extract process-relevant patterns of association among the typically highly-correlated control measurements, using these patterns to build a fingerprint of the system as a whole. The procedure automatically builds a reference set of plant profiles; we use this previous plant experience to decrease operating cost, forecast production yield, monitor process consistency and schedule maintenance.

Lots of data, no one looks at it

We have access to a lot of data generated by sensors on the process line. Plant personnel have little time to spend examining these data closely; they typically trend only a few key parameters. Beyond simple trending, this is an information-rich data set. A ten month history of plant output in Figure 1 shows variability driven by target production levels based on market demand coupled with downtime, scheduled and unscheduled.

These different throughput levels let us see how the plant operates under a variety of normal and upset conditions. The data is sufficient to extract the role of process line efficiency in the production variation. To determine efficiency's effect, we assemble the raw process data for an exploratory data analysis to quantify multivariate trends and groupings. The plant tends to operate in modes that can be characterized using these statistical tools. Results shown in the following pages present the use of all the data to build a better working knowledge of the plant.

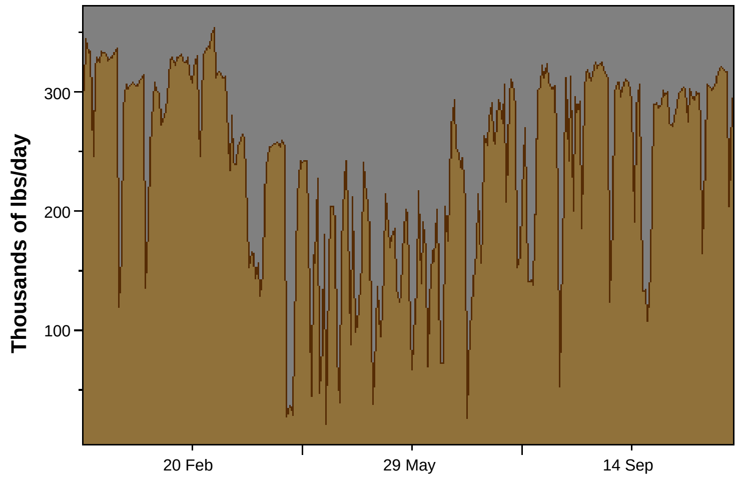

Figure 1: Production levels (in thousands of pounds per day) for the first 300 days of 1996.

Multivariate Plant Profiling

We define daily plant profile as the assemblage of all measurements collected on plant processes each day. The multivariate technique Hierarchical Cluster Analysis (HCA) compares profiles pairwise and constructs a graphic which groups days on the basis of their similarity. The result, shown in Figure 2, reduces thousands of pieces of data to one, very informative picture.

Each day of plant operation is represented by a terminus on the left. Days with similar profiles are linked by vertical lines to form clusters. The horizontal length of the branches connecting clusters indicates degree of dissimilarity. Thus, the daily profiles in the top half of the dendrogram are significantly different from those in the bottom half. The dendrogram consists of four main clusters, indicating that the plant has four distinct operating regimes, which are differentiated by the colors yellow, red, green and blue. We can use the HCA results to interpret the clusters and better define the driving forces behind this division in plant profiles.

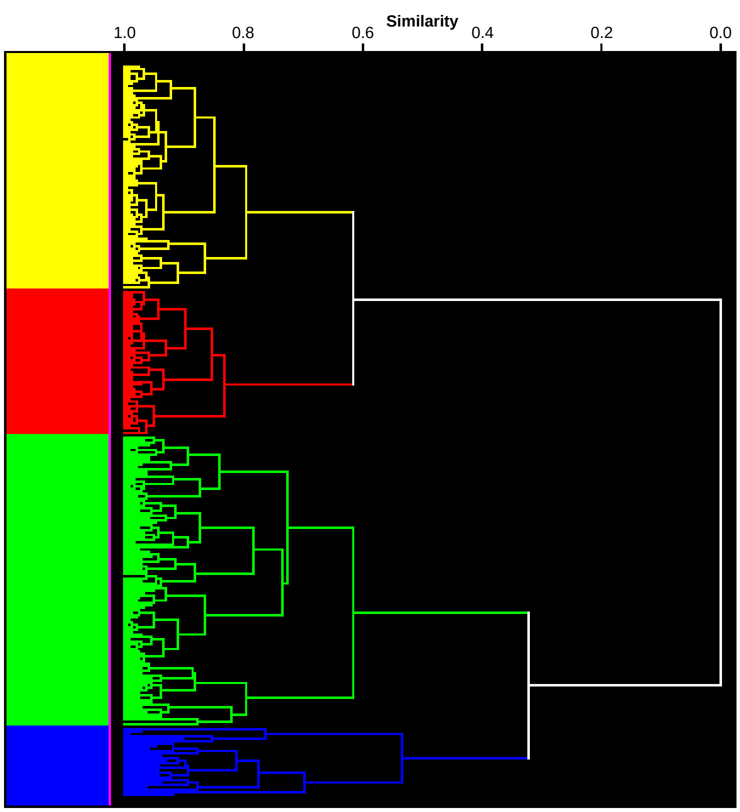

Figure 2: Daily plant profile groupings color coded to show distinct operational regimes.

Applying the profile colors to the daily production data (Figure 3), we can attach a description to each of the four modes of plant operation. The variation in plant profile, as established by HCA, basically follows production level, with red and yellow representing high production, moderate production in green, and low production blue. There is clear differentiation (*i.e.*, two clusters) in the high-production red and yellow groups which separates in time from the first quarter of the year to the third, showing change in how the plant was run that is not dependent on production.

When searching for patterns in data, we employ a second technique, Principal Component Analysis (or PCA), a method which focuses on variation in data instead of HCA's emphasis on similarity. Again, all plant measurements are used to build a plant profile that is represented by a single point. The relationship among plant profiles are revealed by the relative distances that separate points. Such a plot is shown

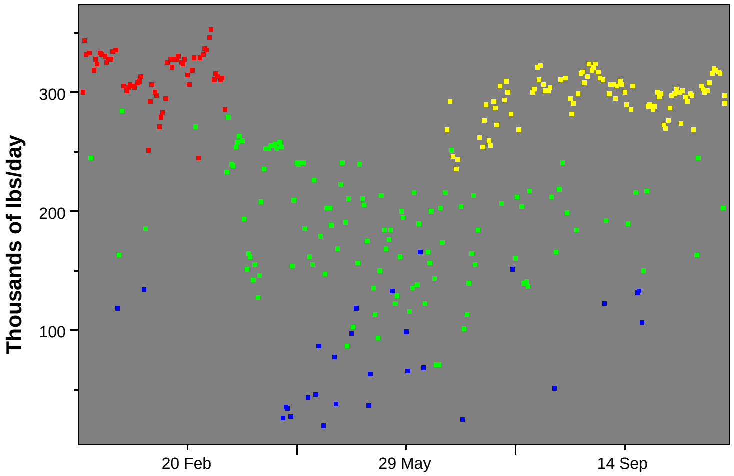

Figure 3: Daily production level color coded by HCA cluster.

in Figure 4; using the same color coding as defined by HCA.

Each day is a single point and close points indicate plant profiles with common patterns. The low, intermediate and high production days group in Figure 4 in the same way as in Figure 2. The large spread in the low-production blue samples is driven by process and chemical fluctuations experienced during production start-up and shutdown. The plant operates within a narrower chemistry realm as production increases (from green to red or yellow). The tight red and yellow regions indicate that the plant runs consistently at high production levels.

If we focus in on only the high production samples, it is possible to get a closer view of similarities in the process fingerprint and better see the distinction between high production plant profiles early and later in the year.

Figure 5's view contains only high production data and the subclustering seen in the plot corresponds to the two families apparent in the HCA view earlier. As seen by the labels, the yellow group is comprised of process profiles collected between mid-June and late-October. The red group represents plant profiles in the first part of the year. Seeing structure in this high-production data is useful because there are significant cost differences associated with running the plant in these two distinct ways. We note that several process variables are different between the first quarter and third quarter plant profiles. The two clusters are characterized by a difference in flow, pressure and temperature.

Because of the relationship between general plant operating data and production, we can look to see how well we can predict production based on the plant profile. In this analysis, we assemble all of the process variables that are not direct measures of production such as pressure, temperature, flow and chemical uptake, and prepare a calibration model using a mathematical treatment called Partial Least Squares (PLS). We use PLS to predict what the production level should be and compare this calculated value to actual production. The results of such a projection are shown in Figure 6. There is a strong connection between the process variables being monitored and the production level of the plant. This is confirmation that the process variables being monitored are appropriate and that the data is being collected in a consistent and conscientious fashion.

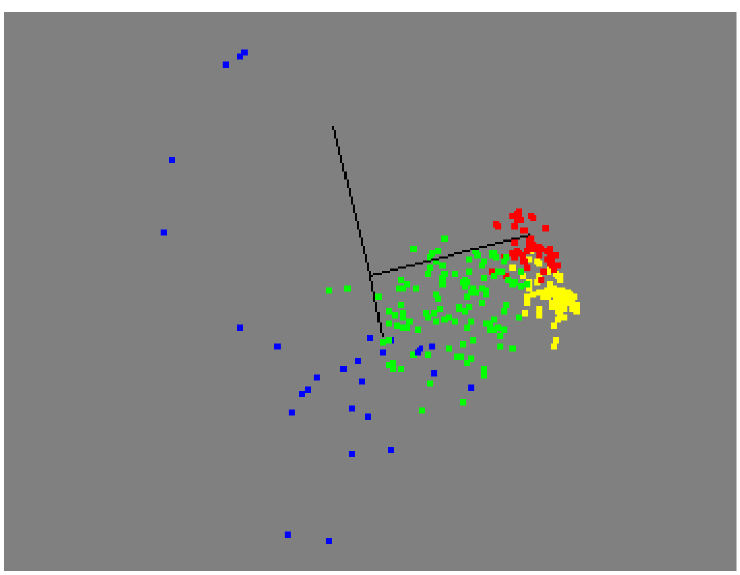

Figure 4: A 3D plot of PCA results for the plant profiles. The low production blue profiles are scattered due to the diversity in plant operation during start-up and shut-down.

The PLS results also find plant profiles where pro-

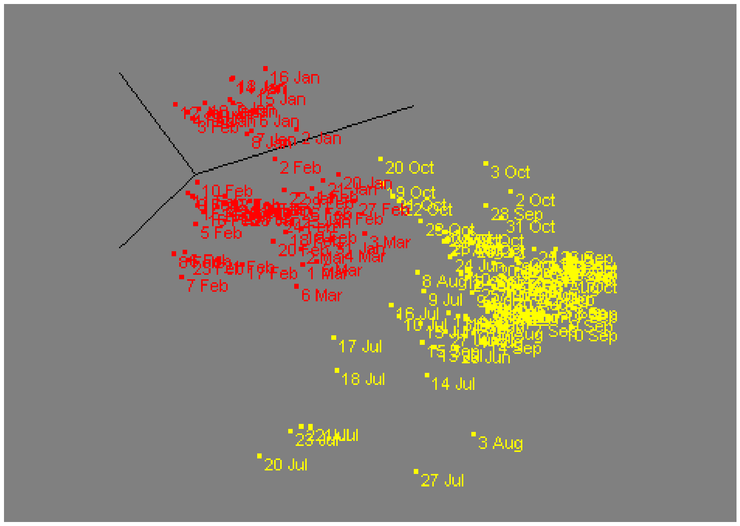

Figure 5: An interrelated set of changes in pressures, temperatures and flows caused a shift in plant profile between first and third quarters, 1996.

duction levels don't fit the expectation. These relate to aberrant behavior such as plant upsets or sensor malfunction. Note that there is significant overlap in the production levels achieved by different colored groupings. For example, some of the "low production" blue group actually show higher production levels than "moderate production" green days. This means that there are some days where the plant profile at 170,000 lbs of production is more similar to the profile of a 50,000 lbs/day rate (blue) than it is to the optimal (green) operating conditions for the 150,000 to 250,000 range. The same relationship is in place in the boundary between the intermediate greens and the "high production" yellow-red days.

What is important is that *we are able to use the existing general process data to develop a reliable library of plant profiles for the facility that allows us to better match the current condition to similar events in plant history.* This match can be done continuously and in real-time. The chemometrics makes archival and retrieval of plant information much more efficient making the experience set useful in controlling the process.

Summary

In the process world, we deal with an arena where data is plentiful but deriving information from this data takes time. In the past, we responded by selecting subsets of the entire data mass and drawing inferences or conclusions based on what we hoped was a collection of representative measures. Today, with the push toward computerized plant historians (*i.e.*, long-term storage of plant historical data), chemometrics provides a vehicle to take a comprehensive approach to mining information from process data.

Pattern recognition methods can be used to improve the flow of information from a process line. Through this, a manufacturing site can achieve an increase in the speed and reliability of translating plant data into process-specific controls. Both the software tools and the support service are in place to incorporate multivariate data mining and routine analysis into a real-time system.

You can't begin to control something you can't measure. Chemometrics allows you to extract and measure process quality parameters that are buried in the myriad data points flowing through a processing plant. More efficient and cost-effective control is the result.

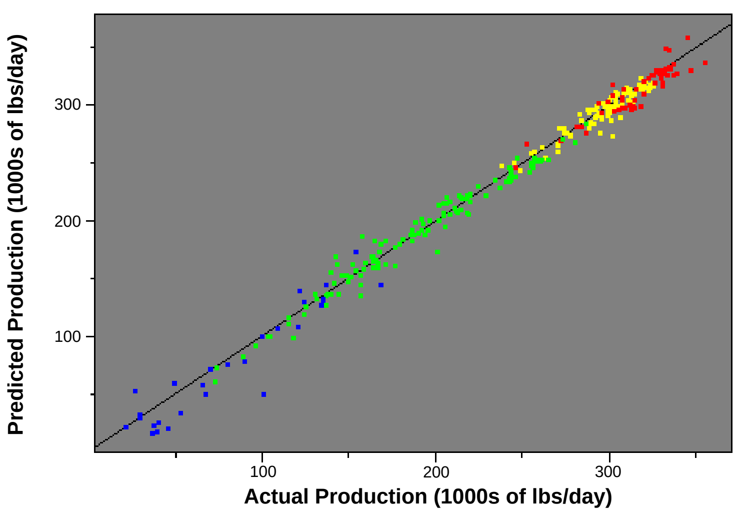

Figure 6: Relationship between general operating parameters and production. Overlap of the colored groupings flags that the plant can be run in several ways for a given level of production and, therefore, optimized by multivariate means.

Infometrix, Inc., P.O. Box 1528, Woodinville, WA 98072 - phone: (425) 402-1450, fax: (425) 402-1040, e-mail: info@infometrix.com