50-1118

Chemometrics

Application Note

Early Detection of Process Upsets via Multivariate Modeling

PCA Calibration for continuous process monitoring
Performed in coordination with Northwest Analytics (NWA)

Background

In the chemicals manufacturing industry, measurement of process parameters provides process controllers and engineers insight into the monitored process. These parameters tend to be monitored individually by upper and lower control limits. However, single-variable thresholds are sometimes insufficient for identifying process upsets, particularly if they are not directly correlated to a specific measurement. For example, a pump overheating is easily detectable when monitoring pump temperature and pressure, but these sensors may not identify pump vibrations or reduced pump efficiency on their own. In significantly more complex systems it may be even less likely that these univariate boundaries can identify complex process dependencies or upsets. In these cases, multivariate analysis can not only detect the upset, often much earlier than a univariate analysis, but can also facilitate an analysis leading to identification of the root cause of the upset.

While investigating a customer's vacuum pump data, Infometrix demonstrated that multivariate analysis can identify system upsets earlier than with univariate monitoring. As discussed below, Infometrix identified issues in a vacuum train a week before on-site personnel identified a potential issue from vibrational measurements. Additionally, by identifying which specific measurements were contributing to the upset, the multivariate model was able to diagnose where the issues were arising, providing concrete data behind what had previously been an operator’s hunch. The model produced was deployed for immediate use by the customer and can identify upsets similar to this in the future.

Often, a multivariate steady-state model (for example, the one developed for this project) can catch other process upsets, not only the obvious upsets that a univariate analysis might detect, but also subtle or slow-moving changes that cannot easily be identified by traditional upper- and lower-bound control charting.

Introduction, Data Description, and Data Processing

Process monitoring of linked systems with multiple sensors and indicators is a complex task. When one sensor is able to monitor an entire process, only that one sensor needs to be monitored. This can be accomplished with ease by monitoring the system on a control chart, utilizing upper and lower control limits to flag process upsets. Two sensors could be monitored with two similar charts. But as systems grow more complex, and when sensors monitor inherently correlated measures of the system, keeping track of multiple univariate charts or the multitude of variable-by-variable charts becomes an untenable undertaking. While univariate control methods can miss subtle, correlated variations that can flag process upsets, multivariate analysis offers statistical methods for revealing inconsistencies in the correlations between variables found in complex processes. Due to this advantage of multivariate analysis, the process monitoring industry is moving towards integrating multivariate analysis into their univariate control systems by working with providers like Northwest Analytics (NWA) to bring these new, advanced capabilities to robust quality control packages.

A customer undertook this project with Infometrix and NWA to determine if integrated multivariate solutions would be able to help with the understanding and monitoring of their manufacturing environment. Pump data was provided with 4-hour averages from January 2017 through September 2018 for a variety of measurement points, including pump temperatures, levels, pressures, etc., totaling 3806 time points (“samples”) with 26 system measurements (“variables”). Approximately half of these data were removed during times when the pumps were off.

Model Calibration and Testing

Commonly, the multivariate tool used to characterize a complex system is PCA (principal components analysis). PCA works by considering all variables at once, re-characterizing the original data as linear combinations of variables, then recasting the samples as 'scores', that is, coordinates in the new space of those linear combinations.

Ideally, a PCA model is built from a dataset that is homogeneous in PCA Scores space. This would indicate that the model is built from a normally distributed set of samples that represents proper working conditions. However, even after removing the data points collected when the pumps were off, we see some clustering in the data. This indicates that we are including data from time points when operating conditions were “different.” In Figure 1, we can see a large central group of samples (highlighted), but also two smaller groups around it.

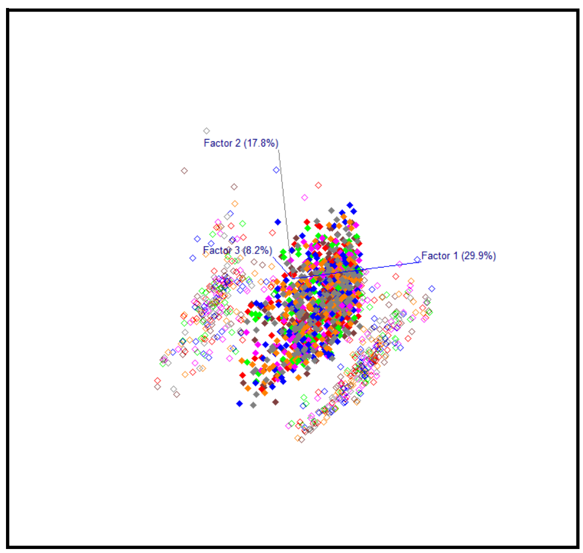

Figure 1. PCA Scores plot of all time points when all pumps were on

Looking at a time-plot of principal component 1 below in Figure 2, we can see highlighted in yellow the time ranges that are part of the central group. The two outside clusters then represent time points in early 2017 and late 2017, when some subtle differences appear to have been present compared to “normal” operations (the grey sections of the plot represent time points that have been excluded from the data subset because the pumps were not on).

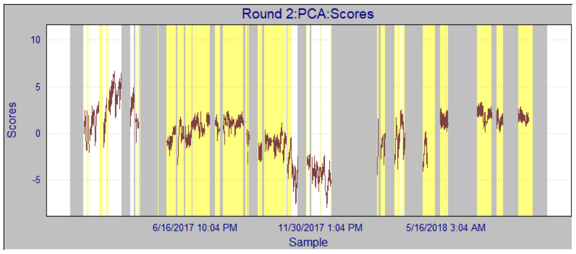

Figure 2. Time plot of principal component 1 of all time points when all pumps were on. Gray regions are when pumps were off. Highlighted points (yellow bands) are part of central cluster from Figure 1

This aligns with information provided by the client, which indicated that there were some pump issues identified in both early 2017 (requiring a component replacement in March) and late 2017 (first detected on August 30th, 2017 but finally requiring a component replacement in December 2017). Excluding these outlying points and performing another PCA plot (below, Figure 3), another small cluster is seen as separate from the highlighted central group.

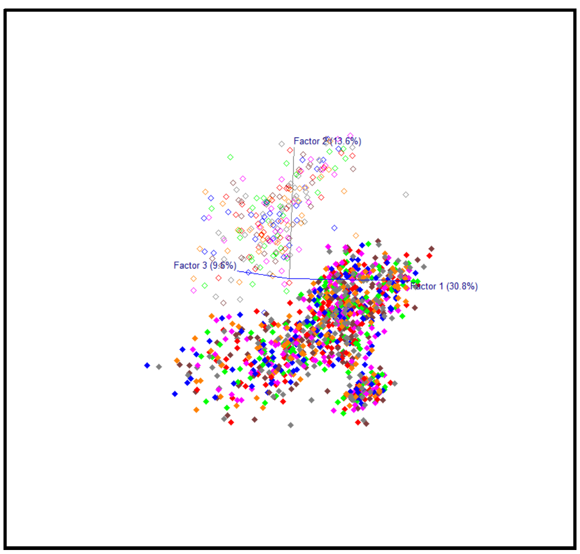

Figure 3. PCA Scores plot of inliers identified from Figure 1

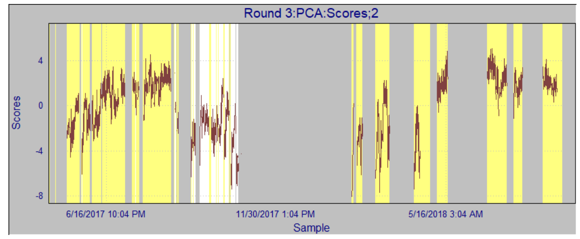

Figure 4. Time plot of principal component 1 of inlier points identified from Figure 2. Highlighted points are part of central cluster from Figure 3

These samples not highlighted in Figure 3 correspond to time points from September and October 2017 (see Figure 4, above), after the pump issues were identified. This indicates that there was a slowly increasing deviation from normal operating conditions in the system beginning some point in late August. This was confirmed by the customer's description of a discovery of vibrations in the pump system by staff on August 30th, 2017.

After once again removing outlier samples and running yet another PCA, a scores plot with minimal PCA clustering results (see Figure 5, below). This final dataset represents timepoints when:

- All pumps were on
- The initial PCA outliers from early 2017 and late 2017 were removed
- The subsequent PCA outliers from September and October 2017 were removed

As far as Infometrix is aware, the timepoints remaining are when the pumps were operating in a standard manner. This "steady state" PCA model can subsequently be used to identify when there are potential issues occurring with the system.

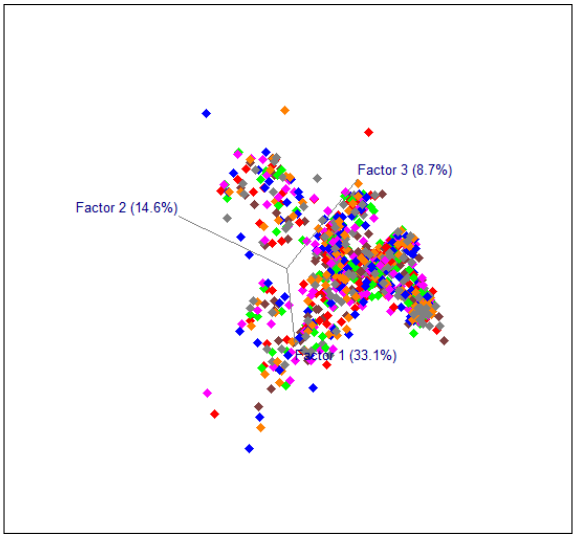

Figure 5. PCA Scores plot of final model

When predicting with a multivariate model, each sample generates outlier diagnostics; two such diagnostics--Mahalanobis Distance and F-Ratio are shown below. We look at these plots to observe when the diagnostics deviate significantly from the baseline. Higher values represent unusual combinations of measurements (not necessarily higher temperature or pressure, but deviations from the norm).

We can also use the final PCA model, with outlier samples removed, to evaluate the times (samples) when these outliers were included, as a way to possibly understand process excursions at those times. We projected the full dataset (all data when pumps were on) onto the PCA model to see if it was possible for the model to identify potential upsets. In Figure 6 below, we can see clear periods of stability, as well as periods of significant deviation from ideal (the brown trace is Mahalanobis Distance, the green trace is F Ratio).

There are two times when we see very large outlier diagnostics. Early on, beginning in late February, the Mahalanobis distance jumps up high, and stays high until late March when maintenance was performed on the system. This is followed by a period of stability, with occasional spikes but nothing sustained.

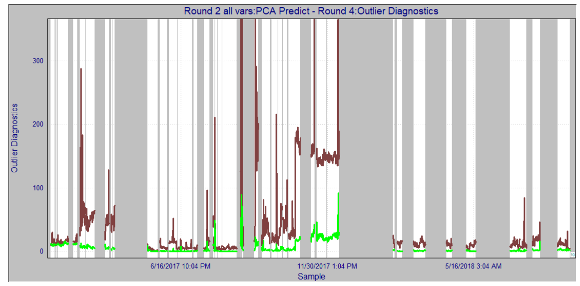

Figure 6. Outlier diagnostics for all samples when pumps were on

Finally, in late August, both outlier diagnostics spike very high. Operations continue with high Mahalanobis distance for a couple months, but then both F Ratio and Mahalanobis distance spike in November and sustain high values until the system is taken down for repairs in December. Figure 7 shows a zoomed-in window for this second upset, which is identified via these outlier diagnostics on October 24th, 2017, a week before the vibrations were noted on-site.

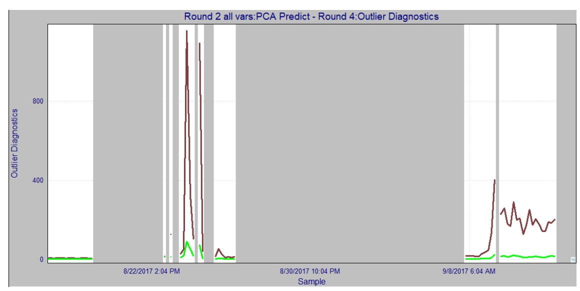

Figure 7. Outlier diagnostics zoomed in for late August / early September 2017

After exporting the model and loading the data into NWA's Quality Analyst software package, we can see below in Figure 8 the alarms in both metrics from early 2017 and late 2017 clearly. We can also see, for the highlighted point in the bottom T-squared plot from August 24th, the variables that most contributed to high outlier diagnostics. Fine-tuning of alarm levels may be necessary, but the early warning from August 24th clearly shoots out well above all other samples over the 20-month history of data provided.

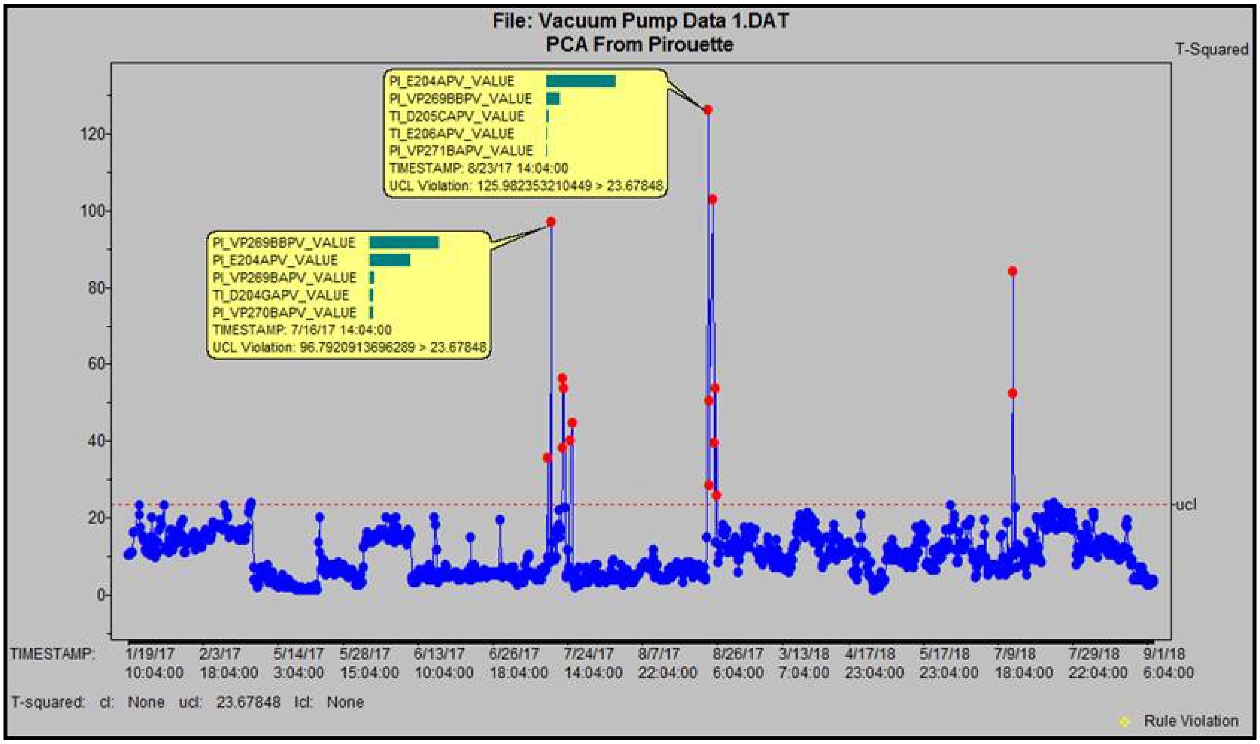

Figure 8. Quality Analyst deployment of multivariate model for vacuum pumps

By creating a steady state PCA model of this manufacturing process, a multivariate tool could be deployed that would allow continuous monitoring of the vacuum pump system. This success demonstrates how a multivariate approach can provide benefits in an industrial setting.