Early Event Prediction with Geometric Process Control: Compressor Surge

Alan Mahoney, Robin Brooks, John Wilson – PPCL, Gerrards Cross, UK

Abstract

A 30-variable geometric operating envelope model of a 6-stage propylene refrigeration compressor was tested for its ability to identify known surge events in historical data. It identified 12 out of 23 events over one hour before the surge event occurred and the remaining events during the last hour. The advantage of these long prediction times is that very small control actions are needed to avoid the event which benefits both the compressor and the stability of the process to which refrigerant is being supplied.

Introduction

Adverse events such as surge in compressors, breaks in paper, glass, and film manufacturing, and pack-offs in oil well drilling cause upsets, unproductive downtime, equipment wear, and can be a major expense in operations. Avoiding these events is a key part of the design and operation of many process systems. Often the otherwise most economic operating conditions (fastest throughput or most energy efficient) are exactly the conditions which increase the likelihood of adverse events. Thus, a common operating goal is to operate near event conditions while minimizing the chances that these events will occur. Further complicating the problem is that the mechanisms leading to events are rarely fully understood and the ideal operating conditions can be very sensitive to process operations.

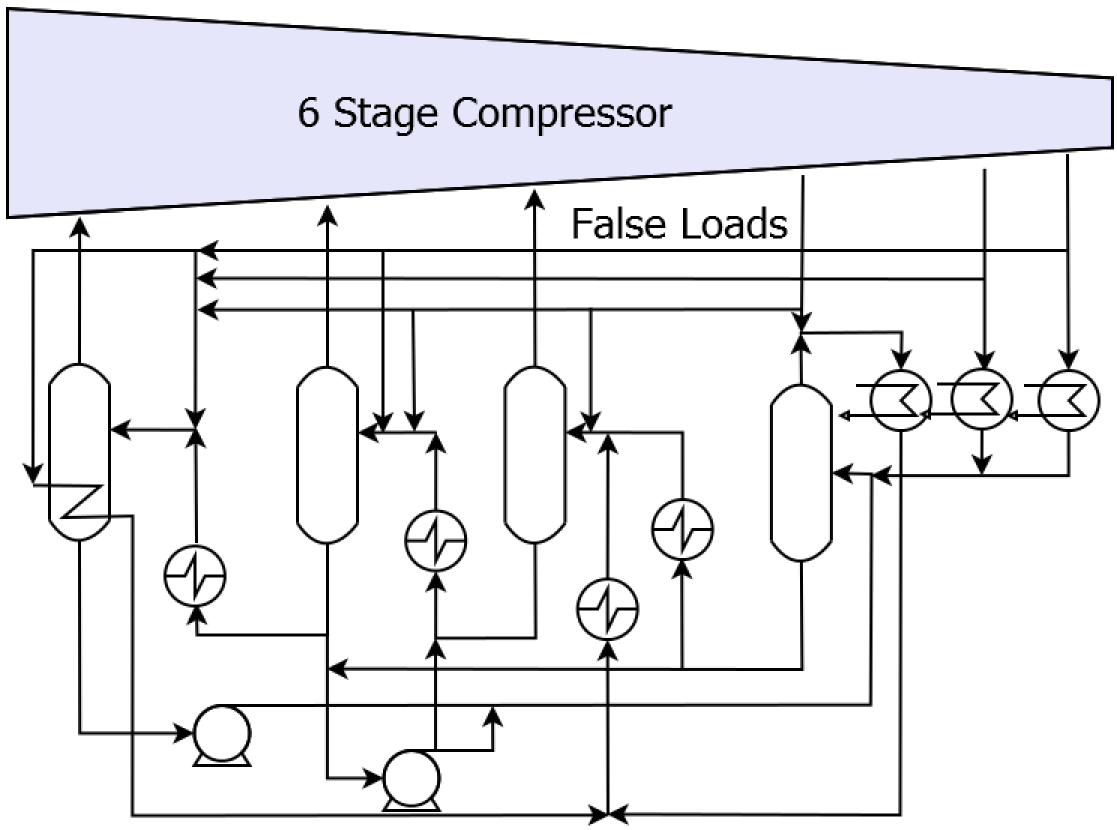

Figure 1: Schematic of a refrigeration system.

The process example is a closed propylene refrigeration system. It consists of a six-stage compressor, heat exchangers for providing cooling duty for a plant and cooling the process fluid, gas-liquid separators and associated piping, valves, and controllers. Figure 1 shows a simplified schematic of the refrigeration process. Ethylene passes through the compressor, condenses in the heat exchangers on the right, and provides cooling to the plant through the heat exchangers in the left half. Liquid-vapour drums keep liquid from entering the compressor and direct it to heat exchangers. The goal is to operate the system at high efficiency while avoiding surge in the compressor. These goals are contradictory under low load conditions, as the highest efficiency operation is near surge conditions. The operation is driven by changing conditions, as the system is coupled to plant behaviour and must react to changing demand from other systems. The combined plant and refrigeration system can take hours to settle following a disturbance.

Surge occurs when the throughput through a compressor is insufficient to overcome the pressure drop. Back-flow occurs through the compressor, equalizing the pressures. This causes fluctuations in flow and pressure as well as taxing the equipment. One strategy for avoiding surge is a local fast-acting control loop and bypass valve, possibly with cooling. Due to the size and complexity of this system, a simple local control loop is not possible, and the equivalent system is a false load system that is included in the operating strategy. This work looks at using Geometric Process Control (GPC) for advising this operation.

In GPC the operation with desired characteristics is represented as a geometrical space. The defining data comes from process history which implicitly captures the interacting behaviour of the system. Rather than focus on local measurements and react quickly, the goal is instead to identify an operating region where surge is less likely to occur and alert the operator when leaving this region.

The GPC model is able to identify behaviour that is leading to surge with at least an hour warning in 50% of the events studied here. This provides time for plant operations to avoid the conditions smoothly, reduces bearing wear and provides more efficient operation.

Geometric Process Control

The parallel coordinate plot allows the visualization of points, surfaces, and regions in arbitrarily high dimensional space (Inselberg, 2009). The application of these graphs to plant data is motivated by visually displaying and exploring the relationships between hundreds of variables present in plant historians. The view allows presentation of data impossible with two-dimensional plots. The parallel coordinate plot allows visualization and exploration of design space and operating envelopes in high-dimensional space that enables geometric process control, a supervisory control technique based on the underlying geometry of the process phase space.

The parallel plot is a graph with the axes drawn parallel to each other. A set of variable values (a point) are represented by a polygonal line connecting the values of each variable on its own axis. Figure 2 shows one point capturing operation at a certain date and time. To be useful, more points need to be shown. Figure 3 adds 170,000 points representing 120 days of operation at 1-minute intervals. A set of points corresponding to normal operation, discussed below, is highlighted in yellow.

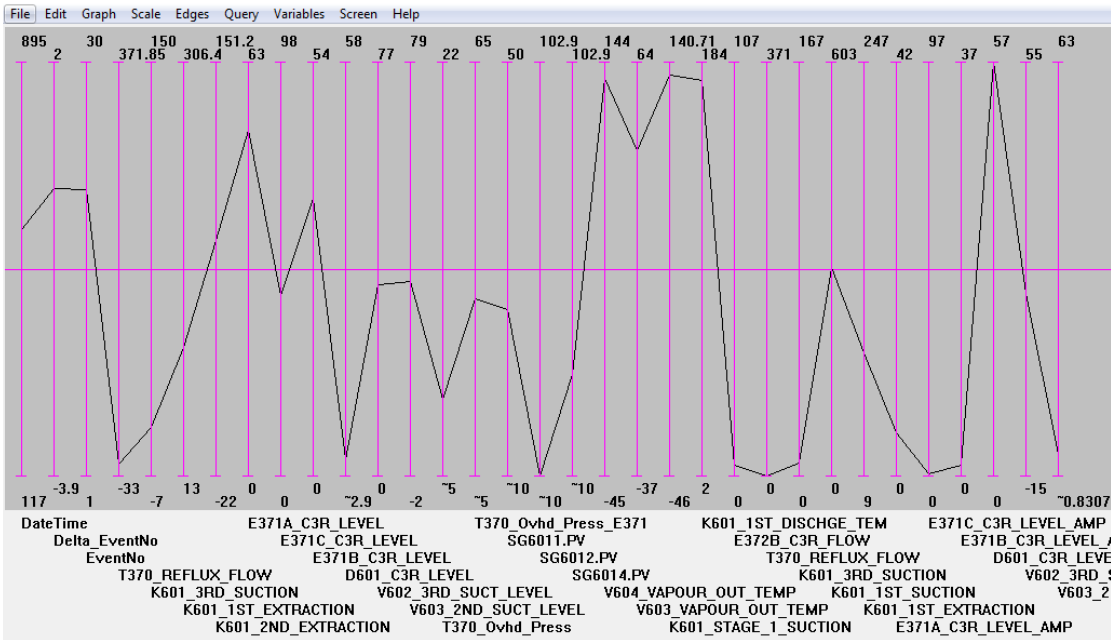

Figure 2: One point on a 33 dimensional plot.

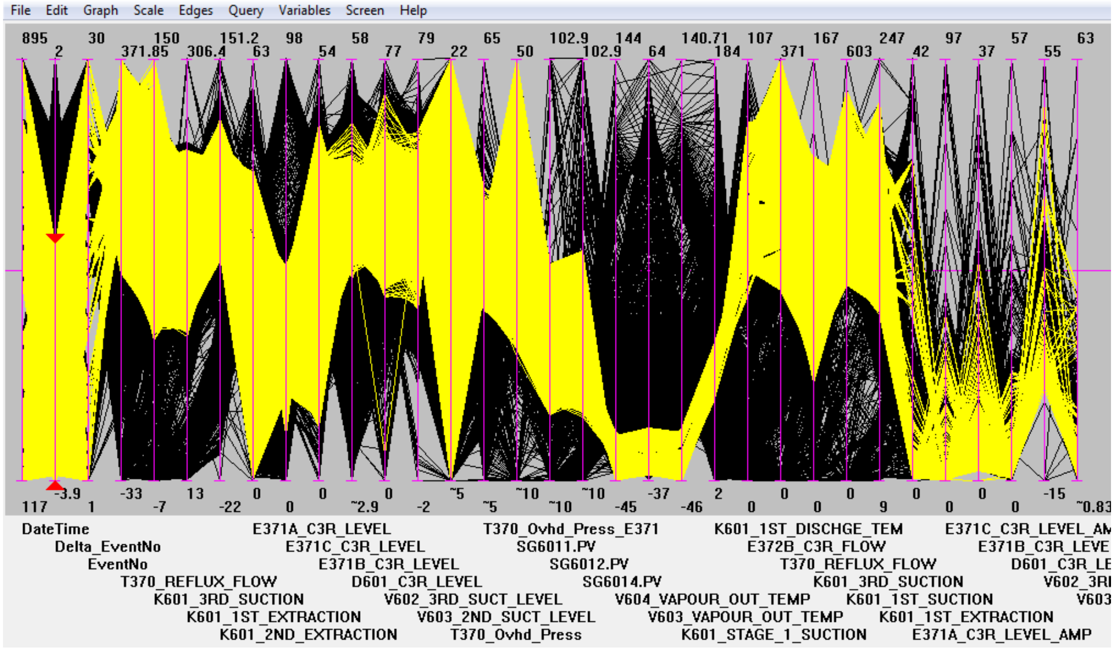

Figure 3: 173,219 points on a 33 dimensional plot.

Using the parallel plot to visualize spaces, surfaces, and points in high dimensional space enables supervisory process control based on the geometric relationship between these spaces. The geometry between target, historical, and other process envelopes creates a math-free model that captures the underlying relationships between feed characteristics, process variables, operation strategies, and targets of quality, throughput, and efficiency.

This is done by first identifying a set of historic operating points that are to define the model. Next, a multi-dimensional solid object that contains these points is identified, and the objective is to maintain operation within this object. This type of modelling is wholly data-driven, and the mathematics governing the relationships between the variables have been captured in the shape of this geometric solid.

Geometric Process Control (GPC) is a method for geometrically capturing and evaluating a model of a process operating envelope. The operating envelope concept is more easily understood in an example low dimension system. Figure 4 shows the operating space of a process with two operating variables. The operating envelope has been defined by restricting the operating space on a third variable, perhaps a quality that is only occasionally available. From constraints on the process history, operating in the bright green space has been found to produce good quality product. In addition to the shape of the envelope, there are a number of interesting aspects of this picture. Absolute minimum and maximum values for P1 and P2 define the light green ranges beyond which the process can never achieve the dark green operating envelope.

However, even when the process is within these limits, there are still values which do not fall within the green operating envelope. If P1 is high and P2 is low, the operating point is at the lower right hand side of the enclosing box, which is outside the envelope, despite being within the absolute ranges. By capturing this interaction, the envelope is approximated much more closely than any set of individual variable limits. This is the basis of GPC.

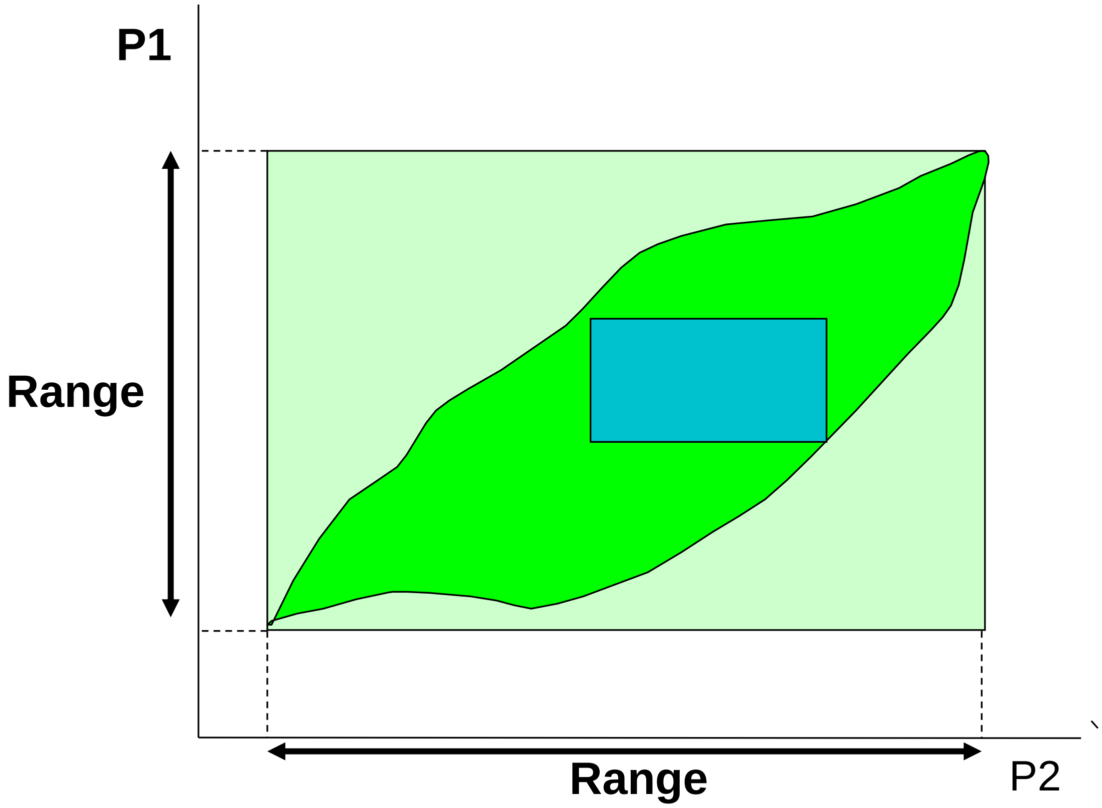

Figure 4: Schematic envelope on two variables. The operating envelope is indicated in dark green, the enclosing box in light green, and one potential inscribed box in blue-green.

In any practical case, the envelope is more complicated than the area in the example and extends into as many dimensions as there are process variables rather than just two, but the principles are the same. The definition of the envelope, that is, the boundary between the green space inside and the space outside, is done by identifying in historic operating data the behaviour that corresponds to desired operation. This is then evaluated in real time to determine whether the current operating point lies within the envelope or not.

The application area of the model, i.e., supervisory control, event prediction, fault detection, etc., is linked to the choices made for model definition. For event prediction, the model should consist of operation that is normal in the sense that it is safe and customary. Additionally, the model should not contain conditions where events are occurring, or the run up to events. When the process deviates from these normal conditions and begins moving toward an event condition, an excursion from the operating envelope will be seen and the operator alerted to take appropriate action. The most important variable that won't be included directly in the model is the time until onset of event. This can be calculated for historic data where we know when events occurred.

Compressor System Data

Two sets of data from the propylene cooling system were provided. Each consisted of roughly 830 tags with most of these tags in common between the two sets of data. The main differences were in frequency and time coverage. The first consisted of 4 years of data at 1 hour intervals; the second consisted of 5 days of data around each of 30 suggested surge events (with some overlap) at one minute intervals. Only the second dataset is used in this study.

Process Engineering went through the full taglist and produced a reduced list of 87 key tags along with role (manipulated/non-manipulated). Of these, five were not present and another three were single-valued. In five other cases, the control output (OP) tag for a variable was listed as important while the current value (PV) was not. These OP tags were replaced with corresponding PV variables, as it was expected that the PV variables were intended rather than the control output and those tags had been erroneously selected based on description without examining tag sub-type. This left 79 tags that were carried forward.

The data covered five days around each event, about four days before surge and one day after. Precise times of surge events were identified by examination in CVE. Time plots around each of the events were investigated, and event times were estimated by non-experts to within 10 minutes by identifying qualitative changes in system behaviour (see Figure 5 as an example). The adjusted event times often preceded the nominal event time by 40-60 minutes and up to 2 hours for some events. One set of data had no apparent surge event and was not used. In three other cases, events occurred within three days of an earlier event. In each of these cases, data before the first event was retained, and the later event(s) were not included as recovery behaviour from the first event could confound the development of the second and later events.

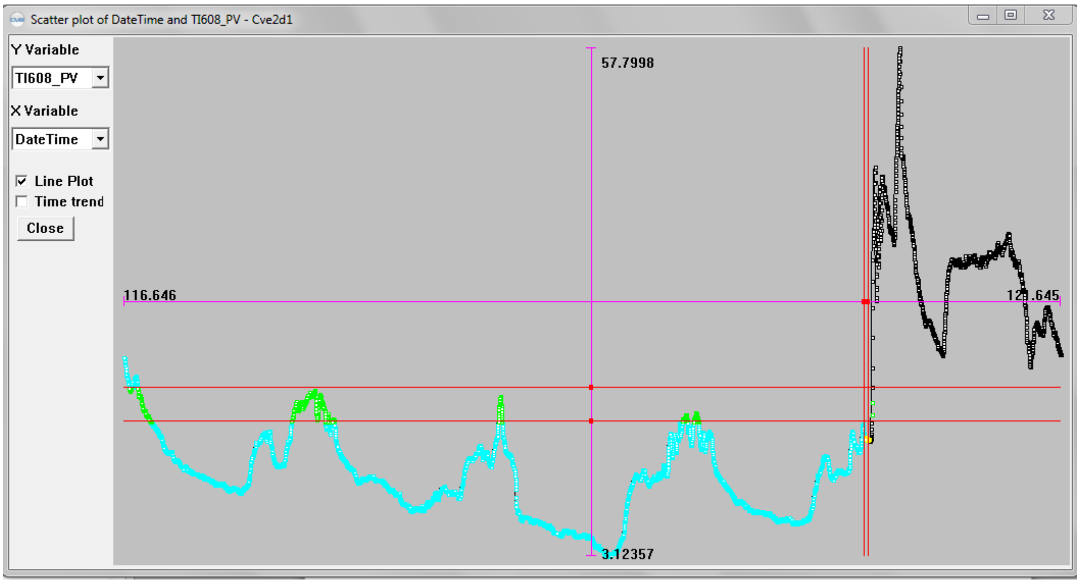

Figure 5: Surge time identified by sudden qualitative change in one or more process variables. Note oscillations with 24 hour period in system beforehand.

The raw process variables were supplemented with variables calculated to capture the frequency and amplitude of variations in each variable. The original and calculated variables were visually screened and categorized. Those that showed qualitative behaviour change in the hour before at least two surge events were kept. Six calculated variability and 24 raw variables were retained for a total of 30 variables in the model. This method would not normally be used to screen variables for a CPM model, but was used in this case to keep the variables in common with an already developed model. This method of identifying variables individually disregards the relationships between variables that form much of the power of GPC, and may remove important variables or leave unimportant ones. An alternative model is suggested in Future Work, below.

Model Formulation

The engineering choices in model definition consist of choosing a set of historical operation that is inside the envelope, that is, the behaviour that is desired and not to be alerted. Depending on the application of the model, different choices are used for normal behaviour. For event prediction, a model that produces alerts when the system moves toward events is be made of normal operation that is far removed from surge events evident in the historic data. This behaviour should be passed without generating operator alerts. Building and validating the model involves selecting appropriate data and testing the model for selectivity and sensitivity. If necessary based on validation, the model can then be refined iteratively.

The first step is to identify and remove abnormal operation. This is behaviour that, even if it did not lead to surge, does not represent normal operation. This includes the behaviour while surge is occurring and during plant recovery, start-up, shut-down and idle periods, and other abnormal events. This behaviour is not included in the model, as it would expand the envelope of normal operation artificially and the model would be less sensitive to surge events. It is also not useful for model comparison, as every model, regardless of quality, will generate alerts, so it does not provide discrimination. This behaviour is eliminated, along with obvious outliers as early in development as possible. The focusing operations in CVE make this easy to do, and also make it easy to later remove additional abnormal operation discovered at later points in the model-building process.

The next step is model selection. For event prediction, the model is selected by considering a time horizon for possible prediction. If this horizon is too short, the model may include behaviour that anticipates a developing event, making the envelope less sensitive. Even though some of the rejected behaviour is expected to be within the envelope of normal operation, most of the normal operation remains to form an envelope. Figure 6 shows the process of model selection. Behaviour more than 12 hours before each event is picked out in yellow to form the model. The red triangles show the queries that pick out the model and test data. A few model outliers have also been removed directly here.

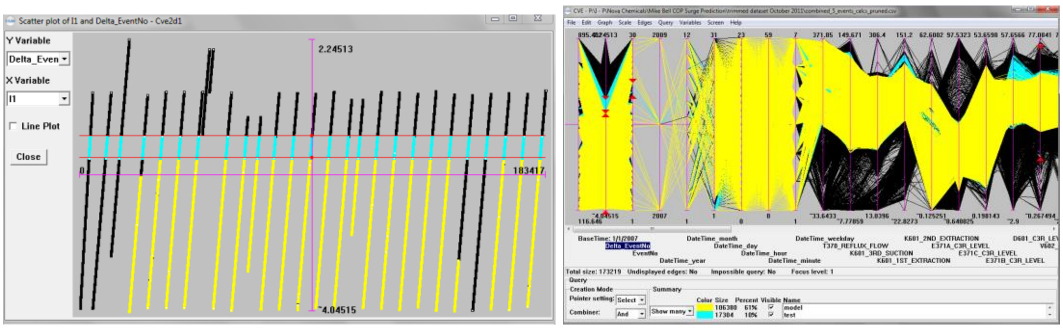

Figure 6: Model selection. Left graph shows time until event on y axis. Right graph is a parallel plot. Normal operating envelope is yellow, 12 hours before event until event is in cyan, after event is in black.

Finally, test data is used to validate the model. In this example, the data in cyan in Figure 6 is used. This data includes the behaviour from 12 hours until surge was seen in historic data. During this period, the initial data is expected to be represent normal operation. If the system moves out of the normal operating envelope, alerts are expected. These alerts will be concentrated in the last hours before the historic surge event. By comparing the alert rates in the first and later periods, the sensitivity and selectivity (false positive/false negative) rates can be estimated, and the patterns of alerts preceding surge examined. The model is expected to perform similarly in the future as it does with historic events, so testing with the historic data validates the model.

Application and Performance

The Process Modeller evaluates whether the current operating point is within or outside the model envelope. During offline model evaluation, the same geometric algorithms are applied to a historical dataset to simulate real time data. This allows the sensitivity and specificity to be evaluated with historic data.

From the split of the data (above), the data that precedes each event by 12 hours or less was used to evaluate the model. Of course, this data was not included in the model envelope. For evaluation, the model tolerance was set at 10%, dictating that real-time values must exceed allowed ranges by 10% before the point is judged to be outside the envelope. The tolerance controls the balance between sensitivity and selectivity, a lower tolerance increases the number of alerts, reducing the number of false negatives and increasing the number of false positives. The value of 10% was chosen to eliminate most false positives for the initial study.

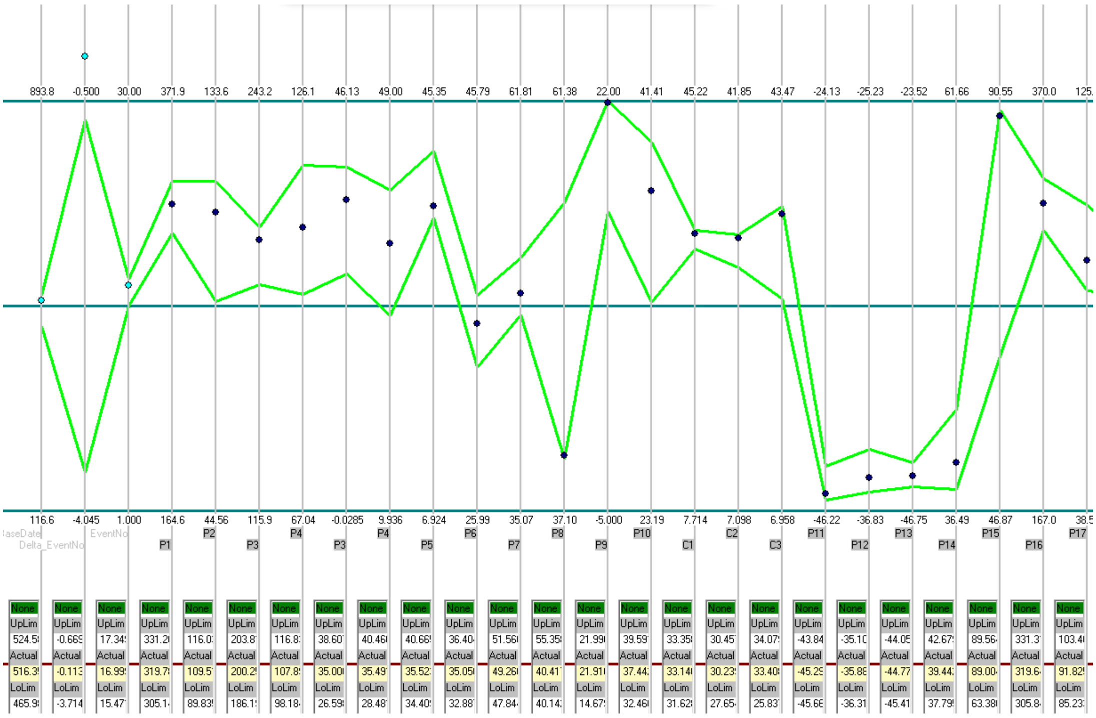

Figure 7: CPM screenshots showing point inside the operating envelope.

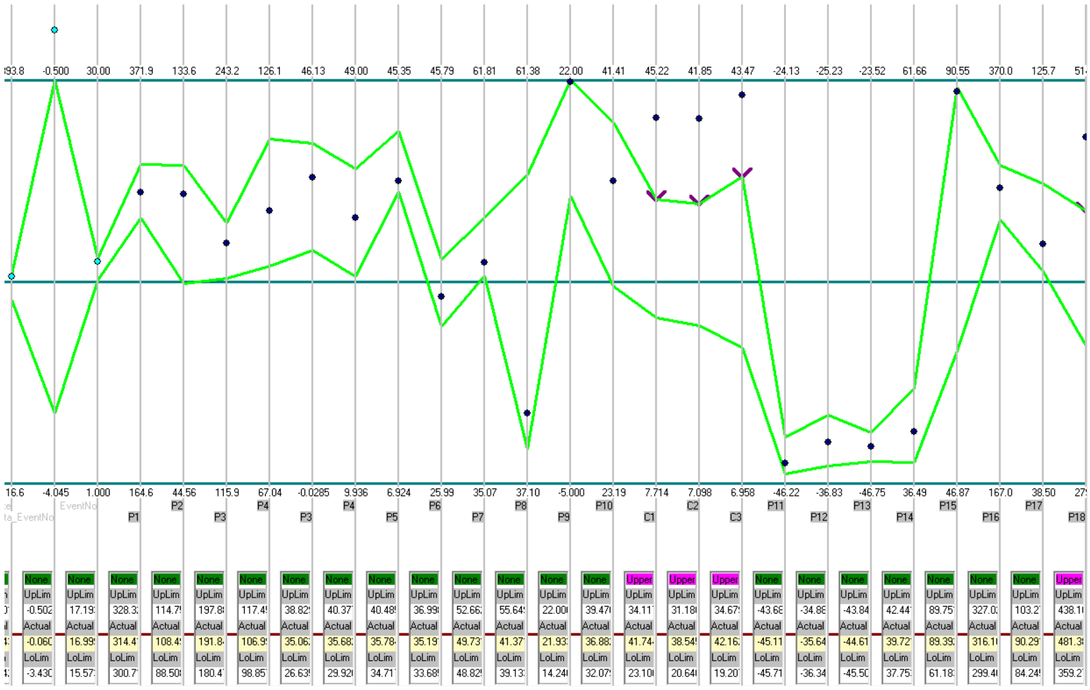

Figure 8: CPM screen shot showing an operating point outside the envelope. The variables in violation are alerted in purple.

Figure 7 and Figure 8 show the display of the process modeller when operating respectively inside and outside the envelope in the run-up to event 17. The black dots represent the current values of the process variables. The scales on each variable run from the minimum to maximum acceptable under any conditions, capturing the limits of the enclosing operating box of Figure 4. The operator can see at a glance whether the point is inside or outside the operating envelope, the distance outside and the range of operating space left inside. Figure 8 includes four variables in alarm. In every case, they are still within the absolute limits of the top and bottom scale, but are too extreme in relationship to the current values of the other process variables. This shows the ability of GPC to detect an envelope violation that would not be seen with individual variable limits. An envelope violation has been detected that is still within the absolute limits of the variable ranges, but indicates a violation of the normal relationship between the variables.

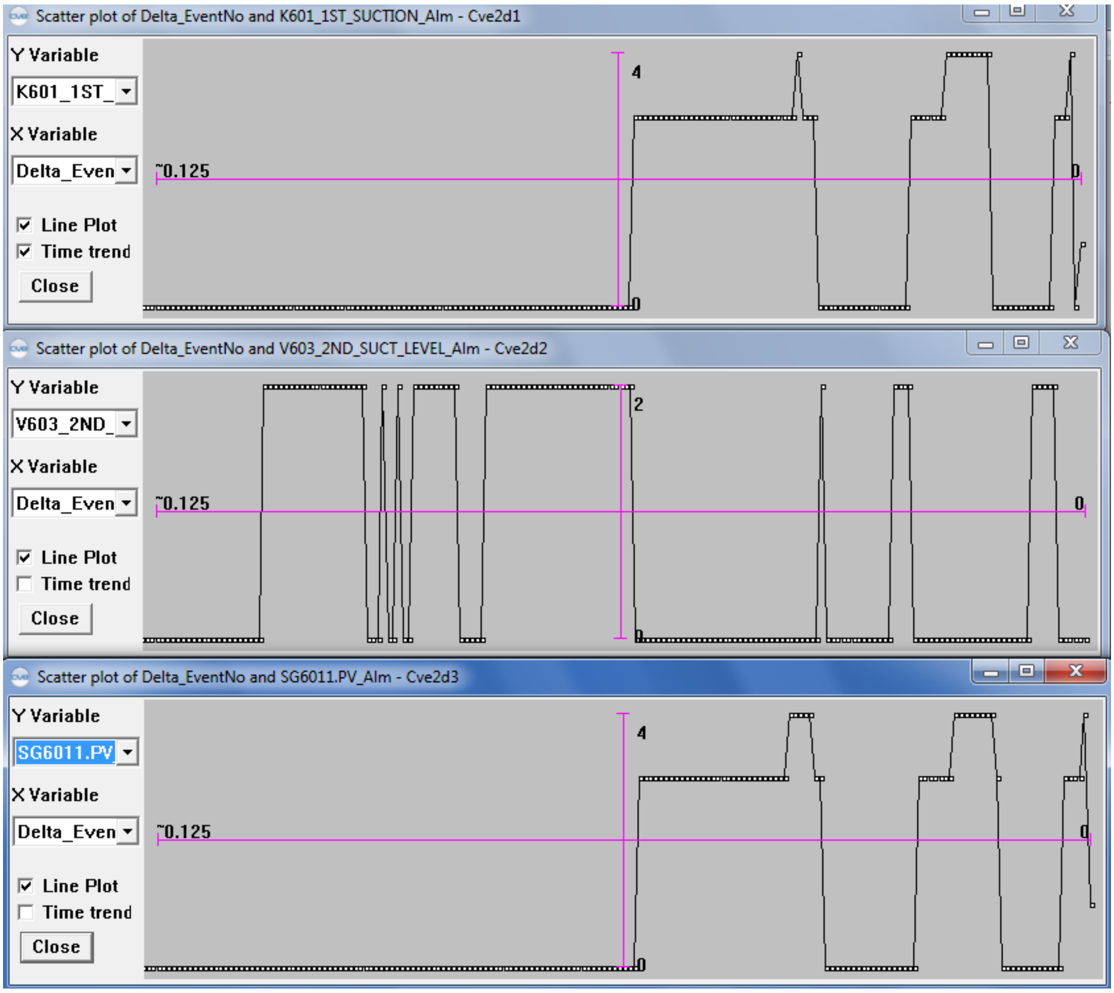

Figure 9: Alerts on three variables in 3 hours approaching event 17. Vertical scale is alert type, 0: none, 1: lolo, 2: lo, 3:hi, 4:hihi.

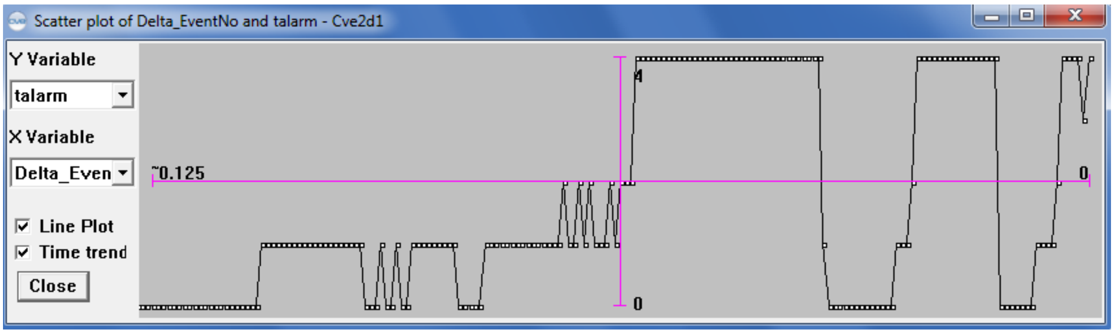

Figure 10: Total number of standing alerts in the 3 hours preceding event 17.

Figure 9 shows the alert behaviour of three variables in the three hours approaching event 17. The horizontal axis is time, in days, increasing from three hours before the event to the beginning of the event. The vertical axis is a coded flag, zero if the variable is not in alert, 1 or 4 for low and high absolute limits (univariate violations), and 2 or 4 for low and high relative alerts, those that occur when the envelope is violated but the system is still within the enclosing window. The first absolute violations, those that would be seen with conventional variable limits, are not seen until the last hour before the event, while relative alerts begin appearing two and a half hours before the event. Figure 10 shows the total number of variables in alert for the three hours approaching the event, demonstrating that both the frequency and number of alerts increases as the event approaches.

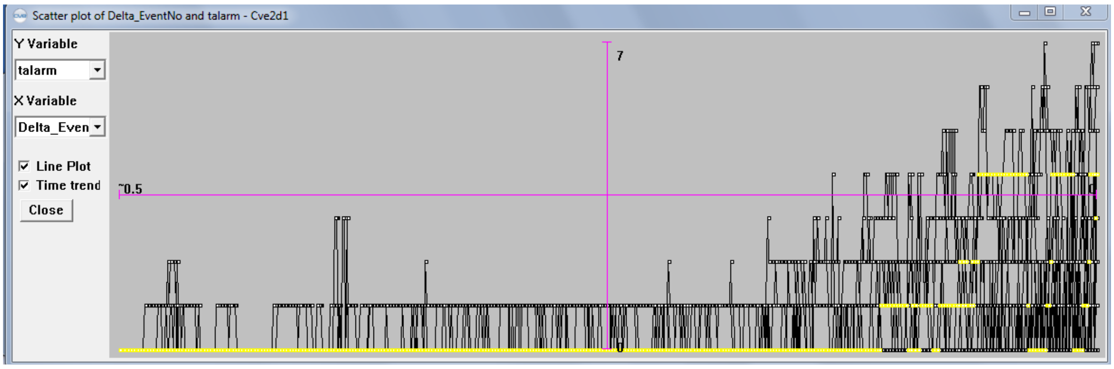

Figure 11: Alert rates per minute approaching event at right (time 0). Event 17 is shown in yellow.

Figure 11 shows in yellow the number of variables in alert in the time preceding each event. Note that the time scale here runs from 12 hours before the event. Event 17 is in yellow, but the graph also shows the traces for the other 22 events. The general trend of increasing frequency and number of variables in alert, sometimes 4 hours in advance, is clear.

For evaluating the warning horizon and false alarm rate, the trigger alert level is taken to be two; that is, a single alert is ignored as part of the baseline normal operation. This level will depend on the variables present in the model and the quantity of data available for building the model, and may be revised with future models. The first six hours (i.e., 12 to 6 hours before the event) is expected to be normal operation with no alerts. More than one variable in alert here is considered a false positive. This occurred on three occasions and may result from normal behaviour that wasn't present in the model dataset. This corresponds to 80% of alarms leading to a surge event. The number of false positives decreases as more data is available for model building.

The final six hours before the event is expected to show increasing number and frequency of alerts as the event develops. In this case, alerts before the last hour are considered early warning. Events that are not alerted until the final hour preceding the event are considered unpredicted. This conservative buffer period ensures that only events predicted well in advance are counted, as well as allowing for any errors in identifying the actual time of event onset.

Of the 23 events, 12 were detected before the last hour. Of these, three were detected 4-6 hours before the event, six were detected 2-4 hours early, and 3 were detected 1-2 hours. The remaining events were detected only in the last hour. This may have been due to quick movement of process variables (either operator driven or due to sudden physical failures or other causes). If the calculated variables are removed from the model, performance is reduced to 9 of 23 events being detected.

Discussion and Future Work

This work has shown that a model can be constructed of normal operation that rejects significant behaviour preceding surge events. The model is built from historic data without plant tests. Selecting the model was based only on time separation from historically identified surge events.

The datasets used here were restricted to the time around surge events, as these had sufficient data frequency for the secondary calculated variables. A more robust model would be built and tested on the full two year dataset. As some abnormal operation unrelated to surge is likely to be included, detailed data investigation to remove these periods from the model will be required, and how the model reacts to these non-surge events can be investigated. The larger dataset will enable testing on time periods far removed from surge to give more accurate false alarm rates. These rates are also expected to be lower, as operation in the days before surge is more likely to give false alarms that that removed by weeks or months.

Variable screening was done by visually inspecting time trends of individual variables. While this has produced good performance, it was done without considering the variable relationships that CPM uses. It may be valuable to bring variables that were eliminated into the basis. Similarly, some of the variables included may not contribute to the alarm rate (or may contribute disproportionately to false alarms). This normal part of model revision will be done with the full dataset.

As time goes on and operators have model feedback and become better at avoiding surge, the additional operation data can be added to the model.

The warning times for alerted events are significantly longer than are currently available to operators due to the model looking for unbalanced operation across the plant which may lead to unstable operation, rather than trying to detect it locally within the compressor which provides very little warning at the onset. The additional reaction time can allow more gradual process changes avoiding process disturbances. The reduction of surge events may also prolong equipment life or allow extended periods between maintenance.

The general methods for event prediction can be applied in other systems where there are undesirable events for which there is sufficient history available in the plant historian. By constructing operating envelopes that avoid events, and the precursor behaviour to events, it is possible to both generate process understanding about the process that lead to them as well as constructing online detectors to aid operators in avoiding them. Current areas of interest include paper/glass/plastic film breaks and oil well drilling.