

OPERATOR ALARMS ARE THE FIRST LINE OF DEFENCE

Robin W. Brooks, PhD*, Alan Mahoney, PhD, John Wilson, PhD, CEng, Na Zhao PhD
Process Plant Computing Limited (PPCL), PO Box 43, Gerrards Cross, Bucks. SL9 8UX, UK
*Corresponding author: Robin_Brooks@ppcl.com

Operator Alarms should be the first line of defence in every plant but all too often are more of a nuisance than an aid to the operator. This exposes safety alarms to more process excursions with the consequent increase in probability of a Failure upon Demand detracting from the plants overall safety capability. Poor operator alarms also contribute to poor process economics. The situation has arisen because there has never been a fundamental understanding of how alarm limits relate to process control and to process operating objectives.

We have identified that the Operating Envelope of a batch or continuous process is the missing factor that unites all three topics. Operator Alarms are inter-related by positioning them on the boundary of an operating envelope which is today approximated by an alarm window or hypercube.

Understanding and using the geometric relationship between an operating envelope and its approximating hypercube eliminates many false alarms. This substantially improves the credibility of the alarm system to the operator and allows earlier annunciation with more time for the operator to respond. The new alarms give the operator earlier and positive warning of deviation from whatever combination of business, environmental and process performance objectives are the operating windows chosen objective thus contributing to the economic performance of the plant and so earning the alarm system a share of the business case for further investment.

Process control has, for the first time, a well-defined boundary within which to operate so that it can deliver business objectives not measurable in real-time.

The method complies fully with the project methodology and the achievement of EEMUA 191 and ISA SP18 ergonomic objectives and in addition is interactive and predictive so that alarm system performance can be predicted for any set of alarm limits being considered to provide immediate, interactive feedback during the alarm review process.

This makes alarm reviews fact-based instead of opinion-based and so saves considerable time in the review process for all involved.

Alarm Floods are investigated and recognised as the process moving into a different operating envelope. This suggests separate sets of alarm limits for normal and flood modes of operation. Some floods can be separated into a disturbance period and a prolonged process settling period caused by interacting circulation or recycle loops such as in an ethylene process and may be the bulk of what is remembered as ‘the alarm flood’.

There can be many causes that lead to similar settling period behaviour in which case providing a separate set of alarm limits for the settling period can be a straightforward way to increase the overall performance and acceptability of the alarm system. The necessary state-based and/or mode-based logic can in many cases be incorporated into the analysis so that performance of even sophisticated alarm systems can be predicted interactively during the alarm review process and before implementation in the process control system.

There are two major alarm systems in a process plant. The first is the Safety Alarm System responsible for taking control and shutting down the process in extreme process excursions which both the process control system and the operator have been unable to prevent. Its role is to prevent an extreme excursion from turning into a disaster with liabilities and costs that can run into hundreds and even thousands of millions of dollars. Its costs are viewed as an insurance premium against a disaster that most plants will never experience.

The second is the Operator Alarm system intended to draw the process operator’s attention to a situation beyond the capability of the process control system to prevent and requiring application of the operator’s considerably greater human intelligence to resolve and correct before the safety system intervenes and shuts down the plant. Automatic plant shutdowns are expensive in lost production time and possible consequential plant damage. Operator alarms give the operator time to intervene and correct the situation to avoid a shutdown. They have an ‘insurance premium’ value in reducing the demand upon the safety system and thus the small possibility that it will fail when called upon.

Operator alarms are sometimes called ‘Economic Alarms’ because they are commonly believed to be intended to help the operator in the achievement of the plant’s economic objectives by assisting him in keeping the plant inside the operating space where these objectives can be achieved. The EEMUA 191 guidelines advise positioning operator alarms ‘on the boundary of where the plant normally operates’ but without giving any guidance on how to locate the boundary. Most plants accept that ‘Normal’ operation refers

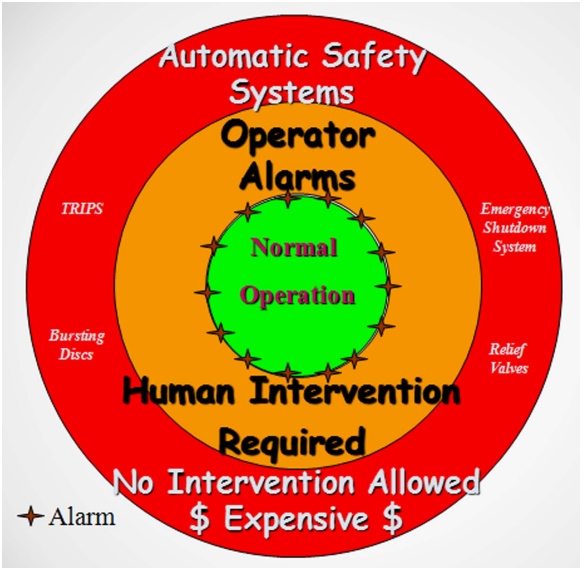

Figure 1. Operator alarm limits at the boundary of where the process normally operates

to the Operating Envelope within which desired economic results are achieved similarly to Figure 1 and place the operator alarm limits where they imagine the boundary to be. This would suggest that (a) alarm limits are ideally the same as operating limits and (b) the economic cost of violating an alarm limit is the delta cost between the material produced and operating costs of desired and undesired operation.

But the first practical problem behind the advice to 'put the operator alarms on the boundary of where the plant normally operates' is that there has been no way to determine the location of the boundary of normal operation when the operating objective is that of meeting all KPI's, including those that cannot be measured in real-time, at all times. The consequence of this is that Figure 2 is a representation of alarm limits as they really are in practice. Some alarm limits are set in the orange recovery space where they will, at best, annunciate late, giving the process disturbance more time to grow and requiring a larger corrective action, or in many cases are set so wide that they can never annunciate. Other alarm limits are set inside the green 'normal operation' space where they will annunciate unnecessarily some of the time creating false alarms and leading to their being labelled as 'bad actors'. Without knowledge of the location of boundary or of how alarms relate to each other there is little that can be done to cure a bad actor other than to push the alarm limit 'outwards' towards or past the guessed position of the boundary.

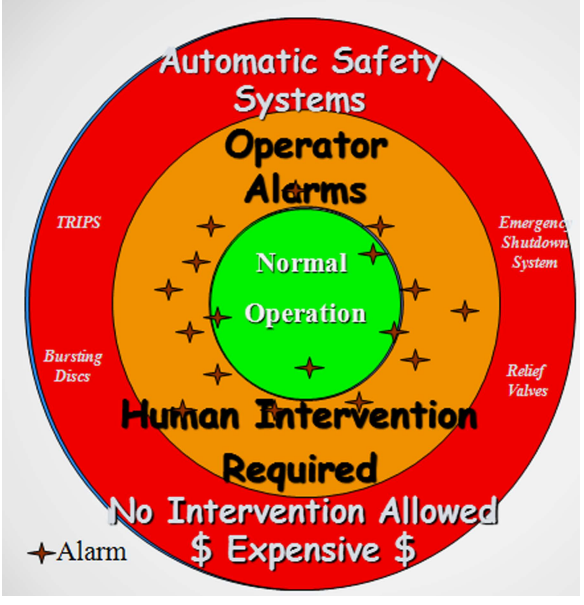

Figure 2. Operator alarm limits as they usually are since the boundary of normal operation is unknown

OPERATING ENVELOPES AND OPERATING WINDOWS

The boundary of normal operation is actually the boundary of the Operating Envelope defined by the objective of complying with all KPI objectives.

The 'Operating Envelope' noun-phrase has been used by generations of chemical engineers to describe a closed boundary with different properties of something inside and outside the boundary. In two dimensions the boundary is a line that separates two areas. In three dimensions the boundary is a surface separating two volumes. So if we are using variables as dimensions and have more than three variables what then is the boundary and what is it separating? The reality has been that a schematic such as Figure 2 was as close as one could get to describing the nebulous concept of an operating envelope involving more than three variables. Figure 3 shows a simplified view of an operating envelope for just two process variables PV1 and PV2. There are many constraints that affect a process and some are shown, over-simplified, as straight lines. Some constraints, such as the production planner's desired quantities and qualities of product, design limits, equipment operating limits and such are known and understood but there are many other constraints buried deep in the chemistry, thermodynamics or hydrodynamics of the process that are not understood and perhaps not even known to users. But at any time, the operating envelope to achieve the objective of meeting all objectives is given by the red shape bounded by the innermost constraints.

However, the constraints are functions of many more variables than just PV1 and PV2 so will move as the other variables move with the result that a little later the operating envelope might look like that in Figure 4.

We can begin to imagine the envelope as a tube of varying cross-section in 3-d by introducing a time axis into the picture as in Figure 5.

Alarm limits (and Operating Limits) are almost invariably defined today as high and low values on a

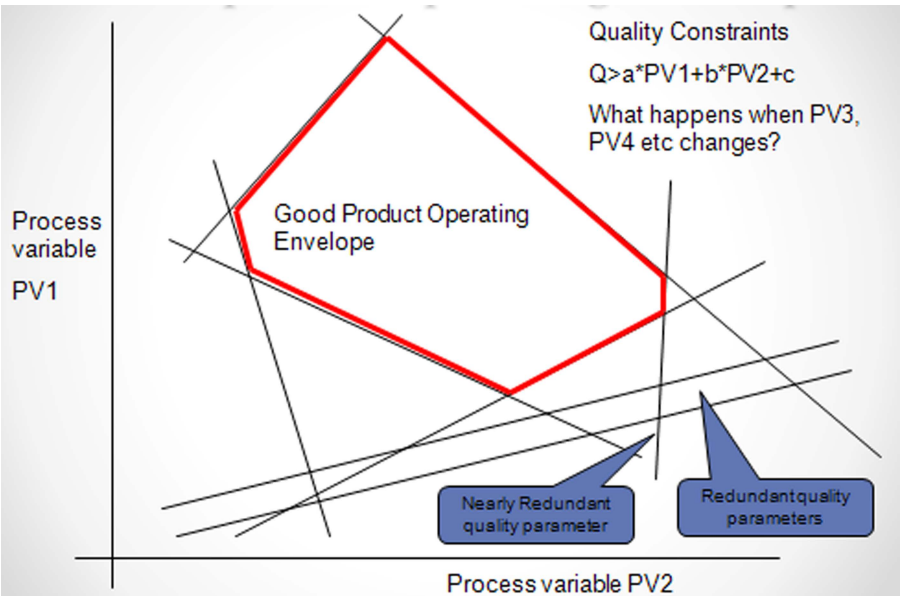

Figure 3. Two-dimensional simplified operating envelope

single variable so are by definition independent of any other variable. Consequently they appear as vertical or horizontal lines when drawn onto our picture of a simple operating envelope. Figure 6 shows Operator Hi and Lo limits for both PV1 and PV2 in blue and two Hi Equipment Operating Limits in brown.

It is clear that the rectangle formed by the blue or brown limits (also known as an Alarm Window or Operating Window) is not the best approximation to an operating envelope.

This simple 2-d picture is useful for forming concepts but we know that for any real process there must be many more than two process variables in our envelope and now have to address problem of not knowing how in Figure 5 to draw the fourth let alone the four hundredth axis.

Fortunately the solution is at hand in the form of Isenberg's[1] parallel coordinate transformation where the axes are drawn vertically and parallel to each other instead of orthogonally as in Figure 6. Each variable has its own axis just as in a conventional graph. A set of related variable values, most commonly related by a moment in time, can then be placed individually each on its own axis and the intersections on the axes joined with straight lines to give the zigzag line which indicates that these values are part of a set and 'belong' together. The zigzag line is in fact a

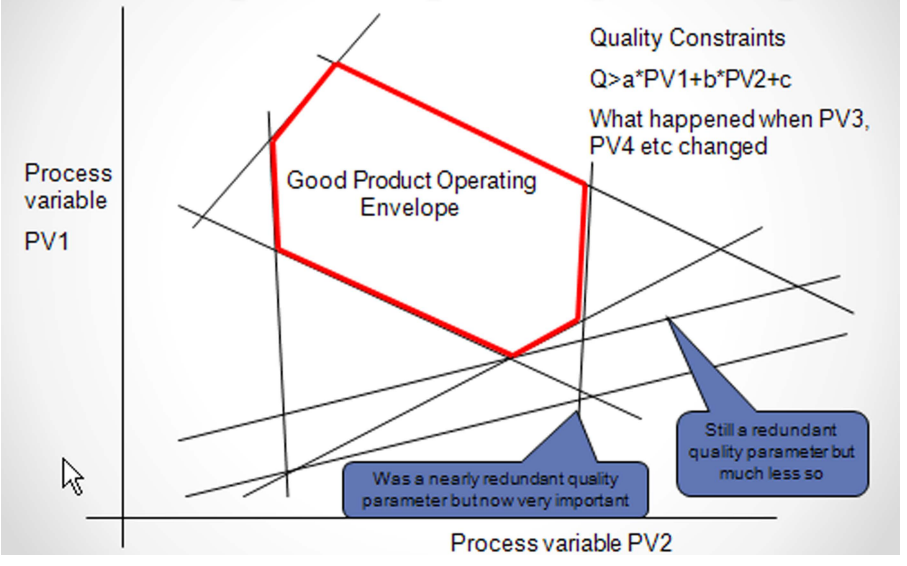

Figure 4. The same envelope a little later

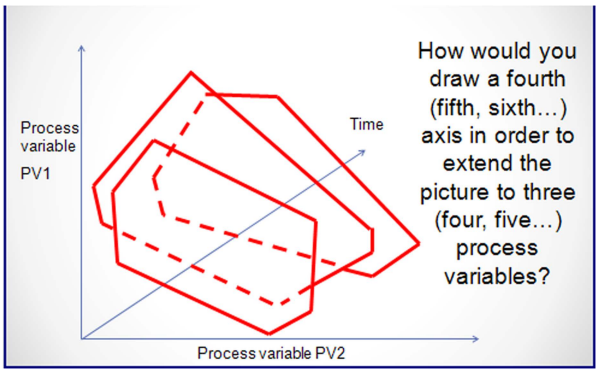

Figure 5. A 3-d operating envelopes cross-section varies in time

representation of a point in an imaginary high-dimensionality space and has as many coordinates as there are dimensions or variables. The individual variable values are the coordinates of the high-dimensionality point.

The purpose of a graph is to allow the plotting of many points so that patterns of lines, curves and shapes are formed. Human brains are very good at recognising and learning to interpret patterns. Plotting many points on the parallel coordinate graph, as in Figure 7, produces distinctive patterns and for the first time gives the ability to see where the process has operated and where it hasn't as well as many non-linear aspects of process behaviour that were not easily seen in the past.

This data in Figure 8 represents 3 months of operation of an oil refinery hydro-desulphurisation (HDS) unit and was taken from the process historian database at 10-minute intervals to give a total of 13,444 'points' or zigzag lines. There are 175 variables in the dataset. Patterns are formed and immediately start to be explainable. Solid black areas suggest that the process frequently operates in those areas and the black bands suggest there are two or more different Modes of operation. Light scatterings of points are easily understood as infrequent process excursions or as infrequent transitions from one Mode of operation to another.

Super-imposing the existing alarm limits on the graph as red triangles as in Figure 9 it is immediately apparent that this is not a good set of alarm limits. Some are inside the solid black area so will give false alarms at least some of the time; others are so far outside the black area that they will, at best, annunciate late in the event of an excursion out of past operating experience and may never annunciate at all.

The role of an HDS Unit is to convert sulphur atoms in hydrocarbon molecules into hydrogen sulphide gas which is easily separated from the liquid product for further treatment in another unit. An HDS unit can process Kerosene or Light Gas Oil (LGO or Diesel) at different times and, when there is no feedstock, has a 'Standby' Mode when it simply recirculates material through the reactor. It is these 'Modes' of operation which largely account for the prominent black bands seen in Figures 8 and 9.

The Envelope of all operation in Figure 9 encloses other Envelopes with more specific objectives such as the 'Kerosene Operating Mode Envelope', the 'LGO Mode Operating Envelope' and the 'Stand-by Mode Operating Envelope'. These are shown in Figure 10 in pink, blue and black respectively and now some sense can be seen for some of the alarm limits inside the black areas of Figure 9. For instance, three low limits commencing with the "Ultrafin Reactor Outlet Pressure" have been positioned, possibly after several cycles of adjustment followed by trial, to indicate the lower edge of the LGO envelope. Many other

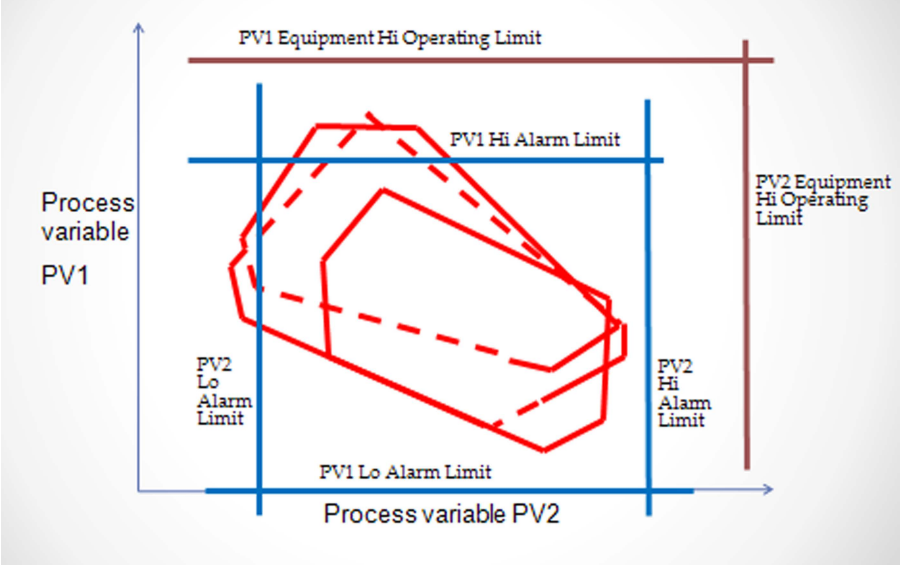

Figure 6. Operator HiLo alarm limits in blue and Hi Equipment operating limits in brown

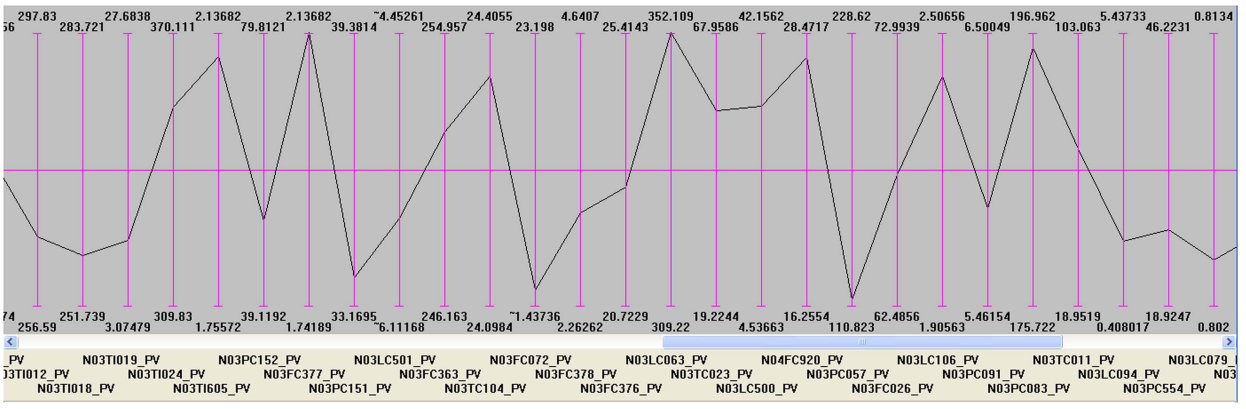

Figure 7. The parallel coordinate representation of the values of 27 process variables at a single moment in time

alarm limits in Figure 10 are still inside the operating space of one or more Modes creating false alarms some of the time and not well-related to a Mode envelope. One has to wonder whether all operators would have the same mode-dependent knowledge of which alarms are to be believed and which to be ignored and of the variability of response to the same alarm which could ensue.

Annunciation Rates and Alarm Counts calculated from process history data after removal of 'Bad Actor' alarm limits caused by instrumentation faults are shown in Figure 11. Ignoring Standby Mode, the annunciation rate is typically in the region of 4–5/hour and the number of alarms visible on the operators alarm list display is similarly in the region of 4–5. That this performance is acceptable against the EEMUA 191/SP 18.2 human factors guidelines emphasises that meeting these human factors guidelines is not by itself sufficient to create a good operator alarm system. There is a "measure of goodness" constraint missing.

Using a single set of alarm limits across all three Modes is 'Lumped Mode alarming' and is the norm in most plants today because of the difficulty and cost of setting any alarm limits with today's method. A simple way to create a set of lumped mode limits, but which doesn't easily or adequately allow the incorporation of KPI objectives is to move the alarm limits to the boundaries of where the process has previously operated. Then start to move alarm limits 'inwards' on Figure 12 to eliminate known process excursions, which you would want to be alarmed if they were to occur again, while watching the affects upon alarm performance in Figure 13. Equipment Operating limits can be shown as a second set of limits and should always enclose the set of operator alarm limits.

Much better is to define a set of limits for each Mode and use multi-Mode alarming. The alarm limits can then be made consistent with an operating envelope which includes objectives based upon variables not measured in real-time so that alarms will contribute to the achievement of operating objectives and not conflict with them as often happens today.

The Lumped-Modes Limits will be further improved during the Alarm Review by requiring justification or removal of far-out limits which may help reduce the size

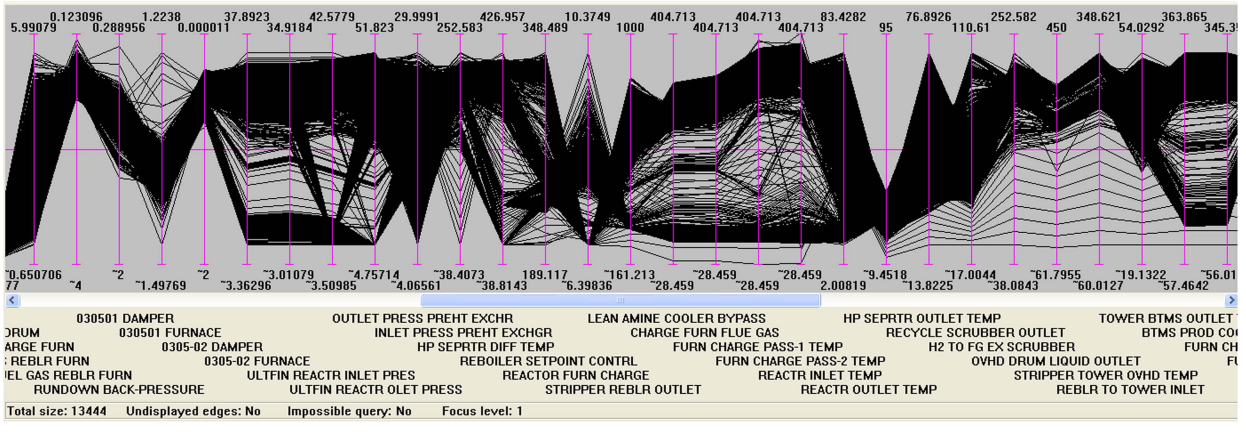

Figure 8. 3 months of operation sampled at 10-minute intervals giving a total of 13,444 points or zigzag lines. This is the Operating Envelope of 'All Operation'

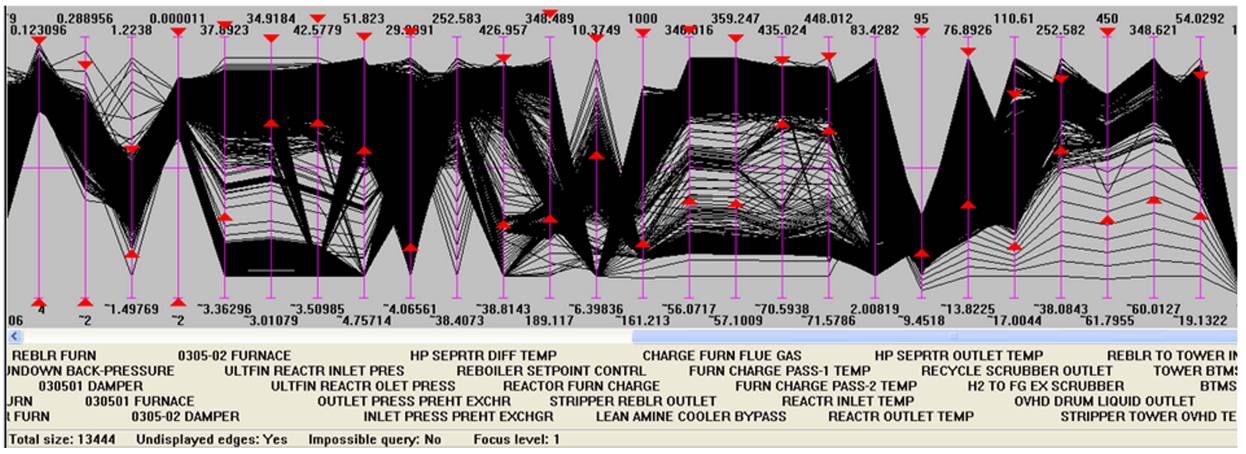

Figure 9. Existing HiLo alarm limits superimposed upon three months of operating data

of alarm floods. The much better operating environment that results will give confidence and a realisation that the alarm system can be improved to assist operations.

ALARM FLOODS

Referring back to Figure 3 the problem with the advice to 'put the alarm limits on the boundary of where the process normally operates' is that it can't apply when the process is not operating 'normally' such as when there is a process disturbance or alarm flood.

Figure 14 extends Figure 8 by showing the trajectory of an operating point that at time t_0 is well inside the normal operating envelope so that neither of PV1 and PV2 is in alarm. The operating point moves to t_1 which is outside the normal (red) envelope but inside the blue box formed by the alarm limits on PV1 and PV2 so neither is in alarm.

At time t_2 a Hi alarm is present on PV2 but not on PV1. If the operator responded quickly and 'turned' the process back at this stage it would possibly be called an 'excursion'. At time t_3 Hi alarms are present on both PV1 and PV2 which might, in this 2-d view, be called a 'flood' rather than an excursion.

What is happening is that some event has occurred to change the capability of the process so that different constraints have come into effect and the effective operating envelope has become that shown by the orange lines instead of the red ones. The orange and red envelopes must overlap and if we had enough operating data to define the orange envelope we could create sets of alarm limits for both of them and change from one to the other to minimise the number of alarms as the process moves fully into the orange envelope and its associated green alarm limits box.

Recognising this underlying reality, it is apparent that using one set of alarm limits for both normal and flood behaviour, as is the majority practice today, will always be a compromise with poor alarm system behaviour in one or both states.

A common practice today is to widen the limits for normal operation in the hope that this will somehow moderate the size of floods. All that this practice achieves is to prevent some smaller disturbances from annunciating at all and

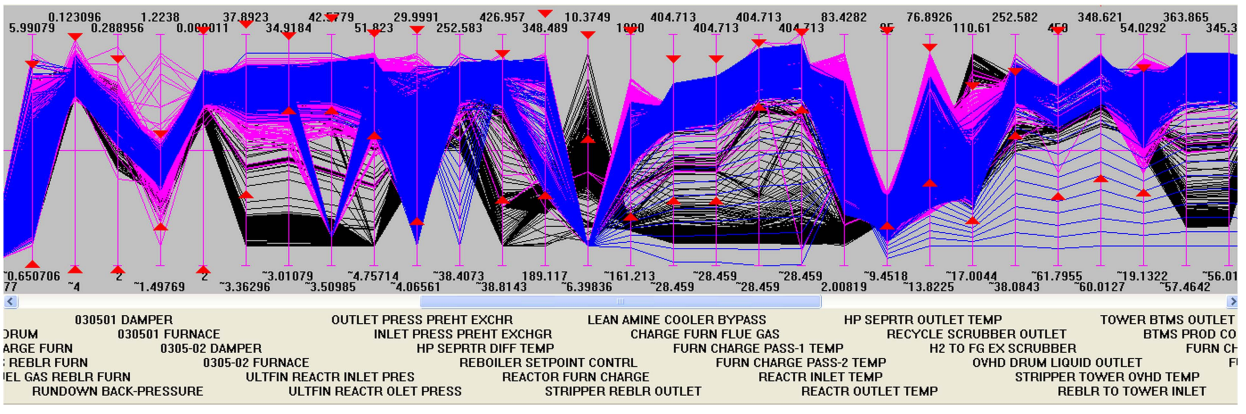

Figure 10. The All-Operation Envelope separated into its sub-Envelopes of Kerosene Mode in pink, LGO Mode in blue and Standby Mode in black

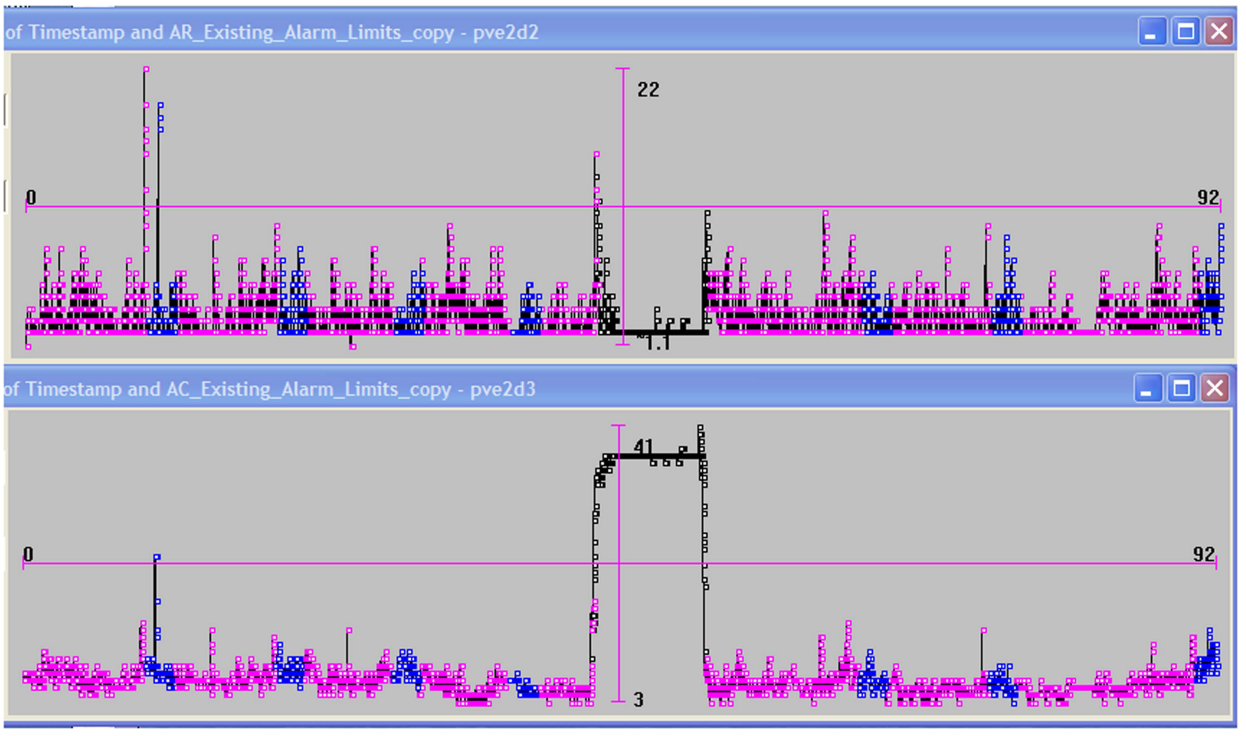

Figure 11. Existing alarm performance Annunciations per hour (top) and Alarm Count (bottom) vs. time for three month of operation

to delay annunciation in both normal and flood operation. Dependence on operator vigilance is increased and he has less time to respond. He also has to deal with a disturbance that has had more time to develop so may need a larger corrective action with a likelihood of over-correcting and propagating the disturbance thus extending the apparent duration of a flood.

As an illustration of these points, Figure 15 shows the history of a very large alarm flood lasting around 8 hours in the separation section of an ethylene plant. It has been divided into zones for convenience of description. The colours indicate the three alarm priorities in use. Cyan are the lowest priority 'Caution' alarms, magenta are medium priority 'Warning' alarms and yellow are highest priority 'Critical' alarms.

Zone A was normal operation before the disturbance event with no alarms present.

In Zone B the event has happened and disturbance is visible on the normally flat-lined process variable but no alarms are present.

Five Low priority Caution alarms appear in Zone C, delayed 1 minute and 35 seconds after the first indication of a disturbance by wider than necessary alarm limits in normal operation, and are sufficient to determine the cause

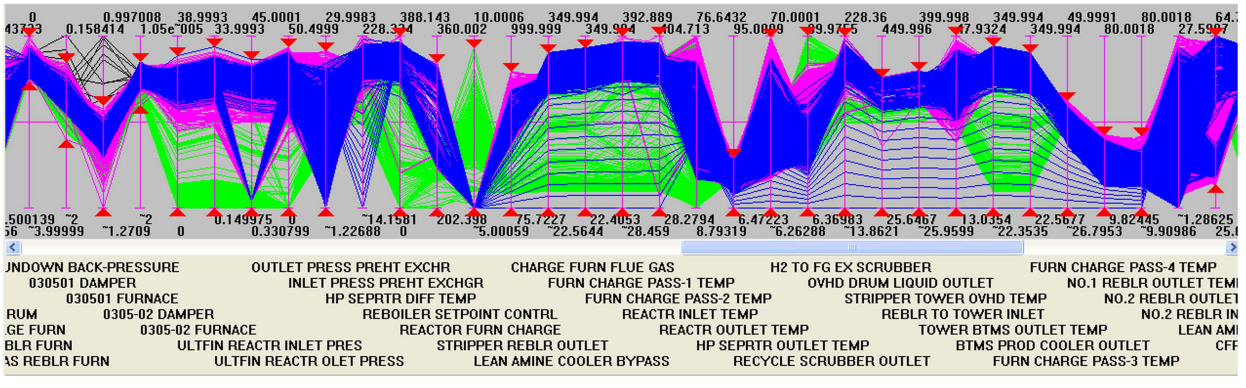

Figure 12. Kerosene Mode is in pink, Gas Oil Mode in blue and Standby Mode in green. One set of alarm limits (the red triangles) set at the boundary of where the plant has actually operated will be used for all three Modes. This is 'Lumped Mode' Alarming

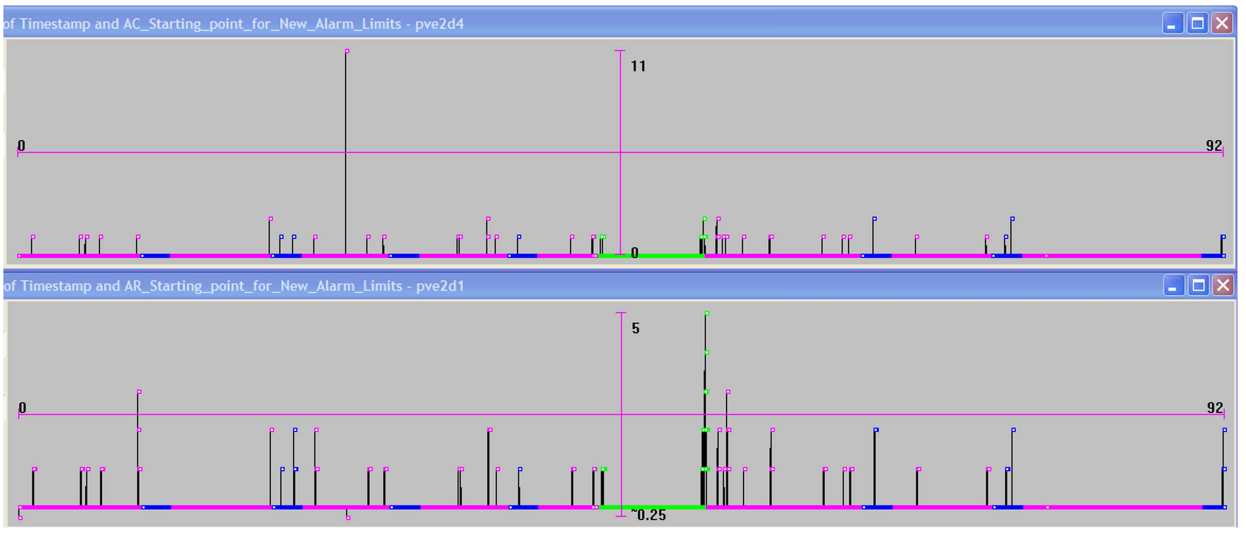

Figure 13. Standing Alarm Count (top) and Annunciation Rate per hour (bottom) for the 'lumped-mode' alarm limits of Figure 10

of the disturbance. Most of the alarms after Zone C are therefore consequential rather than informative in that they do not add to the operators understanding of the situation but raise the possibility that a second unrelated disturbance event has occurred.

In Zone D there are 3 more low priority Caution alarms and 1 medium priority Warning alarm.

In Zone E the first high priority Critical alarm appears after 5 minutes and 45 seconds just when it appears that the corrective action taken by the operator is bringing the process back.

From here on the process is settling out which is prolonged in an ethylene separation process, such as in Figure 16, by the considerable interaction between the process stream and the demand-driven refrigerant stream. Both streams have significant time constants, so that disturbances in either cause disturbances in the other which can take many hours to settle out. It seems likely that what is remembered as 'the alarm flood' is probably the many alarms in the settling period of several hours duration rather than the disturbance event of 10 alarms in 6 minutes between the first alarm and the operator regaining control.

Although the causative events can differ the settling period is similar for many of them which raises the possibility that a 'settling period' operating envelope and a corresponding set of alarm limits could be found that would reduce annunciations during settling but still be able to detect any new causative event that occurred.

Returning to the parallel plot to inspect the whole of the operation it is apparent in Figure 16 that the envelope of flood operation is much larger than that of normal operation but it has distinct boundaries. By comparing previous occurrences

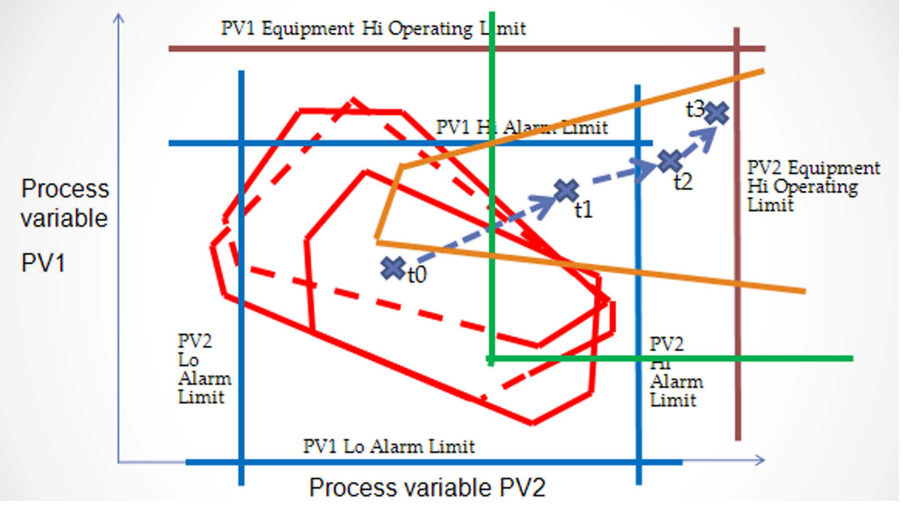

Figure 14. What happens during a process disturbance leading to an alarm flood

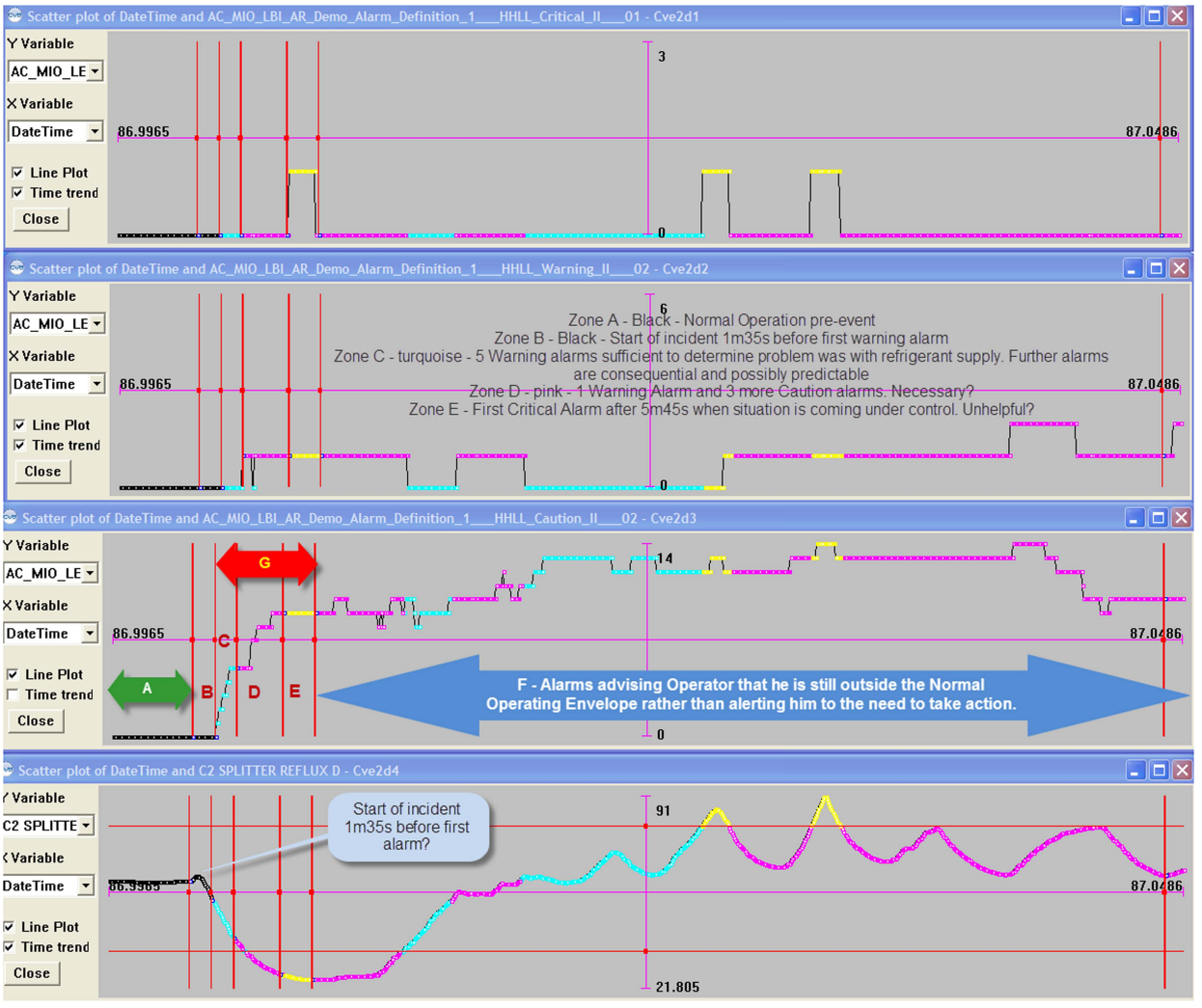

Figure 15. Life-history of an alarm flood with duration of several hours. The upper three trends show the number of alarms by priority on the alarm list display at any moment with highest priority at top. The lowest trend is of a particularly stable process variable which normally 'flat-lines' so indicates the amount of disturbance still present in the process

of alarm floods as in Figure 13 and in Figure 15a set of alarm limits could be found that would be activated by the operator after he regains control, such as in Zone E of Figure 15.

This is recognising the 'settling period' as a different operating Mode just as was done earlier in the HDS-Unit example for Kerosene and LGO operating Modes.

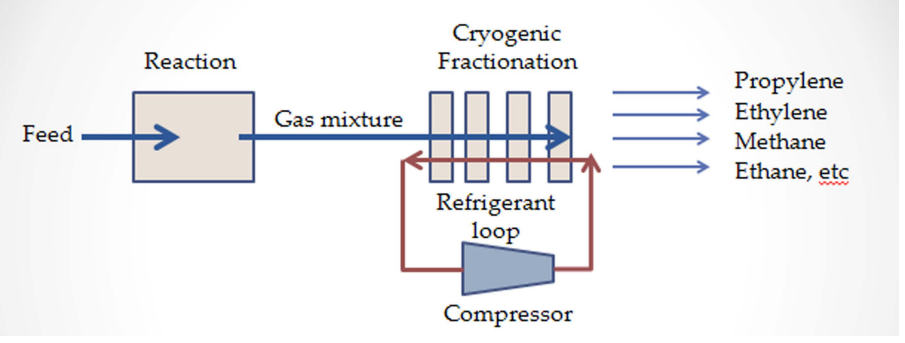

Figure 16. Interacting loops in an Ethylene separation process

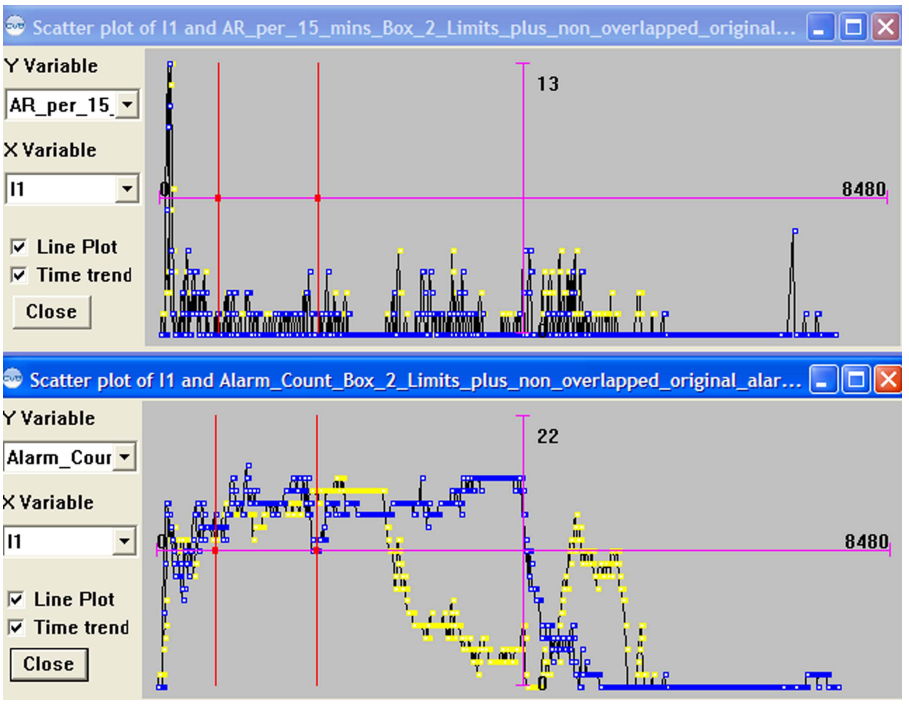

Figure 17. Two 'flood' events aligned by time-shifting to allow easy comparison by overall annunciation rate and alarm count

The distinction between State-based control and Mode-based control is one of scale. State-based control usually operates over a small range of equipment and detects state-changes automatically to suppress alarm annunciation. A typical example of state-based control would be a primary and standby pump with state-based control only allowing an alarm to annunciate when both pumps were stopped and perhaps suppressing other consequential alarms in the immediate vicinity. Mode-based alarm might refer to the whole process unit, would be unlikely to have an automatic state-detector because of the overlap between different mode-envelopes, and would probably have an

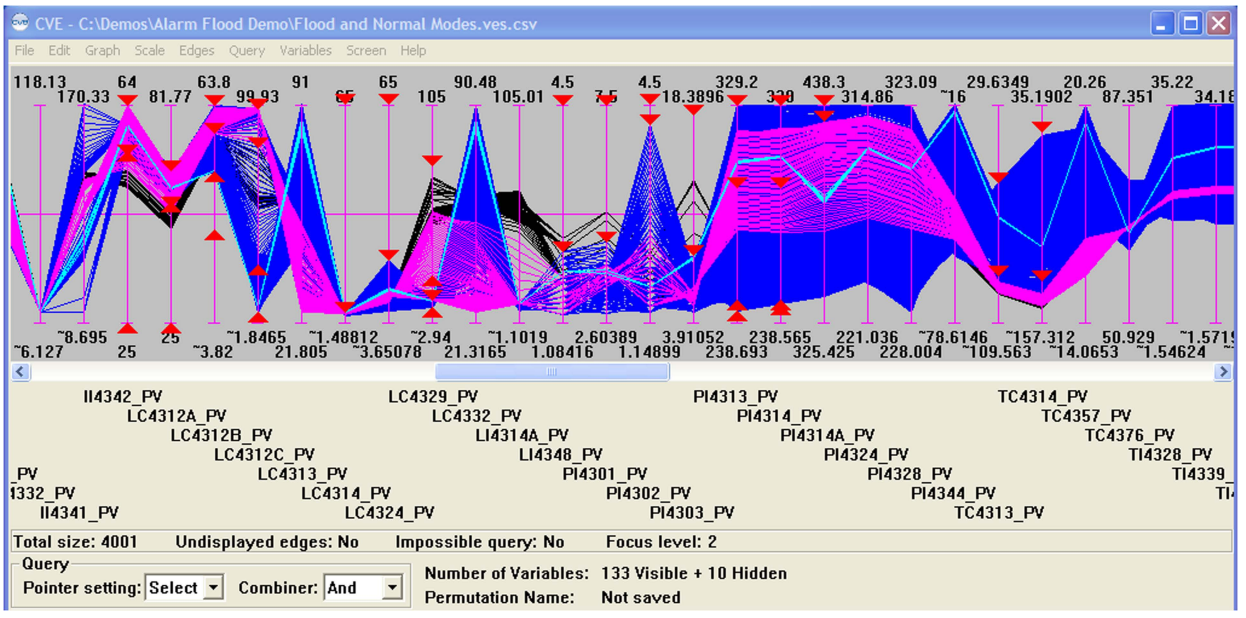

Figure 18. Sections of the Normal Operating Envelope in pink and the Flood Operating Envelope in dark blue

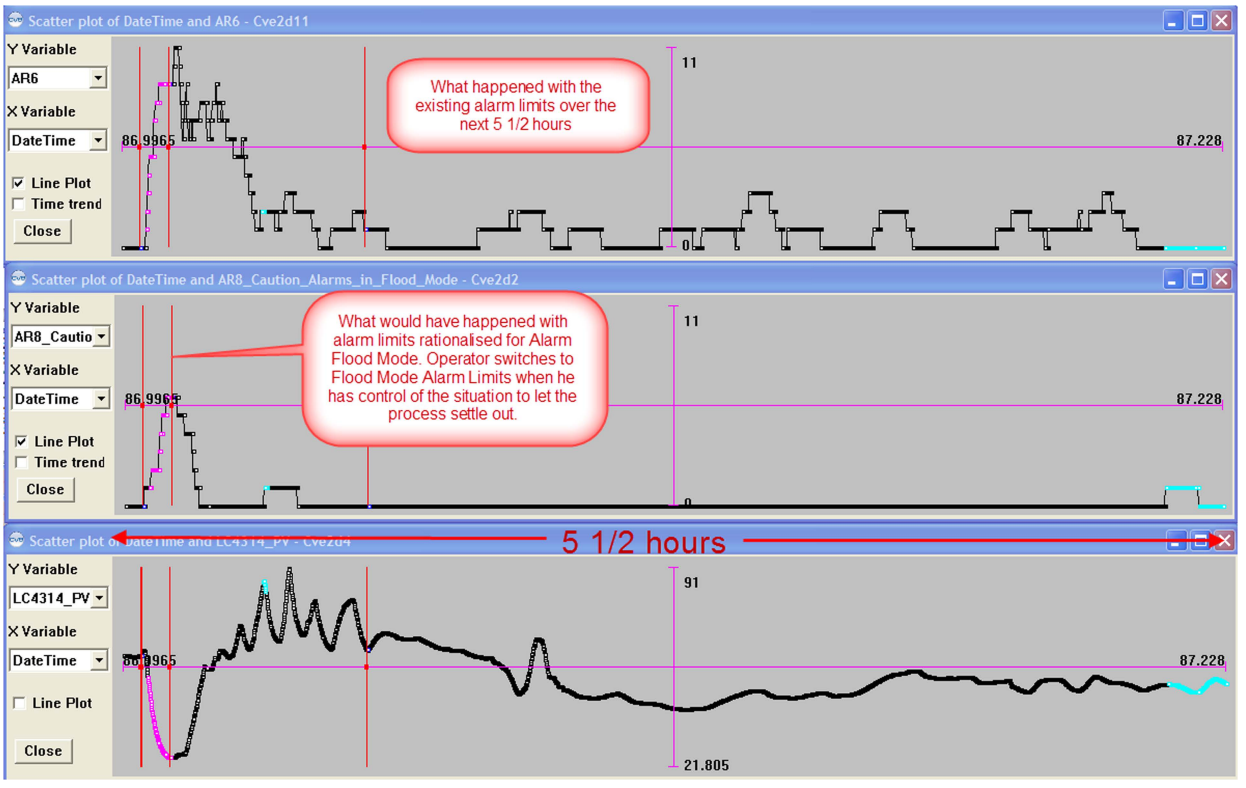

Figure 19. Comparing Normal and Settling-mode alarm limits

entirely different set of alarm limits as in this settling-period example.

Continuing the ability to predict the performance of any set of alarm limits by calculating the annunciation rates and alarm counts that will result, the calculations are easily extended by the parallel plot to predict the effects of existing or proposed state-based and mode-based controls before implementation in the DCS as in Figure 17.

Figure 17 shows the annunciation rate during the settling period with the original alarm limits and below it the annunciation rates with alarm limits set to the boundary of the settling-mode envelope with just one alarm limit tightened slightly and occasionally annunciating in cyan. In the bottom trend is a process variable gradually settling out as the process stabilises. The duration of the plot is 5.6 hours. The annunciation rates have been shown in this case based on the most-recent 10 minute interval rather than discrete 10-minute intervals as this is believed to better match human perception of alarm system performance.

Further improvement is possible by using a better approximation of an operating envelope than a hypercube to take into account variable interactions and implement predictive alarming. This was introduced in an earlier paper[2] where it was noted that One model can handle multiple Modes of operation by including the Mode number as a variable in the model.

REFERENCES

1. Inselberg, Parallel Coordinates, DOI 10, 1007/978−0-387−68628−8_5, Springer Science + Business Media 2009
2. A New Method for Defining and Managing Process Alarms and for Correcting Process Operation when an Alarm Occurs. Brooks R.W., Thorpe R., Wilson, J.W. Journal of Hazardous Materials 115(2004) pp 169−174.