Finding patterns in process signals

by Carl Sandrock

Anyone who has tried to understand events at a chemical plant will have spent some time looking at a multitude of process trend graphs. The reason for doing so is often related to process failure. When one is trying to determine why something went wrong, or to establish whether things are likely to go wrong in the near future, it is helpful to build maps of the situations that are related to failures.

Process trends represent the values of measurements taken all over the plant as they change over time. Figure 1 provides an example of a graph to illustrate chemical process data trendlines. There are a couple of ways in which such a graph may be interpreted to understand what the curves on the graph reveal. For example, one may look at the position of all the measurements at a particular time, at how signals are correlated, given some shift in time, and at the patterns that seem to be related, even though they occur at different times.

Classification and clustering

A first approach would be to see the data as clouds of points in a multidimensional space, defined by the different process variables. These points are connected because they take place at the same time. However, time is not taken into account in any other way. Figure 2 shows how one might attempt to classify points using this approach.

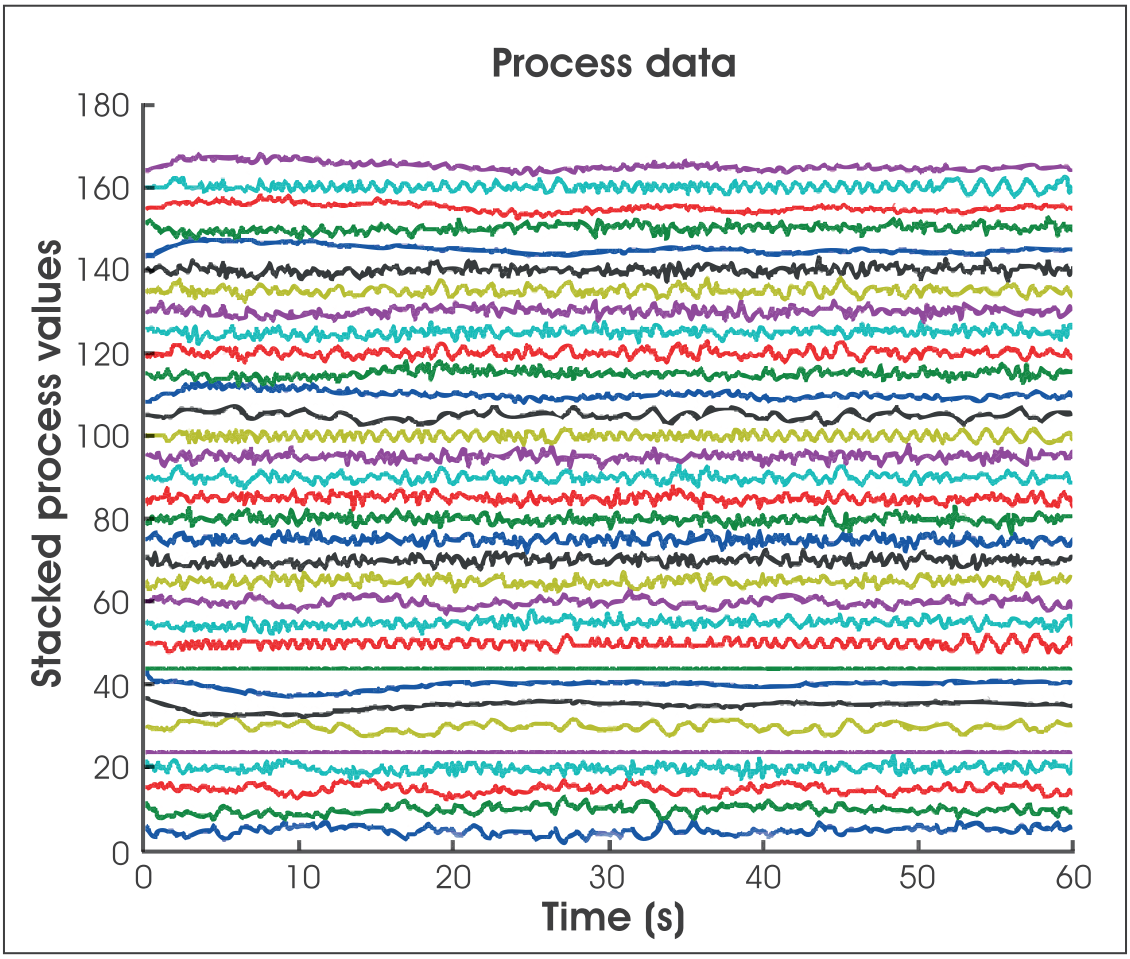
If it were known that the black points were failure points, it would be useful to see the picture in Figure 2, showing how the values of the different variables are related to their status. This problem is known as the classification problem and there are many methods that can be used to build “classifiers” or algorithms for placing a new point in the correct class, based on some known points. But what if the points shown in Figure 2 were all black? Would one be able to figure out that there were two discrete groups? This is the clustering problem: trying to find different clusters of points in this space. There are a similarly large number of clustering algorithms to address this type of problem. Typically, clustering is followed by classification.

The relevance of time

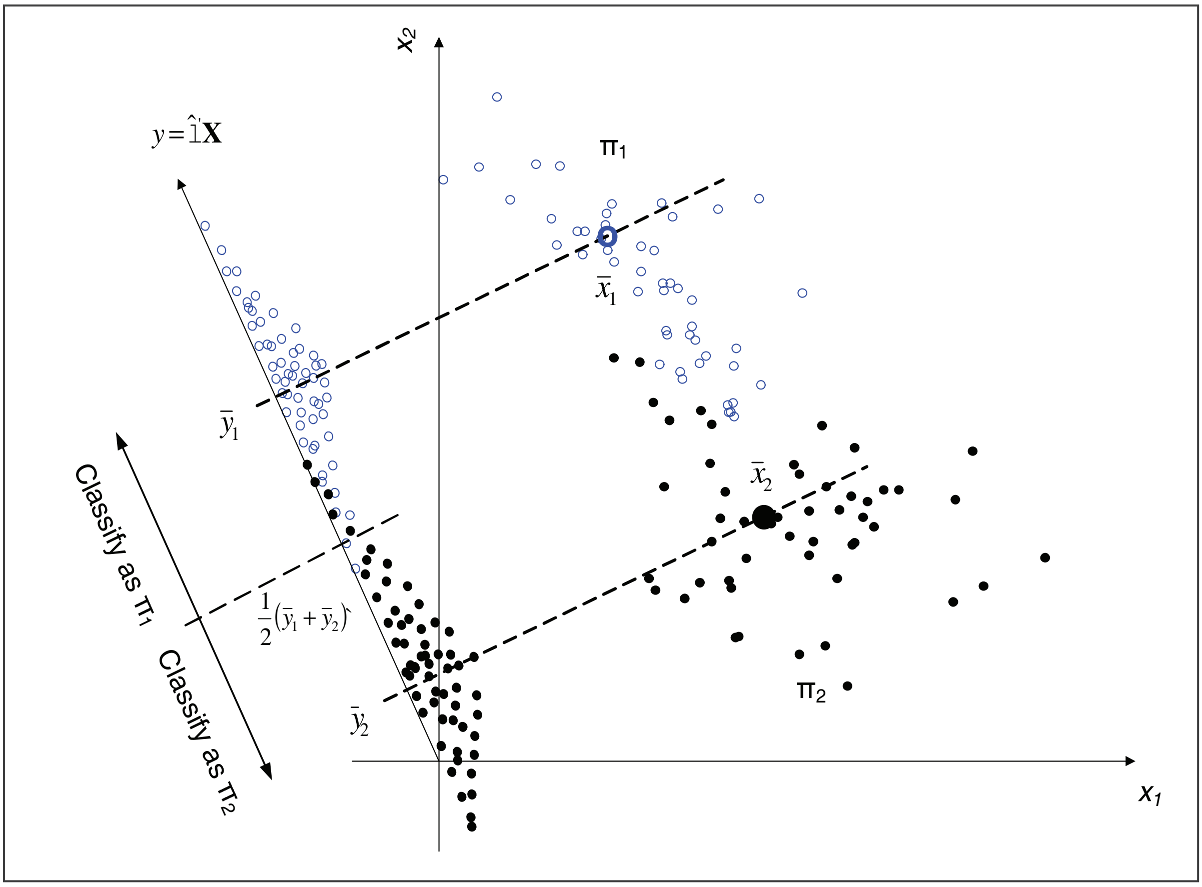
So far, the effect of time has been ignored, except as an index to see which measurements are related to one another. Unfortunately, it may not be the best idea to group the points in this way, as the same event may cause changes at different points in time. The Process Modelling and Control Group in the Department of Chemical Engineering at the University of Pretoria has developed a method of using the correlation between different signals to shift the signals in such a way that the effects all line up, as indicated in Figure 3. Shifting the times in this way allows both clustering and classification to identify single-event causes of failure propagating through the system, rather than just finding out which states are good or bad.

Patterns in time

So far, the patterns formed in the state space of the system have been discussed without looking at the patterns that appear in the signals as they are plotted in Figure 1. The combination of peaks and troughs that is visible on those trends should also be informative. In order to find patterns in this way, it is possible to look at signal processing techniques like spectral or wavelet analysis, but this frequency information is often difficult for plant personnel to interpret. It is also possible to fit simpler patterns to a time series using segmentation algorithms. The result of this is a sequence of particular patterns, like “rising”, “falling” or “steady”. If these are converted to letters in an alphabet (A=rising, B=falling and C=steady), one could search for repeating patterns in sequences like AB**ABC**ABAAA**ABC**. Figure 4 shows another way of assigning letters based on the values of the variable (low, medium or high).



1. Chemical process data trendlines.

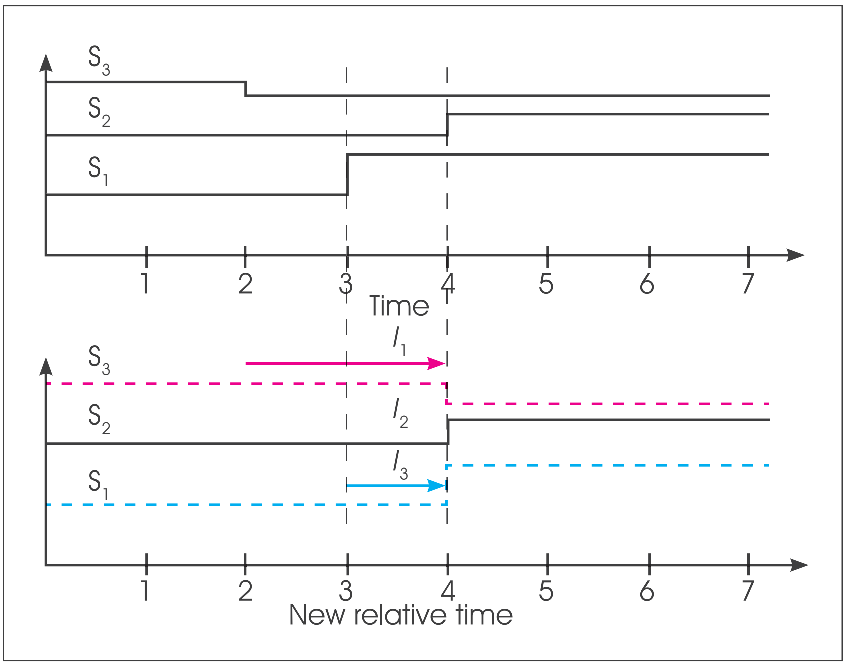


2. Classifying data based on their positions in space.

What is the relevance of finding patterns in process signals?

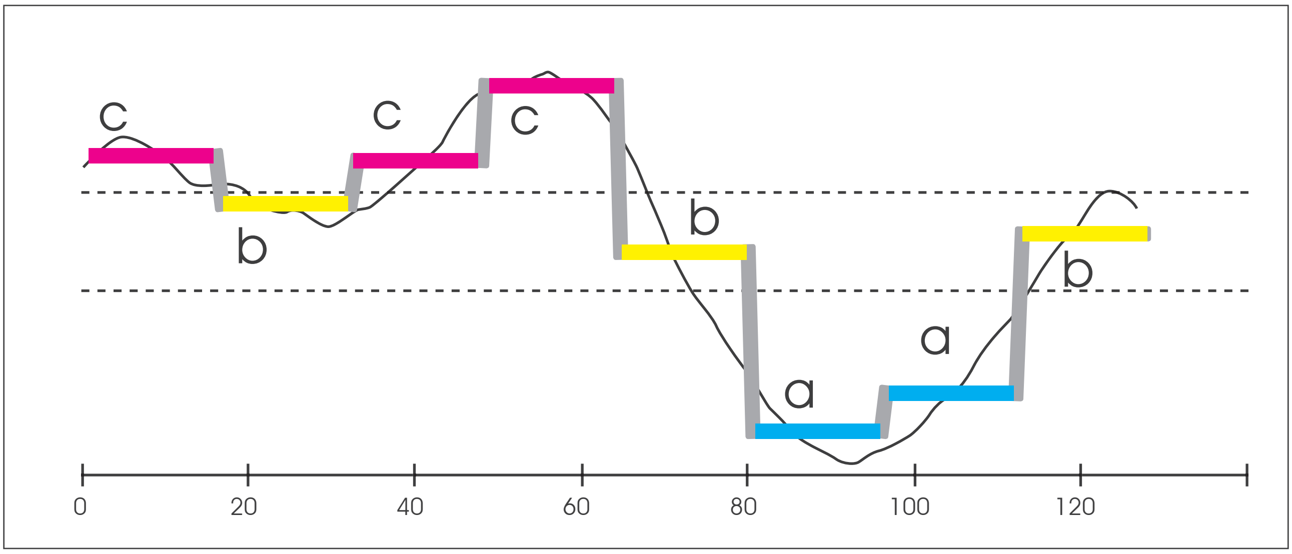
The Process Modelling and Control Group has been finding patterns in process data since 2000. This has yielded several master's dissertations and articles on the subject, working closely with industry to provide methods to learn about the failure modes of processes. By automating the discovery of patterns, the number of variables that can be analysed and the likelihood that patterns will be discovered have increased. Interest in this research is increasing every year, due to the increased availability of measurements and more stringent safety and energy requirements.

1. Chiu B., Keogh, E. & Lonardi, S. 2003. Probabilistic Discovery of Time Series Motifs. *9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* August Vol. 24–27. United States of America: Washington. Pp 493–498.
2. Labuschagne, P.J. 2008. Automatic clustering with application to time-dependent fault detection in chemical processes. Master's dissertation. South Africa: University of Pretoria.
3. Phillpotts, D.N.C. 2007. Nonlinear fault detection and diagnosis using Kernel based techniques applied to a pilot distillation column. Master's dissertation. South Africa: University of Pretoria.



3. Shifting the signals on the original time scale allows the events to line up at the same time on the new time scale.

***Carl Sandrock** is a full-time lecturer at the University of Pretoria and is currently completing his PhD. His research interests centre around the application of modern computational technology to the modelling and control of chemical engineering processes.*



4. Assigning letters to different values in a time signal for pattern search.