GET VALUABLE INFORMATION FROM THE DATA GRAVEYARD

M. Bauer[1]

[1]Department of Electrical, Electronic and Computer Engineering
University of Pretoria, South Africa
margret.bauer@de.abb.com

ABSTRACT

Continuous measurements such as temperature, pressure, or flow are recorded and stored in the plant data history of industrial chemical plants. Frequently, data is accumulated and kept in storage without anyone drawing further conclusions from it about the state of the process. This paper shows how historical data can give insight into the operation of a chemical plant. The time and frequency analysis methods are explained via an industrial process at Eastman Chemical Company, Tennessee. A procedure for systematic data analysis is given. Combined with expert knowledge of the process, causes of disturbances affecting the process can be identified.

OPSOMMING

Kontinue meting van veranderlikes soos temperatuur, druk en vloei word aangeteken en bewaar as 'n omvangryke dataversameling by 'n industriële chemiese aanleg. Heel dikwels gebeur dit dat sodanige inligting nie verder gebruik word vir die bepaling van prosestoestand nie. Díe navorsing toon hoedat sodanige inligting nuttig aangewend kan word vir die bedryf van die proses. Tyd- en frekwensie-analisemetodes word voorgehou vir 'n nywerheidsproses by Eastman Chemical Company van Tennessee. Sistematiese data-analise word volgens prosedure uitgevoer. Deur laasgenoemde te kombineer met proseskennis kan die oorsake vir prosessteurings uitgewys word.

---

[1] The author was a Ph.D. student at the Department of Electronic & Electrical Engineering, University College, London.
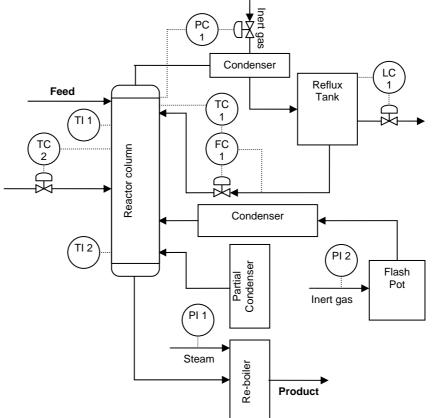
1. INTRODUCTION

Data acquisition systems gather data from measurement and control instrumentation in the plant, such as distributed control systems (DCS) or programmable logic controllers (PLC). Often in this context the term SCADA (supervisory control and data acquisition) is used for data capturing systems in industrial plants. Most SCADA systems have facilities to record, present, and store data. However, the quality of the human machine interface decides the system is to be used for process monitoring tasks. In the chemical industry, one of the commonly used data historian systems is the PIsystem, a software product by OSIsoft [2]. A frequent problem of data storage systems is data graveyards – that is, the situation where data is kept in storage and not further examined and used for performance investigation – which has been discussed and addressed [3].

The historical data can be used to give valuable insight into disturbances that affect the process, and that should be removed. If a fault occurs in a part of a continuous process, it often results in a variation in the process measurements closest to the root cause [4]. If such measurements can be identified using historical data, the root cause can be pinpointed and the cause of the disturbance can be removed. In this paper, several data-driven analysis methods are reviewed and applied to an industrial process. A classification of recent analysis methods is given by Thornhill and Horch [5]. Some important characteristics and objectives of data-driven analysis methods are as follows [6]:

- **Turn data into concise targeted information:** Analysing the process data often results in information overload. Analysis measures should extract the most important signatures in the process and discard irrelevant features.
- **Streamline and reduce troubleshooting time:** After the plant personnel flag a persistent disturbance, the origin of the disturbance is sought. This is often done by retrieving data from all measurements of the process.
- **Yield information to enhance maintenance efforts during plant shut-down:** Processes that run continuously are shut down at regular intervals to perform maintenance. During shut-down, equipment is tested and replaced. Any explanatory information about plant problems assists in focusing the maintenance efforts.
- **Discover problems not found with traditional "fight today's fire" approaches:** Plant problems may exist of which neither plant personnel nor process engineer are aware. Analysis methods can identify disturbances that have escaped the standard investigative tools.

In this paper, a systematic approach for the use of data-driven methods is illustrated using an industrial process. The process is part of a larger production facility at Eastman Chemical Company in Kingsport, Tennessee, and has been described previously ([7], [8]). A number of disturbances affected the process. These disturbances will be investigated in detail in the following sections. First, available information about the process is reviewed and interpreted. In Section 3, the time trend given in the form of the historical process data is investigated, and two simple but effective methods are described that can assist in identifying the disturbance. In Section 4, the use of the frequency spectrum for process analysis is discussed. The

results obtained from these analysis methods are consolidated and discussed in Section 5.

Figure 1: Process schematic of a reaction process at Eastman Chemical Company

2. PROCESS INFORMATION

Most industrial production processes are well known to the plant personnel, and in-depth expert information is often available to make sense of the current operation of the plant. This existing information platform – comprising process schematics, mathematical models, or written plant descriptions – is relevant for analysis of disturbances and process performance. Not incorporating the process information in the analysis would be a waste of useful process insight.

2.1 Process Schematic

Information about the process exists in the form of process and instrumentation diagrams (P&IDs)[2] which are often simplified in a process schematic. Process schematics show the most important parts of the process: all actuators such as control valves and pumps, and the controllers acting on the actuators. The main reactions and flows can be seen from the process schematic. The process schematic of the case

---

[2] P&ID design handbook, http://www.engineeringtoolbox.com/p&id-piping-instrumentation-diagram-44_446.html

study is given in Figure 1. The feed enters the reactor column as the first reaction component at the top. The reaction takes place inside the column, and the lighter reacted material exits the column at the top while the heavier material, which will become the product, exits the column at the bottom. The lighter material is recycled, and a portion of the recycle is fed back into the column via a condenser and reflux tank. Further reactants enter the column from external streams. The product that exits the column is available for further processing after it is heated in a sequential re-boiler. Thus, the product flow is from the top of the column to the bottom and onward to the re-boiler. Root cause analysis is complicated by the recycle path via the reflux tank. Because of the recycle, disturbances can travel not only from top to bottom but also in the other direction.
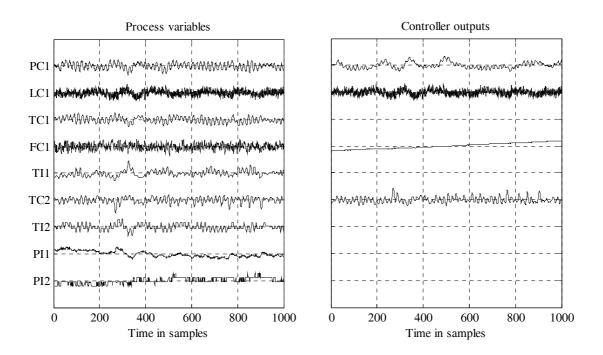

Figure 2. Time trends of process variables and controller outputs for the measurements indicated in Figure 1

2.2 Process measurements and controller setup

Process schematics indicate where and which types of measurements are taken. The regulating control schemes in place are usually also indicated in the schematic. In the industrial example (see Figure 1), nine process variables are measured and recorded: four temperatures around the column (TC1, TC2, TI1, TI2), the level at the reflux tank (LC1), the flow rate of the recycle stream (FC1), as well as the pressure at the top of the column (PC1) and the inflow pressures PI1 and PI2. Five of the measurements are used to control flows via control valves. Since the temperature in the reactor is an important quantity, it is controlled by two temperature controllers (TC1 and TC2) via the input streams and a cascade loop with FC1. The level of the reflux tank is controlled to avoid overflowing, while the pressure in the column is controlled through the inflow of inert gas in the recycle path.

It is important to note the sampling period. If the sampling interval is too long, then

fast dynamics might be lost when capturing the data. Normal industrial sampling intervals are 30 seconds or 1 minute for logged data. The data in the industrial example is in these terms logged relatively fast, with a sampling interval of 10 seconds. A set of six samples covers one minute of operating time.

Tag name	σ_x %
PC1.PV	0.6%
LC1.PV	0.2%
TC1.PV	0.3%
FC1.PV	0.1%
TI1.PV	1.1%
TC2.PV	0.3%
TI2.PV	0.1%
PI1.PV	0.5%
PI2.PV	0.8%

Table 1: Plant variability statistics as percentage of average

3. TIME TREND INFORMATION

Process insight can be gained from simple visual inspection of the time trend. It is useful to combine a number of time trends in one graph, as shown for the industrial example in Figure 2. Plotting the time trend of several measurements in one graph shows similar features that are not easily seen when examining one time trend at a time. The time trends of the process variables and controller outputs in Figure 2 show the presence of a fast oscillation in all the temperature measurements around the reactor column as well as in the reactor pressure PC1. The setpoint is not shown in Figure 2, since no changes in setpoint occurred within the captured time frame. Level LC1 exhibits a high frequency noise and pressure PI1 long term fluctuations. Pressure PI2 and temperature TI2 appear to be highly quantized – that is, only discrete amplitude levels are adopted. Quantization occurs if the analogue-to-digital converter (ADC) has a large measurement quantity assigned to the least significant bit (LSB) relative to the total variation. If quantization appears in the time trend, data-driven analysis has to be taken into consideration, as the time trend is distorted. Quantization is often a problem in temperature measurements, due to low accuracy.

3.1 Variability analysis

Assessing the variability is useful to evaluate the extent and severity of a disturbance. However, if used as an absolute value, variability can be misleading, and it must therefore be compared to the average value of the variable. The variability is measured by the standard deviation, as follows:

$$\hat{\sigma}_x = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x_n - x_{avg})^2} \tag{1}$$

where x_n is the process variable at time sample n, N is the number of samples, and

x_{avg} is the average value of the process variable. The standard deviation is interpreted in relative terms – that is, as a percentage of the average of the process variable (% coefficient of variation).

$$\sigma_x \% = \frac{\hat{\sigma}_x}{x_{avg}} \%$$
(2)

Table 1 lists the standard deviation for the industrial case study. The plant variability is less than 1% for most variables, and therefore not necessarily severe. The highest variability can be observed for temperature TI1 at the reactor column. For large differences in variability, the process variable that exhibits the highest standard deviation may most likely be the root cause. In the industrial example only small differences in variability can be observed.
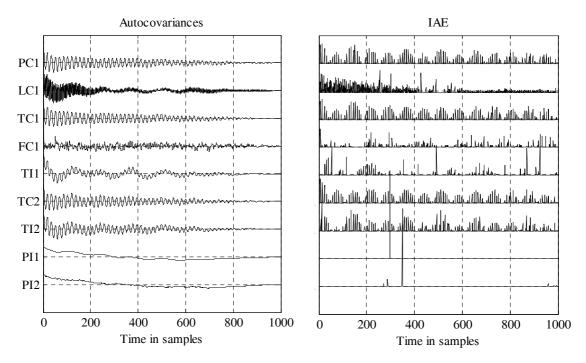

Figure 3: Autocovariances and integrated absolute error of process variables

3.2 Oscillation analysis

Common disturbances in the time trend that give particular rise for concern are oscillations. Oscillations can be caused by inappropriate tuning or by hardware problems such as valve stiction [9]. A real-time oscillation detection method presented by Hägglund [10] investigates the time between zero-crossings, and thus determines whether a time trend exhibits oscillation. The main computation of the statistic is the calculation of the integrated absolute error (*IAE*) defined by the following expression:

$$IAE_i = \sum_{k=k_i}^{k_{i+1}} |Y_k|$$
(3)

150

where Y_k is the autocovariance of the controller error signal at time shift k. Furthermore, k_i and k_{i+1} are the times of successive zero crossings of Y_k. The autocovariance is used instead of the time trend to remove high frequency noise effects. The autocovariance measures the similarity of a signal with a time-shifted version of the same signal, and can be derived from the process variable time trend x with N samples, as follows:

$$Y_k = \frac{1}{N} \sum_{n=0}^{N-k-1} x_{n+k} x_n$$

(4)

An interval between two zero crossings is defined as $\Delta k_i = k_{i+1} - k_i$. Regularity is assessed by the use of a statistic q which is defined as follows:

$$q(\xi) = \frac{R_{avg}}{\hat{\sigma}_R}$$

(5)

where the ratio R between adjacent intervals Δk is as follows:

Tag name	q	T_p
PC1.PV	**4.15**	**16.9**
LC1.PV	0.63	4.2
TC1.PV	**3.20**	**16.9**
FC1.PV	0.77	8.8
TI1.PV	0.64	24.5
TC2.PV	**4.97**	**17.0**
TI2.PV	**1.84**	**16.9**
PI1.PV	0	0
PI2.PV	0.49	153.1

Table 2: Oscillation index q and period of oscillation

$$R_i = \frac{\Delta k_{i+1}}{\Delta k_i}$$

(6)

and from which the average value R_{avg} as well as the standard deviation $\hat{\sigma}_R$ are derived as in Equation (1). ξ in Equation (5) is the threshold for detection, and should be set to $2/\pi$ if sinusoidal oscillations with unit amplitude are present in the presence of noise and r.m.s value of 1 [11]. A regularity index q that is significantly larger than 1 indicates oscillation. The oscillation period of the signal can be calculated from the detected intervals between the zero-crossings, as follows:

$$T_p = 2 \cdot \Delta k_{avg}$$

(7)

Figure 3 shows the autocovariances and the *IAE* for the process variables. A sinusoidal oscillation with a similar period can be seen in the temperature

151

measurements, as well as in the pressure in the column, PC1. The IAE shows the position of all zero crossings that occur at regular intervals for the variables with the sinusoidal oscillation. Table 2 shows the oscillation index q and the period of oscillation T_p. Four variables have an oscillation index significantly larger than 1, namely PC1, TC1, TC2 and TI2. The period of oscillation for these variables is approximately 16.9 samples, that is, around 169 seconds, or slightly less than 3 minutes. A three-minute oscillation is considered fast in most chemical processes. TI1, however, clearly shows the same rapid oscillation in the autocovariance function. The three-minute oscillation is, however, superimposed on a slower oscillation with higher amplitude, and it is therefore not picked up by the oscillation index.

4. FREQUENCY SPECTRA

The advantage of the oscillation index described in the previous section is that oscillations can easily be detected, and the index can be used as an online monitoring tool. In fact, the frequency spectrum analyses the frequencies contained in a time series. The frequency spectrum is derived via the discrete Fourier transform (DFT), as follows [12]:

$$X_m = \sum_{n=0}^{N-1} e^{-2\pi imn/N} x_n$$

(8)

where N is the number of samples of time trend x_n and i indicates an imaginary number. The resulting sequence X_m, $m = 1\ldots N$, is complex – that is, it has an imaginary and a real part. Only the absolute values of X_m, $|X_m|$, will be regarded in the following, and are often referred to as the power spectrum. The power spectrum reflects the intensity with which a sinusoidal function with frequency $2\pi m/N$ is contained in the time sequence x_n.

Figure 4 shows the power spectra of all process variables and controller outputs in the industrial example. The fast oscillation that was detected with the oscillation index can be seen as clear peaks in the temperature measurements TC1, TC2, TI1, and TI2, as well as in the column pressure PC1 and, to a lesser extent, in the reflux tank level LC1. The same frequency peak can be seen in the power spectra of the controller output signal of PC1, LC1, and TC2. The power spectra also reveal a rapid oscillation in the reflux tank level LC1 and the reactor pressure PC1. In the following, several approaches to analyzing the frequency spectrum in order to extract information about the process will be discussed.
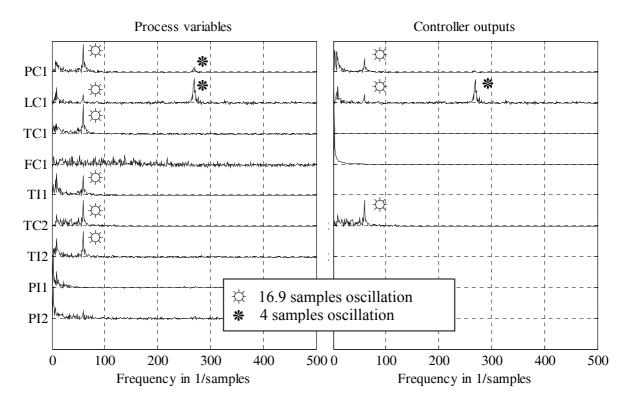
Figure 4: Power spectra of process variables and controller outputs for the measurements indicated in Figure 1

4.1 Drill-down tool

The frequency spectrum can be used to identify similar process variables, and thus be used as a drill-down tool to focus on a smaller number of process variables that show oscillations with the same oscillation period. The root cause of the oscillation is expected to be close to one of the oscillating process variables. In the industrial example, the analysis can be focused on the temperatures in the reactor column (TC1, TC2, TI1, TI2) as well as the pressure PC1, since the oscillation is most prominent in those measurements.

4.2 Root cause analysis

The height of a frequency peak in the frequency spectrum can be seen as the strength of the oscillation present in a signal. High amplitudes in the oscillation can be an indication that the variable is close to the root cause, since the oscillation is often attenuated as the disturbances travel through the process. This guideline, however, has to be viewed with caution, since in a first signal only one disturbance may be present, while in a second signal, several disturbances might be present. The frequency peak in the first will automatically be lower, no matter whether the variable is closer to or further away from the root cause. In the industrial example, for which the frequency spectrum is shown in Figure 4, the highest frequency peaks for the 16.9 sample oscillation can be observed for pressure PC1 and temperatures TC1 and TC2. Since these are all measurements at the top of the reactor column, it is likely that this part of the process is closest to the root cause. Further investigation

showed that the oscillation was caused further upstream and entered the process through the feed into the reactor. The 4 sample oscillation shown in Figure 4 can be seen both in LC1, the reflux tank level, and in PC1, the pressure at the top of the column. Since the oscillation is significantly stronger in the level measurement, this variable is more likely to be closest to the root cause.
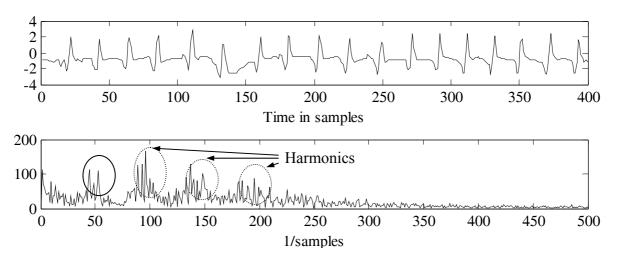

Figure 5: Non-sinusoidal, oscillating signal and frequency spectrum

4.3 Analysis of several oscillations in one signal

One advantage of frequency analysis is that several oscillations present in the same time trend can be detected and investigated. For example, LC1 shows the 16.9-sample oscillation, the 4 sample oscillation, and some low frequency components that can be clearly identified in the frequency spectrum. The analysis of multiple oscillations could not be achieved by the oscillation index. The frequency spectrum, however, is an off-line analysis tool, since all the data has to be gathered to conduct the analysis.

4.4 Harmonics as indication of nonlinearity

Oscillations that are not of a sinusoidal nature show frequency peaks not only at the oscillation frequency, but also at a multiple thereof. The frequency peaks at the multiples of the oscillation frequency are called harmonics. Process variables with harmonics are often nonlinear [6] and tend to be close to the root cause. The reason for this is that the higher frequency components are often removed as the disturbance travels through the process. This is because most processes act as low pass filters, which let low frequency components pass while high frequency components are suppressed. Since none of the disturbances in the industrial case study were significantly different from a sinusoid, and hence showed no harmonics in the frequency spectrum, a further process variable is used to show the harmonics. The top panel of Figure 5 shows the time trend of a process variable with a non-sinusoidal oscillation of around 24 samples. The lower panel of Figure 5 shows the frequency spectrum, and the main oscillation can be seen at around 48 per sample.

However, further peaks – the harmonics – can be observed at 2x48=96 and 3x48=146 per sample.

5. CONCLUSIONS

In this paper, a systematic approach to analyzing historical operational data from a chemical process was presented. First, an inspection of the time trend for disturbances was conducted, and the extent of variability as a percentage of the mean was assessed. Process variables that show an excessive variability were flagged. In a next step, an oscillation index can be computed to establish the oscillatory behavior of a variable. The oscillation index estimates both regularity and period of the oscillation. In the case of more than one oscillation in the signal, a frequency analysis can be conducted. Frequency analysis further assists in the identification of the root cause, and can help to focus on a reduced number of process variables. The data is thereby retrieved from the data graveyard and turned into concise process information.

An industrial case study was presented of a process at Eastman Chemical Company. The measurements in the process were affected by a number of oscillatory disturbances. The variability analysis showed that the disturbances, though persistent, did not have a large impact on the process variables. However, since the disturbance affected temperatures in the central reactor column (which are critical to the reaction), the effect cannot be ignored. The oscillation index identified four variables that were affected by an oscillation with the same period of 16.9 samples. Frequency analysis confirmed the result, and identified a further rapid oscillation in the reflux tank level. The analysis indicated that the disturbance originated from the top of the reactor column. Further investigation showed that the disturbance was caused further upstream in the plant by an incorrectly tuned split-range controller, and entered the process through the feed. The data-driven analysis methods could therefore successfully identify the disturbance origin.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] **Avlonitis, S.A., Pappas, M., Moutesidis, K., Avlonitis, D., Kouroumbas, K. and Vlachakis, N.** 2004. PC based SCADA system and additional safety measures for small deslination plants, *Desalination*, 165, 165-176.

[2] www.OSIsoft.com, June 2006.

[3] **Estler, M, Hilpert R and Soravia, S.** 1998. Exploring the data world: How can we extract the maximum information from product and process data?, *Chemie-Ingenieur-Technik*, 70 (11), pp 1388-1392.

[4] **Thornhill, N.F.** 2005. Finding the source of nonlinearity in a process with

plant-wide oscillation, *IEEE Transactions on Control System Technology*, 13, 434-443.

[5] **Thornhill, N.F. and Horch, A.**, 2006. Advances of new directions in plant-wide controller performance assessment, *Proceedings of ADCHEM* 2006, Gramado, Brazil.

[6] **Bauer, M.** 2005. *Data-driven methods for process monitoring*, PhD thesis, University College, London.

[7] **Bauer, M., Cox, J.W., Caveness, M., Downs, J.J. and Thornhill, N.F.** 2006. Finding the direction of disturbance propagation in a chemical process using transfer entropy, *IEEE Transactions on Control Systems Technology*, in press.

[8] **Bauer, M., Cox, J.W. and Thornhill, N.F.**, 2005. Measuring cause and effect of process variables, *IEEE Advanced Process Control Applications for Industry Workshop*, Vancouver, Canada, 9-11 May 2005.

[9] **Choudhury, M.A.A.S.**, 2004. *Detection and diagnosis of control loop nonlinearities, valve stiction and data compression,* PhD thesis, University of Alberta.

[10] **Hägglund, T.** 1995. A control-loop performance monitor, *Control Engineering Practice*, 3, 1543-1551.

[11] **Thornhill, N.F. and Hägglund, T.** 1997. Detection and diagnosis of oscillation in control loops, *Control Engineering Practice*, 5, 1343-1354.

[12] **Mitra, S.K.** 2006. *Digital signal processing – A computer based approach*, 3rd Edition, McGraw-Hill.