contents

Opening the Black Box: Demystifying Performance Assessment Techniques

Robert C. Rice
Control Station, Inc.
One Technology Dr.
Tolland, CT 06084

bob.rice@controlstation.com

Rachelle R. Jyringi
Department of Chemical Engineering
University of Connecticut
Storrs, CT 06269-3222

rachelle@engr.uconn.edu

Douglas J. Cooper
Department of Chemical Engineering
University of Connecticut
Storrs, CT 06269-3222

cooper@engr.uconn.edu

ABSTRACT

Real-time performance monitoring to identify poorly or under-performing loops has become an integral part of preventative maintenance. While some control software packages display performance metrics, it is important to understand the theory, purpose, and limitations since each metric signifies very specific information about the nature of the process.

This paper reviews performance measures from simple statistics through complicated model-based performance criteria. By understanding the underlying concepts of the various techniques, readers will gain knowledge of how to use and implement each of the performance criteria. Basic algorithms for computing performance measures are presented using example data sets. A discussion with tips and suggestions provides guidance for interpreting the results.

INTRODUCTION

Over the past two decades, process control performance monitoring software has become a staple of any successful control engineer's toolbox. The amount of performance tests and statistics that can be calculated for any given control loop can be overwhelming. The problem with controller performance monitoring is not the lack of techniques and methods; it is the lack of guidance and understanding as to how to turn statistics into improved performance.

The performance analysis techniques discussed here are separated into three sections. The first section details methods for identifying process characteristics using batches of existing data. The second section outlines methods used for real-time and dynamic analysis of streaming process data. These are the techniques that are vital for the timely identification of changing process behavior or loop performance deterioration. The third section outlines techniques that aid in the identification of interacting control loops.
The techniques presented here use Microsoft Excel® to calculate the performance measures. Readers may obtain a complimentary copy of the Excel worksheet by contacting bob.rice@controlstation.com .

Presented at ISA EXPO 2005, 25-27 October 2005
McCormick Place Lakeside Center, Chicago, Illinois, www.isa.org

IDENTIFYING PROCESS CHARACTERISTICS

Set Point Analysis

Explained here are techniques for analyzing closed loop data during a set point response. These techniques permit an orderly comparison of process response shapes and characteristics. When analyzing a set point response, criteria used to describe how well the process responds to the change include the peak overshoot ratio, decay rate, set point crossing time, rise time and settling time. The criteria can be used both as specifications for commissioning of control loops and for documenting changes in performance due to the adjustment of the controller or process parameters.

Figure 1 shows a closed loop response to a set point change. To calculate the set point criteria listed above, certain features are assigned the following definitions:

A = Size of the set point step
B = Size of the first peak above the new set point or steady state
C = Size of the second peak above the new steady state

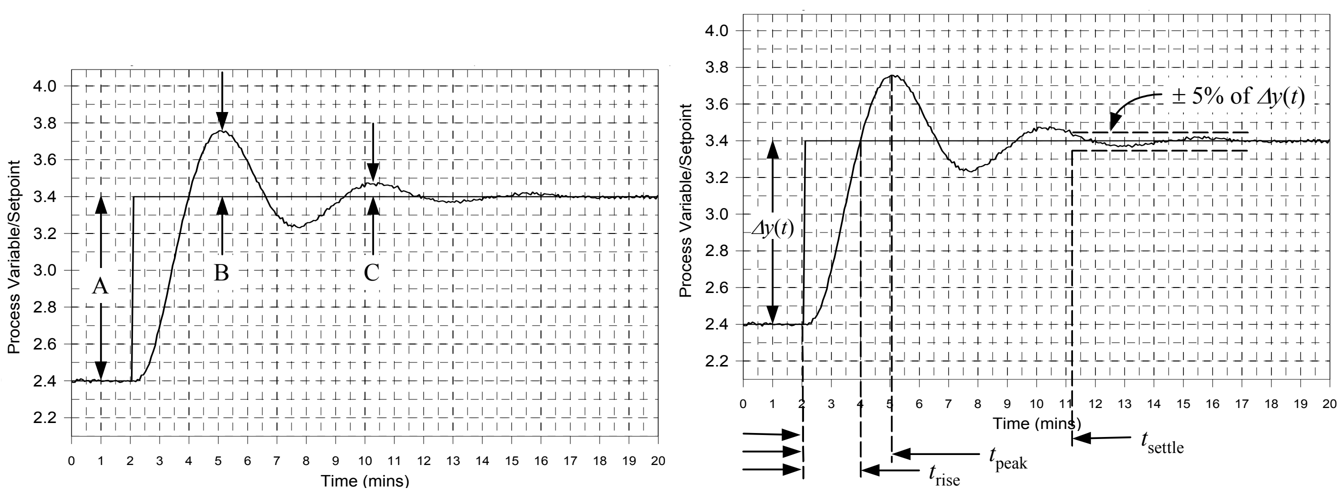

Figure 1 - Process response to a set point change with labels indicating response features

As shown in Figure 1, the time when the measured process variable first crosses the new set point and the time at which it reaches its first peak are used to describe controller performance. Another popular measurement is the settling time, which is the time required for the measured process variable to first enter and then remain within a band whose width is computed as ± 5% (or 3 % or 10 %) of the total change in $y(t)$, labeled as $\Delta y(t)$ in Figure 1.

Additional criteria are summarized in Table 1. Popular values include a 10% peak overshoot ratio and a 25% decay ratio. Also, these criteria are not independent. A process with a large decay ratio will likely have a long settling time. A process with a long rise time will likely have a long peak time.

Criteria	Interpretation	Calculation
Peak Overshoot Ratio (POR)	The POR is the amount by which the process variable exceeds the set point. An aggressive controller can increase the amount of overshoot associated with a set point change.	(POR) = B/A
Decay Rate	A large decay rate is associated with an aggressive controller, and visible oscillations are present in the set point bump. The smaller the decay rate, the faster the oscillations will dampen.	Decay Ratio = C/B
Peak Time & Rise Time	These measurements gauge the time response to a change in the set point. A large peak and rise time could be the result of a sluggish controller.	Rise Time = t_{rise} Peak Time = t_{peak}
Settling Time	The settling time is the time for the process variable to enter and then remain within a band. Time spent outside the desired level generally relates to undesirable product. Therefore, a short settling time is sought.	Settling Time = t_{settle}

Table 1 - Interpretation of Set Point Response Criteria

The integral of error indexes focus on deviation from set point. The Integral Squared Error (ISE) is very aggressive because squaring the error term provides a greater punishment for large error. The Integral Time Absolute Error (ITAE) is the most conservative of the error indexes; the multiplication by time gives greater weighting to error that occurs after a longer passage of time. The Integral Absolute Error (IAE) is moderate in comparison to these two. Additional indexes can be derived depending on the system requirements. Integral Time Squared Error (ITSE) combines the time weighting with the exaggerated punishment for larger error.

The formula for calculating the Integrated Error indexes are listed below.

$$IAE = \int_0^T |e(t)|\,dt \tag{1}$$

$$ISE = \int_0^T e^2(t)\,dt \tag{2}$$

$$ITAE = \int_0^T t\,|e(t)|\,dt \tag{3}$$

$$ITSE = \int_0^T t e^2(t)\,dt \tag{4}$$

Often the above indexes are used as criteria in controller tuning. Users will choose one of the above and define optimal control as tuning that will give the minimum value of the index.

Figure 2 shows the process variable's response to a set point change under various controller tunings ranging from poor/unstable to conservative. The results are summarized in Table 2.

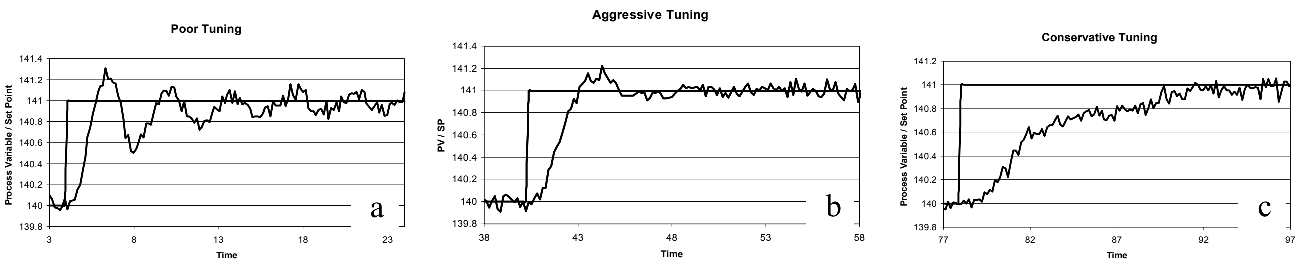

Figure 2 - Set Point Response of a a) poorly, b) aggressively, c) conservatively tuned PI controller

		Poorly Tuned	Aggressively Tuned	Conservatively Tuned
Set Point Bump Criteria	POR	33%	21%	0
	Decay Rate	44%	24%	0
	Rise Time	2.0 min	2.9 min	13.7 min
	Peak Time	2.4 min	4.1 min	13.7 min
	Settling Time	10.8 min	6.5 min	14.7 min
Integral of Error	IAE	3.10	2.28	5.49
	ISE	1.24	1.23	3.17
	ITAE	17.69	8.64	25.40
	ITSE	2.97	1.21	7.85

Table 2 - Results of Set Point Response Criteria, and Integral of Error calculations for Figure 2

Disturbance Analysis

A disturbance is defined as anything other than the controller output signal that affects the measured process variable. In an interacting plant environment, each control loop can have many different disturbances that impact its performance. By understanding the type of disturbance and its impact on the control loop, engineers, operators and technicians can more easily identify the cause and work towards a solution.

Autocorrelation is a method that is used to determine how data in a time series are related [1]. By comparing patterns in current process measurements with those exhibited in the past, the nature of disturbances and how they affect the system can be assessed.

The equation for calculating the autocorrelation relationship is:

$$r(k) = \frac{\sum_i [(y(i) - \overline{y})(y(i-k) - \overline{y})]}{\sum_i (y(i) - \overline{y})^2} \quad (5)$$

Where: y is measured process data

$\overline{y}$ is the set point or the series average if there is an offset

k is time delay in samples

i is sample number (or sample time)

Autocorrelation values are always between negative one and one. If data is random, the values will be approximately zero for all time. Any value that is significantly non-zero indicates the data is non-random. A strong autocorrelation will have an initial value near one or negative one and the trend will be linear; this shows that the each measurement dictates the next. A moderate autocorrelation is one in which the plot begins below one (or above negative one) and decreases magnitude towards zero but displays noise. An autocorrelation of closed loop data can also give an estimate of the response time for an isolated disturbance.

Another performance statistic that is useful for identifying trends in data is the power spectrum. It is calculated by computing the discrete Fourier transform of the process data. A Fourier transform is a mathematical expression of the data as a series of two-dimensional sine waves, and the power spectrum is computed by squaring the complex coefficients determined by those sine waves. The power spectrum shows the frequency at which change is occurring and the magnitude of the change [9].

The shapes and heights of peaks on power spectrum plots give information about the system. The shape of the power spectrum curve yields information about the nature of the disturbances by displaying its frequency. An increase in peak heights compared to historical data indicates that the process has greater deviation from set point or historical mean. Low powers and low frequencies are most desirable, as they are associated with small deviations from set point or lower average values.

Figures 3-6 show four different scenarios in which the autocorrelation and power spectrum can be useful in understanding the nature of the disturbance impacting the system. Only the single pulsed disturbance shown in Figure 3 is noticeable from a casual evaluation of the process data, but using the autocorrelation and power spectrum tools one can identify characteristics for all four disturbances.

In Figure 3, the process is upset with a single pulsed disturbance. The autocorrelation shown in Figure 3c shows an initial peak where the process is responding to the step up then the negative peak occurs approximately 10 minutes later when the disturbance steps back down. This is characteristic of an isolated disturbance. If a second pulse occurred, another similar pattern would appear on the autocorrelation plot. The power spectrum of the process data is shown in Figure 3d. Since the frequency of change corresponds to the frequency of the disturbance, an isolated disturbance is located at approximately zero frequency on the power spectrum plot. There is no other disturbance occurring at any other frequency, so the power quickly drops off and the remaining values are close to zero.

Figure 4 shows the process with no disturbance impacting the system. Neither the autocorrelation nor the power spectrum contains any obvious peaks. In fact, both trends show random values close to zero. This indicates the control loop is undisturbed and performing well.

The oscillating disturbance depicted in Figure 5 yields an oscillating autocorrelation. The power spectrum shows the oscillating disturbance is a single cycle sine wave since there is one strong dominant peak at the wave's frequency. If a second disturbance was acting on the system a second peak would appear.

The continuously pulsed random disturbance depicted in Figure 6 is difficult to identify since the magnitude of the disturbance is within the range of noise. The disturbance is not impacting the system at a regular frequency because the length of time of the disturbance pulses are not constant. Therefore the power spectrum does not show any significant peaks outside the range of the noise. The autocorrelation gives an indication of disturbance other than noise because there is a strong peak at 25 minutes. Also, there are slight clusters above and below the x-axis, especially close to zero, but they are not as regular as the oscillating disturbance. In this situation comparison to historical data and familiarity with the process is vital.

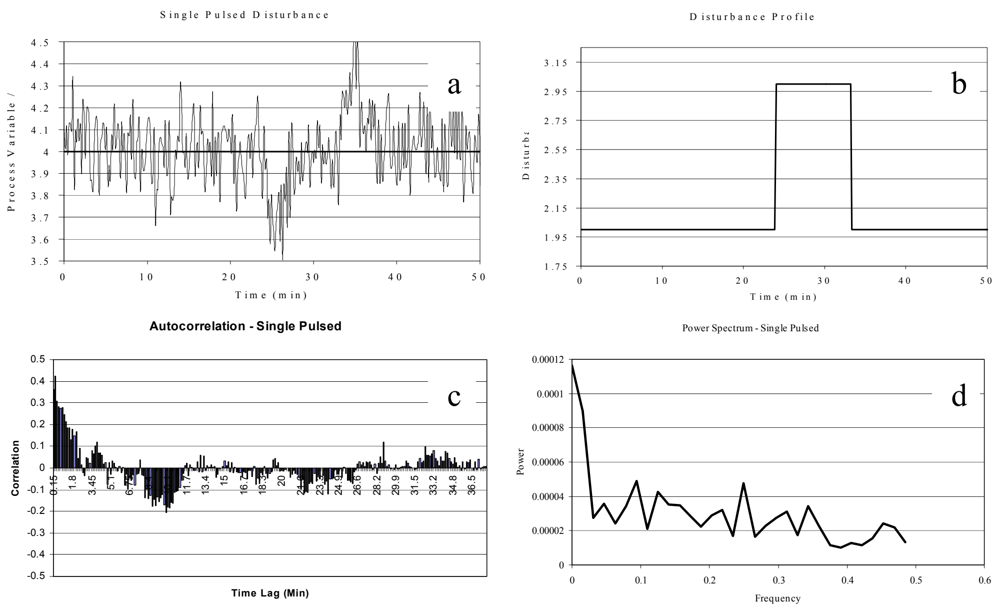

Figure 3 - For a process subjected to a pulsed disturbance here are the a) process variable response b) disturbance profile c) autocorrelation and d) power spectrum plots

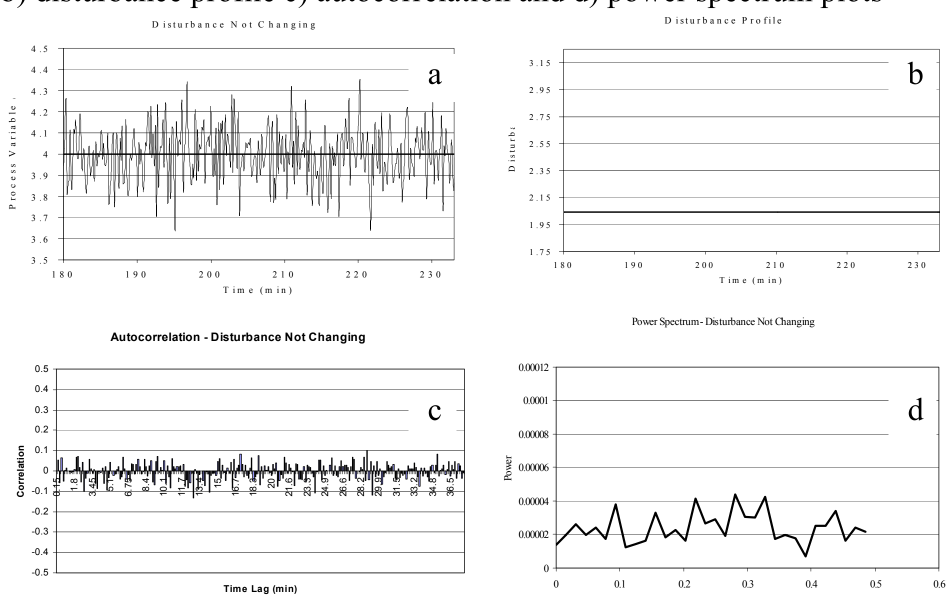

Figure 4 - For an unchanging process here are the a) process variable response b) disturbance profile c) autocorrelation and d) power spectrum plots

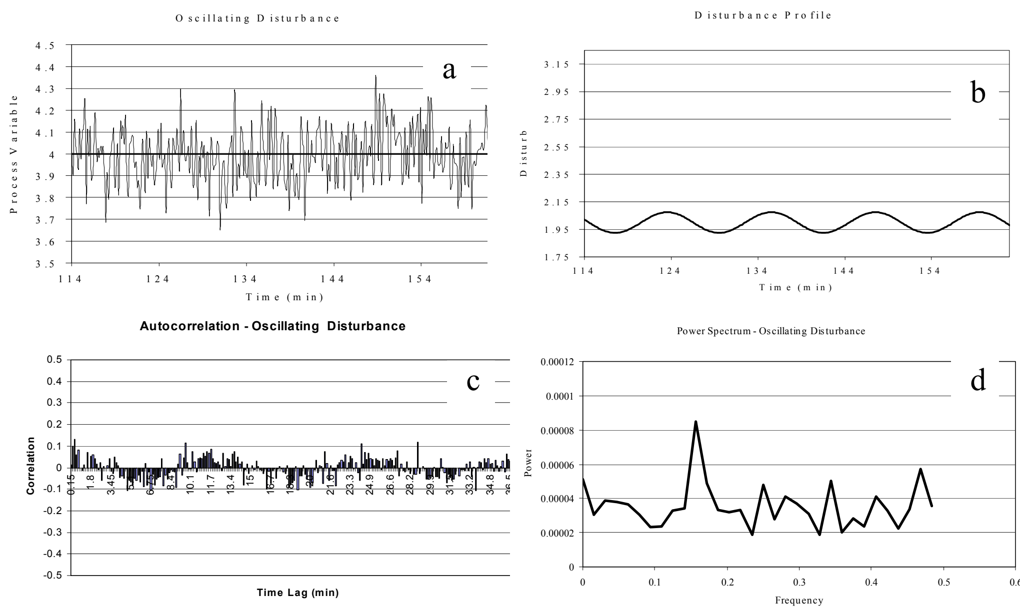

Figure 5 - For a process subjected to an oscillating disturbance here are the a) process variable response b) disturbance profile c) autocorrelation and d) power spectrum plots

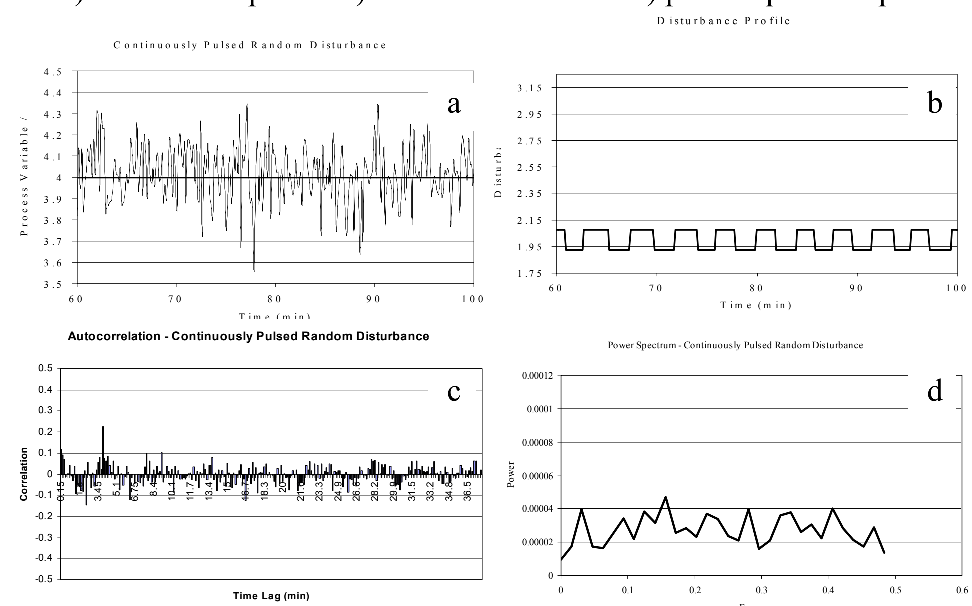

Figure 6 - For a process subjected to a continuously pulsed random disturbance here are the a) process variable response b) disturbance profile c) autocorrelation and d) power spectrum plots

REAL-TIME PERFORMANCE MONITORING

Control loop monitoring is increasingly popular. Many employ the Harris index, based on minimum variance control principles, as the preferred strategy. Presented here is a comparison of the Harris index to simpler strategies for monitoring controller performance. At the heart of every performance monitoring system is the ability to identify a problem within the control loop as soon as possible.

Descriptive statistics are broken into three categories: measures of central tendency, measures of spread, and measures of shape. The mean is the most common measure of central tendency. The measures of spread provide information about the degree to which individual values are clustered or deviate from the mean value in a distribution. The minimum and maximum are the simplest measures of spread and give only the range of values. The variance and standard deviation are other popular measures of spread that provide a more useful numerical value based upon the deviation from the mean. Measures of shape are used to describe the data value distribution; the skewness refers to the degree of asymmetry present in the data set. Each of these descriptive statistics can provide insight into how the control loop is functioning. These statistics are most commonly calculated for the process variable, controller output, and controller error.

Shown in Eq. 6, the Harris Index is a value based on comparing performance under current control to performance if minimum variance control, MVC, were used. In calculating the minimum variance, an autoregressive moving average model is fit to the process data. This is a predictive model that represents action a minimum variance controller would have taken. If the disturbances that affected the process could have been predicted by MVC then the current controller is performing poorly in comparison, but if the disturbances were random then the controller is performing as well as MVC. Since the Harris index is difficult to calculate, it is important to note that under MVC the autocorrelation of the data is zero after the initial process delay; therefore autocorrelation could be used to assess if the system is displaying minimum variance.

The Harris index is computed as [4]:

$$I_H = \frac{\sigma_y^2}{\sigma_{mv}^2} \tag{6}$$

Where: I_H is the Harris index

σ_y^2 is the variance of the process data

σ_{mv}^2 is the minimum variance

When the process displays minimum variance, the Harris index is one. To establish a baseline, the Harris index should be calculated while the system is in peak performance and then used as comparison against future values. The Harris index is useful for assessing the output variance due to stochastic disturbances. It cannot give any specific information about set point changes, known disturbance variables, settling time, decay ratio or stability [10].

The reliability of the Harris index depends on the strength of the model and the estimation of the process dead time. The parameters for the model need to be determined by Box and Jenkins' method [1], prior knowledge, or trial and error. Wrong choice of model or error in estimating the dead time will give misleading values of the Harris index.

An additional performance metric introduced in this paper is standard variation. The standard variation is a normalized measure of deviation of the measured process variable from the set point of a process, and is detailed by Eq. 7.

$$\text{Standard Variation} = \frac{\left(\frac{\sum |PV - SP|}{n-1}\right)}{Average(PV)} \cdot 100\% \tag{7}$$

Where: PV : Measured Process Variable

SP : Set Point

n : Number of Data Points

A smaller standard variation represents less deviation from set point. Some factors that can impact the standard variation include the number of set point changes and the number of disturbances that impact the process. The standard variation can be used to gauge the improvement after a loop is re-tuned, if the standard variation is smaller after retuning, then performance has been improved. It should be noted that when comparing a before and after performance index, the data needs to be collected for a long enough time so that the number of disturbances impacting the system are approximately equal.

Two distinct techniques for computing performance measures include moving and static calculations. Moving calculations are computed on a moving subset of the complete data set. Static calculations would compute the performance measurements on the entire data set. The results of the moving subset calculation are graphed with the performance measure plotted along the vertical axis and time along the horizontal axis. By using a moving subset in lieu of a complete batch calculation, it is possible to identify the point in time when loop performance begins to change. This in turn signals when to begin an investigation into the cause of the change. Because of the ability to identify in real time a changing performance, the moving subset method is preferred for control loop monitoring.

Figure 7 shows the process variable and controller output traces for a time-variant process. A time-variant process is a system whose dynamic behavior changes with time. This change in behavior could be the result of degrading valve performance, heat exchanger surface fouling, catalysts deactivating, or even fluctuating weather conditions. In this example, the system is under PI control and the tuning values are constant during the process transition. By using the moving window technique, the time at which the process began to shift will be clearly identifiable.

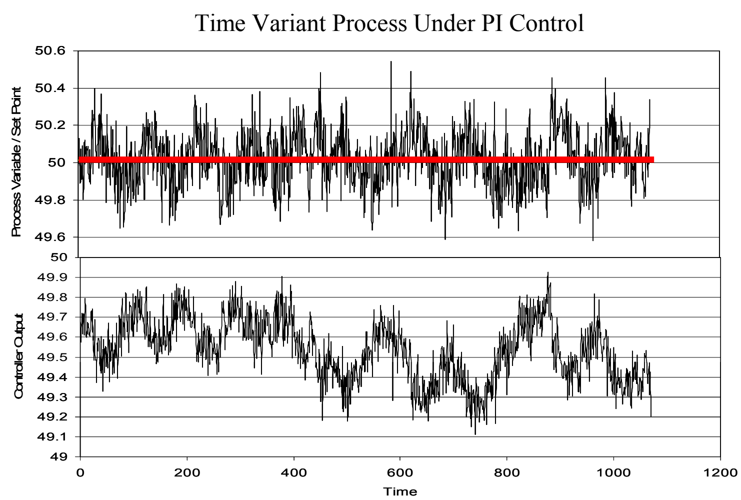

Figure 7 - Process Data and Controller Output

Visual inspection of the process trends shown in Figure 7 does not indicate any significant change in controller performance. The plots that follow based on the methods just discussed reveal that something in the process indeed has changed. By detecting this change before it has a significant impact on controller performance, solutions including updating controller tuning can be considered before alarms are triggered.

Figure 8 shows the Harris index, standard deviation, variance and standard variation of a moving subset of data for the time-variant process. The dotted line in each trace represents the pre-defined baseline value. Since no two processes are alike, each process should have its own baseline or acceptable performance limit determined by measurements collected under normal operating conditions when the system is thought to be running under good control. If the value for any performance criteria moves outside its performance limit for a specified amount of time, then that system has drifted to a warning situation. All four methods show that the process begins to drift from its baseline value at about 550 minutes.

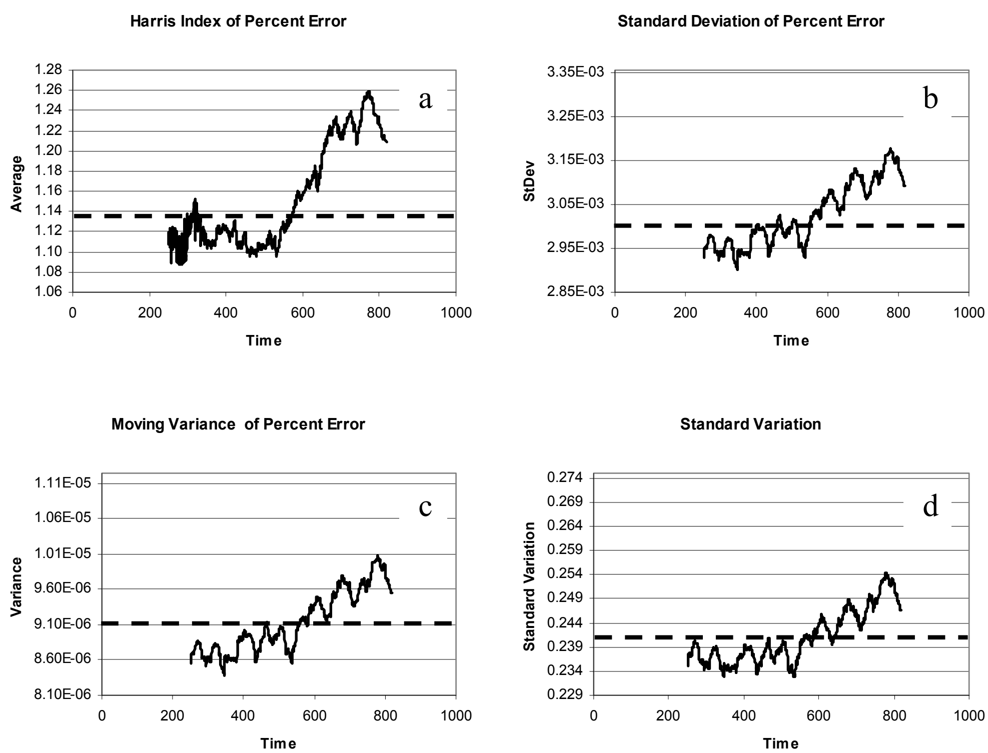

Figure 8 - a) Harris index, b) standard deviation, c) variance, and d) standard variation are calculated by the moving subset method. The time when the process model begins to change is apparent in all plots.

IDENTIFYING INTERACTING PROCESSES

Interacting processes can be troublesome in any manufacturing process. By identifying which systems interact, the disturbances can be counteracted rather than perpetuated throughout the system. Even if an upstream disturbance cannot be eliminated, by identifying the source, a feed-forward controller could be used to improve downstream loop performance.

Cross-correlation analyzes the relationship between two data series. By calculating a set of correlation values at increasing time delays, a picture develops that shows how the data series are related through time. The cross-correlation is calculated as:

$$r(k) = \frac{\sum_i [(y_a(i) - \bar{y}_a)(y_b(i-k) - \bar{y}_b)]}{\sqrt{\sum_i (y_a(i) - \bar{y}_a)^2}\sqrt{\sum_i (y_b(i-k) - \bar{y}_b)^2}} \qquad (8)$$

Where: y_a and y_b are process data
$\bar{y}_a$ and $\bar{y}_b$ are the set point values (or the series averages)
k is time delay in samples
i is sample number (or sample time)

Cross-correlation values are always between negative one and one. Positive values indicate that process A directly affects process B, so that an increased deviation from average in process A causes an increased deviation in B. Negative values indicate an inverse relationship such that an increased deviation in process A causes a decreased deviation in process B. If there is no relationship between the data sets, then the cross-correlation values will be close to zero. Cross-correlation can also be used to determine exactly how much time elapses before the downstream process is reached. At the point when there is greatest impact on the downstream loop, there will be a peak in the cross-correlation trend.

Additionally, cross-correlation is used to identify when disturbances are being caused by a recycle stream. If a recycle stream occurs within a single control loop, an autocorrelation can be used to identify how the recycle influences the system.

Power spectrum is also employed to identify and analyze interacting loops. Interacting loops are affected by the same events and therefore have power spectrum peaks at the same frequencies. Power spectrum cannot identify how long it takes for a change in one system to reach another like cross-correlation can, but it can be more useful when there are many processes separating the suspected interacting loops. Cross-correlation can be muddled when there are many processes in between with varying relationships, but the power spectrum is more sensitive and if processes are affected by events occurring at the same frequencies it will identify the interaction.

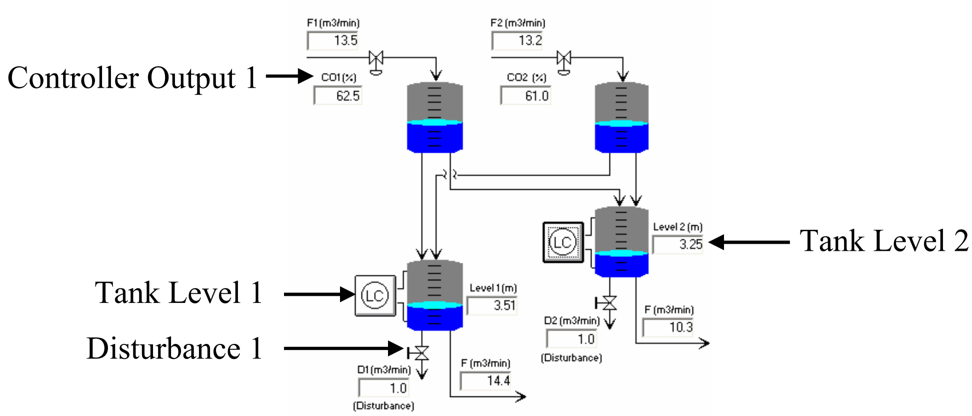

Figure 9 - The interacting tanks process used to demonstrate power spectrum and cross-correlation

To explore the ability of cross-correlation and power spectrum to identify interacting loops, consider the array of tanks shown in Figure 9. The two upper tanks each drain into the two lower tanks. Two controllers connect lower tanks to the upper tanks. If the level controllers on the bottom tanks are put into automatic, a disturbance in one of the lower tanks will affect all four tanks. If the level controllers are left in manual, the tank connectivity is broken and disturbance impact remains local to the particular tank affected.

Consider the system of tanks in manual mode. Figures 10a, 10b, and 10c show the process data when there is a step in the controller output 1 increasing the flow to upper tank 1. Figures 11a and 11b show the cross-correlation of controller output 1 and the levels in lower tanks 1 and 2, respectively. The large peaks on the graphs signify a strong correlation for both and that the maximum effect takes approximately 15 minutes to impact tank level 1 and 30 minutes to impact tank level 2. Figures 12a, 12b, and 12c show process data collected during a step disturbance in lower tank 1. From the autocorrelation plots shown in Figures 13a and 13b, it is clear the disturbance has an almost instantaneous negative effect on the level in tank 1 and no effect on tank 2.

Figure 14a shows the power spectrum of controller output 1, tank level 1, and tank level 2 scaled so they can be displayed on the same graph. All three systems share peaks at the same frequencies and this indicates they are interacting. Figure 14b shows the relationship between a disturbance in tank 1 and the levels in tanks 1 and 2. Disturbance 1 shares similar peaks with tank level 1, indicating they are interacting. Tank level 2 has a unique power spectrum, indicating it is responding to different stimuli.

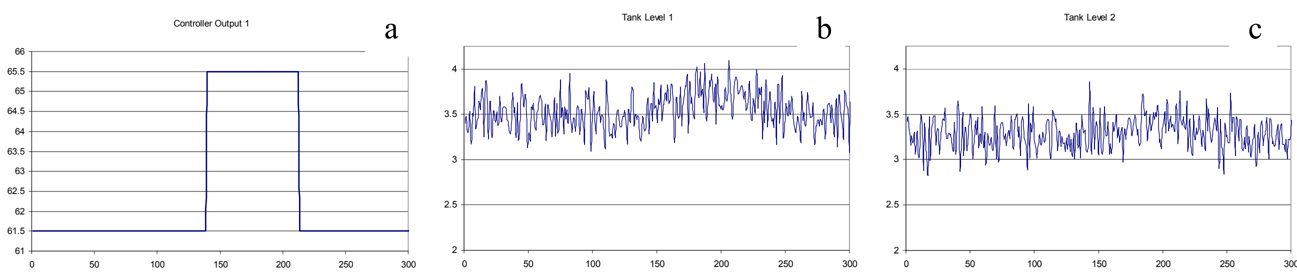

Figure 10 – Process data during a step change of controller output 1 a) controller output 1 b) tank level 1 c) tank level 2

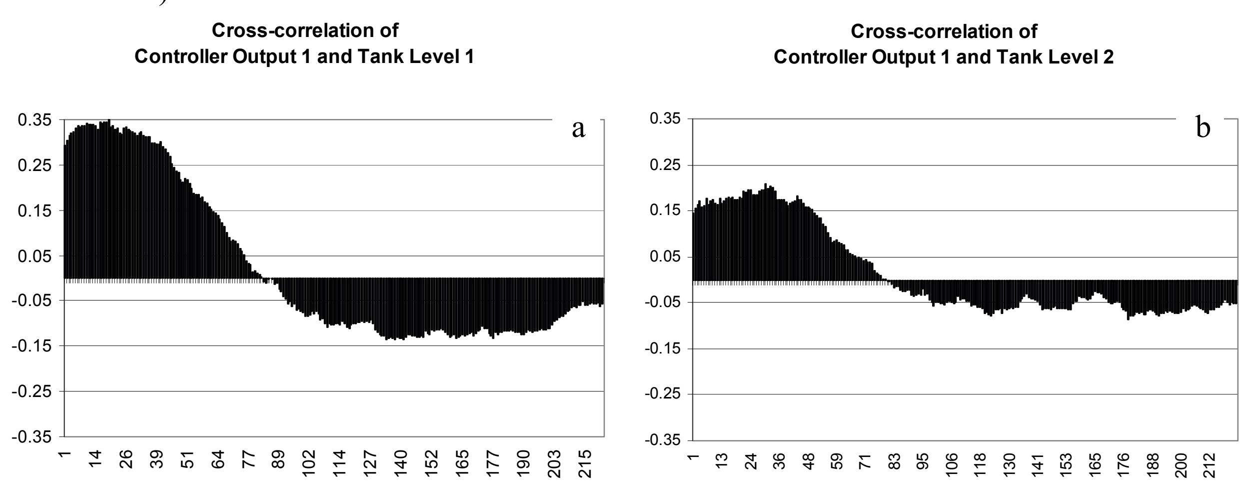

Figure 11 - Cross-correlation diagrams of the relationship between controller output 1 and

Presented at ISA EXPO 2005, 25-27 October 2005
McCormick Place Lakeside Center, Chicago, Illinois, www.isa.org

a) tank level 1 b) tank level 2

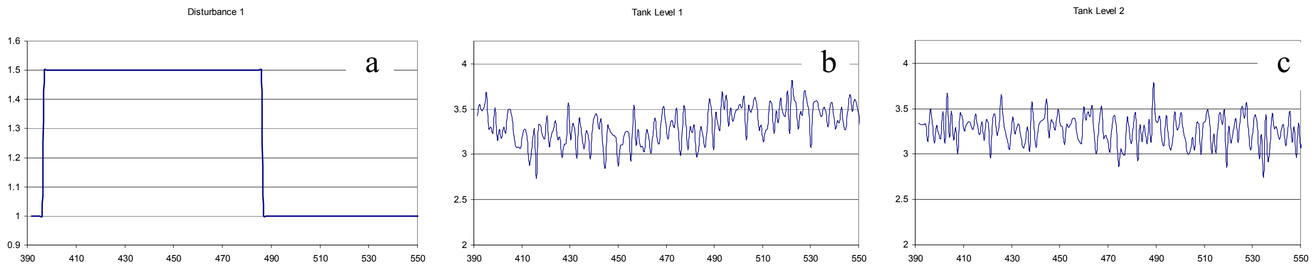

Figure 12 – Process data during a pulse disturbance in tank 1 a) disturbance 1 b) tank level 1 c) tank level 2

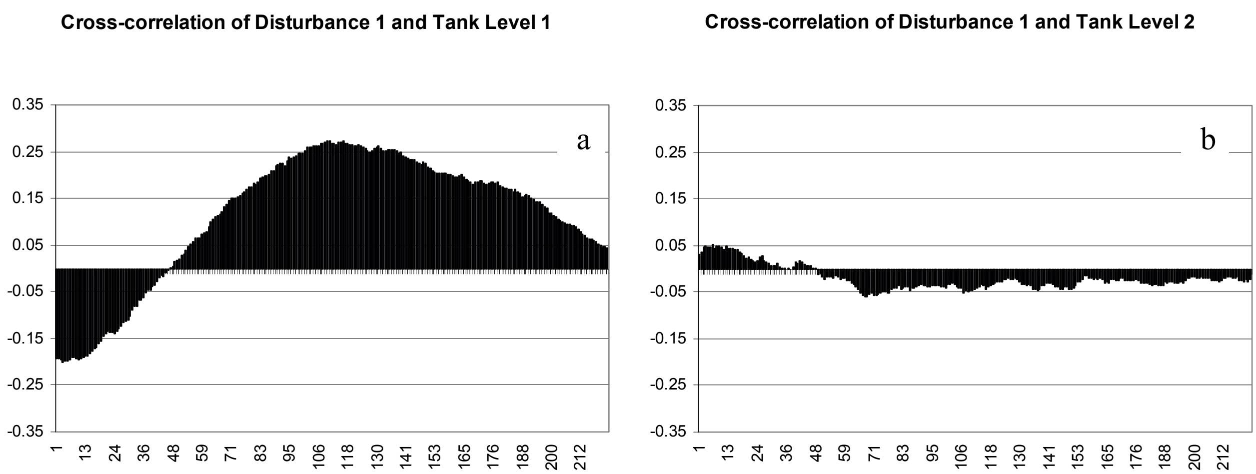

Figure 13 - Cross-correlation diagrams of the relationship between disturbance 1 and a) tank level 1 b) tank level 2

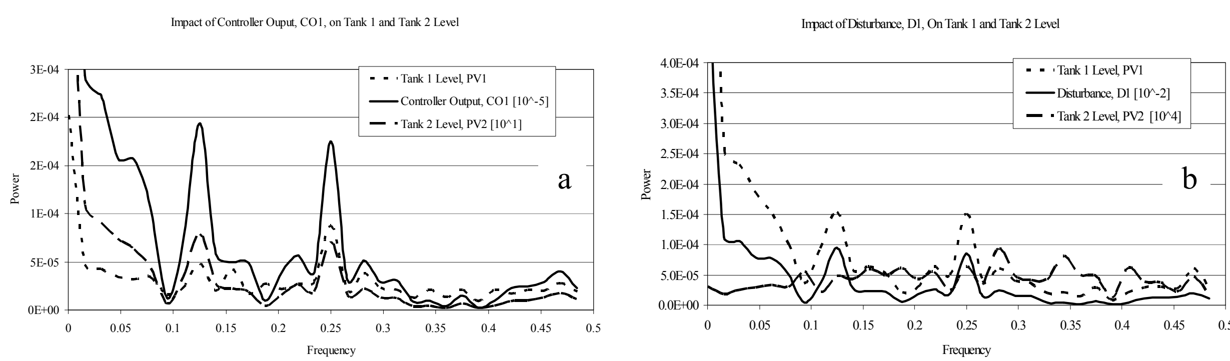

Figure 14 - Power spectrums of a) controller output 1 and tank levels 1 and 2 and b) disturbance 1 and tank levels 1 and 2. The spectrums have been scaled so that they can be view on the same graph.

CONCLUSIONS

Performance measures are an integral part of optimizing and maintaining system performance. Industry and academia are constantly deriving new methods for performance assessment, but the methods are only useful when they can be fully understood and used properly. It is important to understand the theory, purpose and limitations of the measures before relying on their information. In many cases, the performance assessment methods only identify the start of a problem, not the source. By understanding the basic principles and disturbances that impact your system, engineers will know what to expect during normal operation and will be able to identify more quickly what is abnormal operation.

This paper addressed a wide variety of commonly used performance assessment techniques in an attempt to demystify them for better application in monitoring. The techniques detailed in this paper for tackling real-time process monitoring are twofold. First one can identify when a process starts to drift away from baseline operation and towards triggering an alarm. Once a problem is identified, the use of autocorrelation, cross-correlation, and power spectrum can be used for detect the root-cause.

REFERENCES

1. Box G.E.P., Jenkins, G.M., (1970), Time Series Analysis forecasting and control, Holden-Day, Inc. San Francisco, CA

2. Burch, R. (2004). Monitoring and Optimizing PID Loop Performance. ISA Annual Meeting, Houston, TX.

3. Desborough, L. and R. Miller (2001). Increasing Customer Value of Industrial Control Performance Monitoring--Honeywell's Experience. 6th Annual International Chemical Process Control Meeting, Tucson, AZ.

4. Harris, T. J. (1989). "Assessment of Control Loop Performance." Canadian Journal of Chemical Engineering **67**: 856-861.

5. Hoo, K. A., M. J. Piovoso, et al. (2003). "Process and controller performance monitoring: overview with industrial applications." Int. J. Adapt. Control Signal Processing **17**: 635-662.

6. Horch, A. and A. J. Isaksson (1999). "A modified index for control performance assessment." Journal of Process Control **9**: 475-483.

7. Huang, H.-P. and J.-C. Jeng (2002). "Monitoring and Assessment of Control Performance for Single Loop Systems." Ind. Eng. Chem. Res. **41**: 1297-1309.

8. Patwardhan, R. S., S. L. Shah, et al. (2002). "Assessing the Performance of Model Predictive Controllers." The Canadian Journal of Chemical Engineering **80**: 954-966.

9. Press W.H., et al(1986), Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, New York, NY

10. Qin, S. J.(1998), “Control Performance Monitoring – a review and assessment”, Computers and Chemical Engineering, 23, 173-186

11. Thornhill, N. F.(1998), "Performance Assessment and Diagnosis of Refinery Control Loops", AIChe Symposioum Series N° 320, 94, 373-379