Extension of Statistical Process Control (SPC) Methodology to Dynamic Systems Controlled via Output Feedback

Raphael Linker
Faculty of Civil and Environmental Engineering
Technion – Israel Institute of Technology
Israel
linkerr@tx.technion.ac.il

***Abstract*—A model-free statistical method for detecting failures in dynamic systems controlled via output feedback is described. In such systems, the impact of the failure on the output is minimized but the failure causes a change in the control signal, which can be detected. Assuming that historical data of the process with and without failure is available, optimization theory can be used to determine detection thresholds that ensure any desired level of false alarms or detection rate. The proposed method is illustrated on a hypothetical bio-reactor and the results show that the proposed method allows for both rapid detection of the failure and continued operation despite the failure.**

***Keywords*—monitoring; quality control; failure detection; diagnosis.**

I. Introduction

Quality control and failure diagnosis are nowadays integral components of most manufacturing processes. The objective of such procedures is to detect and report any abnormal change in the process as soon as possible while maintaining the number of false alarms sufficiently low. In particular, impressive progresses have been made with regard to model-based diagnosis methods that enable both failure detection (detecting that something is wrong) and failure identification (pinpointing to the most probable failure(s)) (model-based FDI, e.g. [1-3]). However, these methods require relatively accurate analytical models of the process, which for most practical applications are not available. Statistical process control (SPC), which does not require an analytical model of the process, provides the natural answer for such situations. Loosely stated, SPC is based on the comparison of the current output of the process with historical data, assuming that the process and the environment in which it operates are time-invariant or changing very slowly. In other words, historical data is used to determine so-called "control limits" and the new measurements are compared to those limits. Measurements that fall beyond those limits indicate that "something is different", which is typically the case when a failure has occurred in the process or when the operating environment has changed. Such an approach is relevant not only to static manufacturing processes that should produce the same product indefinitely, but also to dynamic processes such as for instance water treatment plants that must produce effluents of acceptable quality.

The most popular SPC methods are based on so-called control charts that present the "process status" in a way that can be easily interpreted by the human operator. One of the main drawbacks of standard SPC methods is that they can not be used "as is" in systems that are controlled via output feedback (also called Engineering Process Control, EPC, in process control and monitoring literature). Clearly, since in such systems the task of the feedback controller is to ensure that the output remains at its desired value, virtually no information can be gained from monitoring this output alone. Due to the obvious advantages of feedback-controlled operation, a number of approaches for monitoring such systems have been suggested ([4-10]). The present study shows that output feedback control and SPC are not mutually exclusive and can be achieved simultaneously after modifying adequately the standard SPC methodology. As in [9], the proposed method is based on the observation that in feedback-controlled systems a failure leads to changes in the control signal, which can be used to monitor the process status.

II. SPC in Open-Loop Dynamic Systems

The basic configuration of SPC is shown in Fig. 1. It must be emphasized that even if the process is continuous, the SPC analysis is based on a sampled signal. As a result, standard SPC tools, such as the $\overline{X}$ graph, can not be used since such graphs were developed for static discrete processes in which the same output (item) can be measured repeatedly. For instance, the $\overline{X}$ graph was developed for data for consist of m samples of n measurements (repetitions) each. In this case, control limits for the sample mean are calculated according to

$$CL = \overline{\overline{X}} \pm 3\frac{\overline{s}}{c_4\sqrt{n}} \tag{1}$$

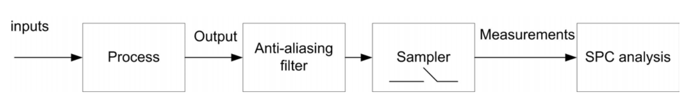

Figure 1. Classical statistical process control (SPC) scheme

where n is the number of measurements in each sample and c_4 is the so-called c_4-factor that depends on n, $\overline{\overline{X}}$ is the grand average of all the previous measurements and $\overline{s}$ is the average standard deviation of the samples

$$\overline{s} = \frac{1}{m}\sum_{i=1}^{m} s_i \tag{2}$$

where s_i denote the standard deviation of the sample i. As noted above, the underlying assumption is that the n repeated measurements are not correlated. Straightforward application of this approach to dynamic systems consists of using consecutive measurements as repeated measurements. Although this is unfortunately done quite often, it must be realized that this is incorrect since such measurements are always correlated, even if the process is driven by white noise of infinite bandwidth. This is due to the fact that this driving white noise is actually filtered by the process itself and by the anti-aliasing filter, so that the sampled signal consists of bandwidth-limited noise ("colored noise"). In particular, since the bandwidth of the anti-aliasing filter is at the most half of the sampling frequency, the auto-correlation of the sampled signal will always be high, regardless of the sampling frequency. Therefore, for sampled dynamic systems, regardless of the mode of control, it is imperative that the control limits be calculated without assuming zero autocorrelation. This can be done for instance using exponentially weighted moving average (EWMA) if the autocorrelation level is approximately known (e.g. [4-5]).

In the present study, a different approach is presented, in which the control limits are calculated based on the probability distribution of the average of each sample ($\overline{X}$) in historical data of the process in the absence and presence of failures. If such data is available, it is possible to determine control limits of guaranteed significance [11]. Denoting the probability density function of the measurement with and without failure as $P_0(x)$ and $P_1(x)$ respectively, the problem of determining the control limits can be formulated as the following optimization problem:

$$\begin{aligned} &\text{Find } u(x_i) \text{ that maximizes } \sum_{1}^{N} P_1(x_i)u(x_i) \\ &\text{Subjected to } \sum_{1}^{N} P_0(x_i)u(x_i) \le A \end{aligned} \tag{4}$$

In (4), $P(x_i)$ denotes the probability of x falling within the x_i-x_{i+1} interval, $u(x_i)$ is the binary decision variable and is equal to 1 when rejecting P_0 (i.e. when rejecting the assumption that there is a failure). According to (4) the probability of correctly accepting P_1 is maximized while the probability of missing a failure is A percent at the most (so-called test significance). Clearly, the role of P_0 and P_1 can be reversed, leading to control limits of known power (known probability of false alarms). This approach has two main advantages over the classical "3σ" approach. First, there is no need to assume normal distribution of the data. Second, the control limits have a clear and well-defined meaning and can be easily adjusted according to the user or process requirements. This approach is illustrated on an example in Section IV.

III. Extension of SPC to Closed-Loop Dynamic Systems

A typical single input-single output system controlled via output feedback is shown in Fig. 2. As mentioned above since in closed-loop systems the task of the feedback controller is to ensure that the output remains at its desired value, very little information relative to the occurrence of failures can be gained from monitoring this output. Indeed, only in cases when the controller is unable to perform its task does the output differ significantly from its desired value. However, the presence of a failure will be reflected in the control signal (the required control effort). For processes that are meant to operate at a time-invariant working point, SPC can be performed on this signal: In the absence of failures, the control signal will vary slightly due to external disturbances and/or changes in the process parameters, while failures will cause larger change of this signal. The example detailed in the next Section shows how the guaranteed power approach detailed above can be used to detect failures via the control signal.

IV. Illustrative Example

We consider a hypothetic biological reactor driven by the following non-linear differential equations

$$\begin{aligned} \frac{dB}{dt} &= \frac{Q}{V}B_{in} + \left[\alpha C - \beta(\theta - T) - \frac{Q}{V}\right]\cdot B \\ \frac{dC}{dt} &= \frac{Q}{V}C_{in} - \left(\gamma B + \frac{Q}{V}\right)\cdot C \\ \frac{dT}{dt} &= \frac{1}{C_v V}(H - \phi B) \end{aligned} \tag{5}$$

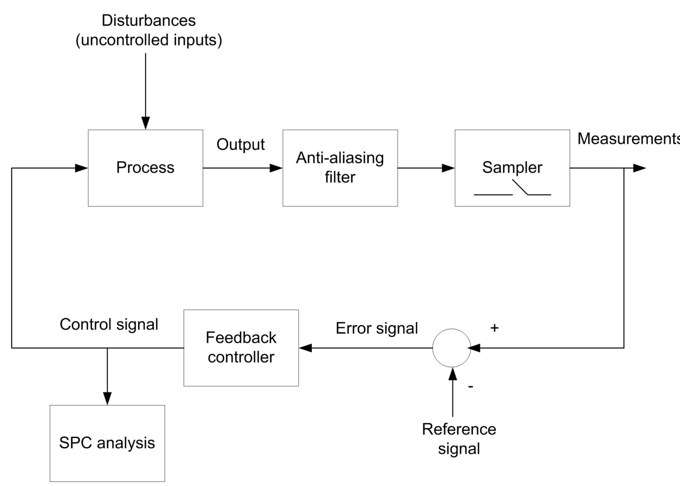

Figure 2. Statistical process control scheme for dynamic systems controlled via output feedback

TABLE I. NOMINAL PARAMETER VALUES

Parameter, units	Nominal value
α, $m^3/(gh)$	1.3
β, $1/(^oCs)$	1E-6
γ, $m^3/(gh)$	0.1
θ, ^oC	120
ϕ, Wm^3/g	70
C_v, $J/(^oCm^3)$	1000
Q, m^3/h	6E-4
V, m^3	1
H (open loop), W	10

The first and second equations describe the rate of change in bacterial and contaminant concentrations (B [g/m^3] and C [g/m^3], respectively) that depend on the inflow concentrations (B_{in} and C_{in}), the flow rate Q [m^3/h] and the temperature T [^oC]. The third equation describes the rate of change of the reactor temperature, which depends on the supplied heat flux H [W] and heat consumption (sink) by the bacteria. The reactor volume (V [m^3]) and flow rate are assumed to remain constant. It must be emphasized that these equations are not expected to describe a real process but merely serve as an illustrative example for the proposed approach.

It is assumed that B_{in}=0 and that during normal operation (no failure), the contaminant inflow concentration remains within the 9.5E-3 – 10.5E-3 range, and we focus on detecting small increase or decrease of the contaminant inflow concentration beyond the normal range: 8.5E-3<C_{in} < 9.5E-3 or 10.5E-3<C_{in} <11.5E-3. For this purpose, we assume that only the temperature T can be measured, since real-time measurements of B and C would usually not be possible in real reactors. It can be easily verified that with the nominal parameter values listed in Table I the system is stable during failure-free operation and the output concentration C remains around 4E-4 g/m^3. Changes in the contaminant inflow concentration C_{in} are reflected by changes of the temperature T, so that T can indeed be used for detecting such changes. Clearly, such an approach would enable only failure detection and not failure identification. However, failure identification is not considered by classical SPC and is beyond the scope of the present study.

The simulation scheme is presented in Fig. 3. Random noise corresponding to ±10% of the parameter's nominal value was added to the nominal value of all the parameters except V and C_v. The system's outputs (B, C and T) were passed through a third-order Butterworth anti-aliasing filter with a 4.44E-7 rad/s (6.94E-8 Hz) bandwidth and sampled at a rate of 1000 measurements/hours. The system was simulated both in open-loop (Gain K=0) and in closed-loop with a simple proportional integral (PI) controller tuned according the Ziegler-Nichols rule. The system was initially close to steady-state and was simulated for 5E6 hours with C_{in} within the no-failure range. At t=5E6 hours a step was introduced in C_{in} to bring it into one of the failure ranges. Altogether, 150 simulations were performed. One hundred of these simulations were used to determine the

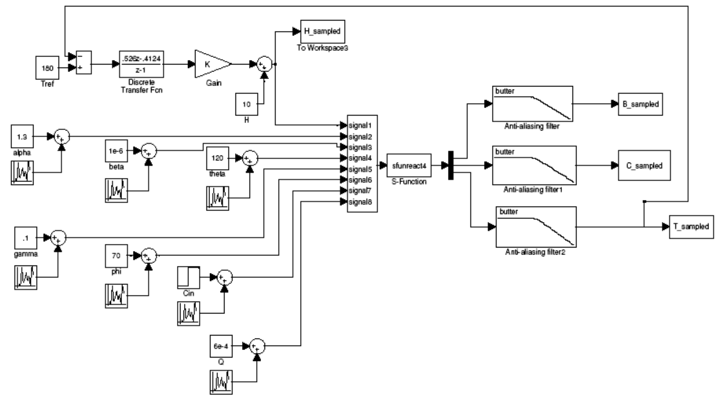

Figure 3. Simulation scheme

control limits (calibration stage) and the remaining simulations were used to test the procedure. To prevent artifacts due to the initial conditions at start-up, in each simulation only the measurements performed after t=1E5 seconds were used in the analysis.

A. Open-loop SPC

Fig. 4 shows the results of 10 typical simulations. The step change introduced in C_{in} at t=5E6 hours causes a rapid deviation of C from its normal value and is also reflected by a strong change of the temperature T.

Fig. 5 shows the probability density functions of the temperature T with and without failure (calibration data only). Based on these historical measurements, the control limits that ensured 0.1 % false alarms were calculated following the optimization procedure outlined in Section 2 (bottom frame, also shown in bottom frame of Fig 4). It must be noticed that the distribution of the no-failure temperature is not normal (top frame), so that a classical "3σ" limits would not be appropriate. Applying these limits to the validation simulations led to the results shown in Fig. 6. The top frame shows the average run length (ARL - the average time between false alarms) for each simulation and it can be seen that only one false alarm occurred during the validation runs. The failure was correctly detected in all 50 cases, with detection times ranging from 1.5E4 hours (3 "samples" of 5 consecutive measurements) to 17E4 hours.

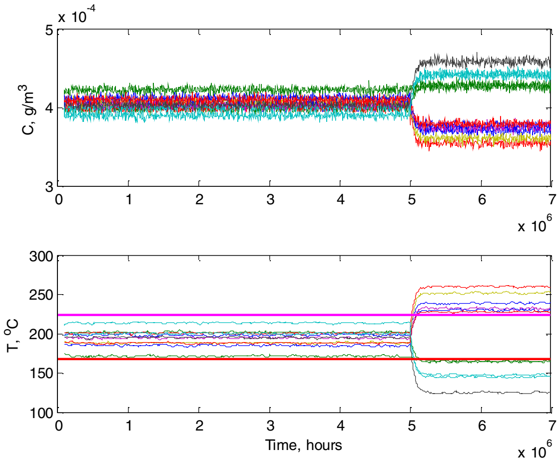

Figure 4. Typical results of open-loop simulations. A step change in C_{in} was introduced at t=5E6 hours. The top frame shows the concentration C and the bottom frame shows the temperature T.

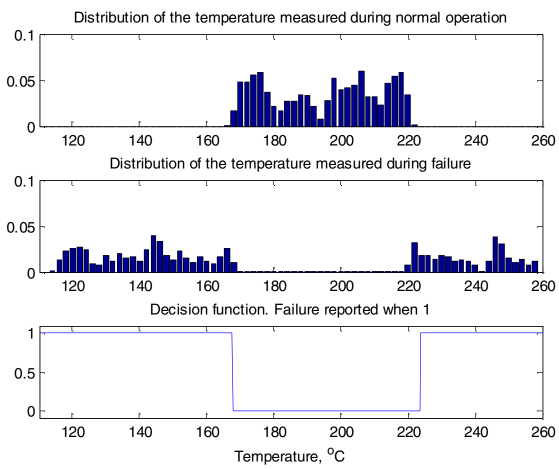

Figure 5. Distribution of the temperature measured during normal operation (top frame) and during failure (middle frame), and control limits (bottom frame).

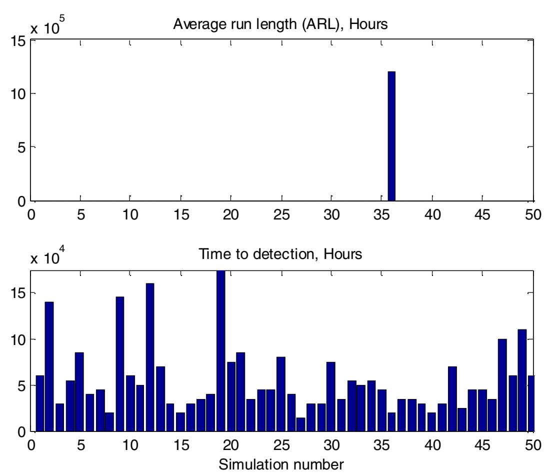

Figure 6. Average time between false alarms (top frame) and time to failure detection (bottom frame). When ARL is not shown, there was no false alarm during the respective simulation run.

B. *Closed-loop SPC*

Fig. 7 shows the signals recorded during a typical closed-loop simulation. It can be verified that the feedback controller is able to maintain the temperature at its desired level (180°C, chosen based on normal open-loop operation). As a result the reactor concentration C is not affected by the failure in C_{in}. It can also be seen that the failure causes a change in the control signal H. The probability density functions of the control signal H in the absence and presence of the failure are shown in Fig. 8 (calibration data only), and it can be seen that, as was observed in the open-loop case, the distribution of the failure-free variable is not normal. The control limits that ensure 0.1% false alarms are shown in the bottom frames of Fig. 7 and 8, and applying these limits to the validation simulations yielded the results shown in Fig. 9. False alarms were reported in only 4 cases, with ARL beyond 4E5 hours. The failure was detected in all 50 simulations, with the vast majority of detections occurring within 1E4 hours (2 samples) of the failures.

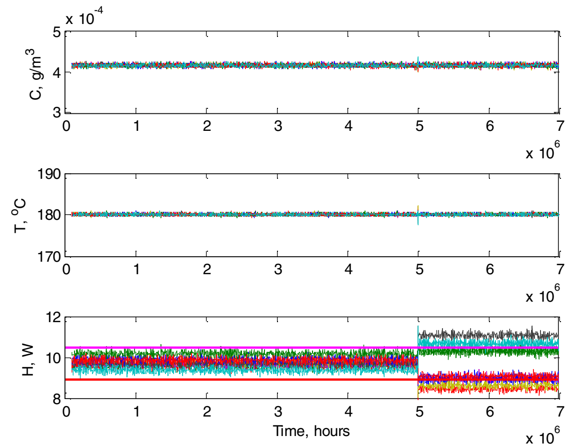

Figure 7. Typical results of closed-loop simulations. A step change in C_{in} was introduced at t=5E6 hours. The top frame shows the concentration C, the middle frame shows the temperature T and the bottom frame shows the control signal H.

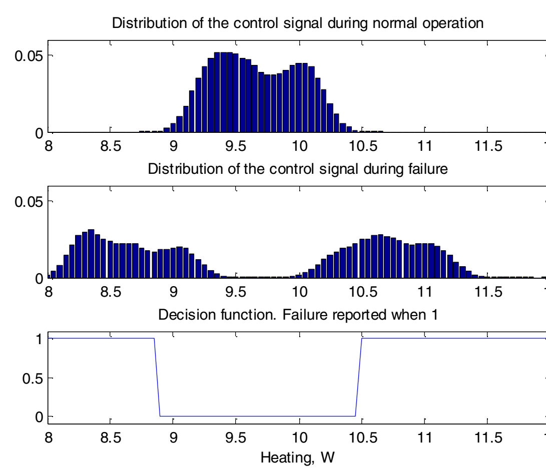

Figure 8. Distribution of the control signal applied during normal operation (top frame) and during failure (middle frame), and control limits (bottom frame).

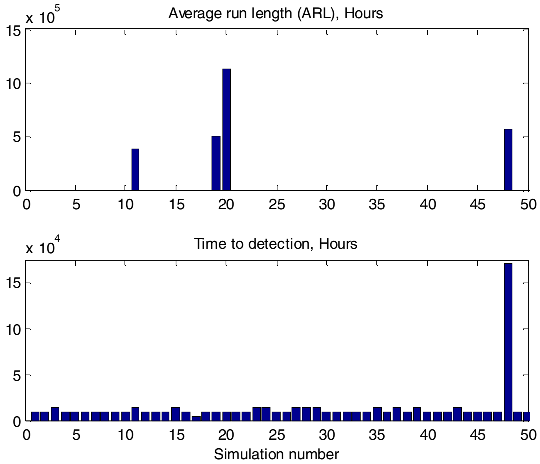

Figure 9. Average time between false alarms (top frame) and time to failure detection (bottom frame). When ARL is not shown, there was no false alarm during the respective simulation run.

V. DISCUSSION – ADVANTAGE OF THE CLOSED-LOOP OPERATION

The results presented in the previous Section clearly show that single input-single output dynamic systems that are controlled by output feedback can be monitored in a statistical way in a manner that is very similar to open-loop systems. The main advantage of the closed –loop approach is that the system output remains virtually unaffected by the failure. In other words, the failure is detected and the system continues to operate and produce an acceptable product despite the failure. Clearly, this is at the expense of an increased (or decreased) control effort and hence a higher production cost. However, in most practical applications, continued production at a higher cost would be highly preferable over system shut-down until repair.

The main advantage of the proposed method is that no assumption is made with respect to the signal auto-correlation or the normality of the measurements. Also, although the method was illustrated with an input failure, it could be applied similarly for detecting any other type of failure (actuator, sensor, or change in the process itself).

REFERENCES

[1] J. Chen, and R. J. Patton, Robust model-based fault diagnosis for dynamic systems, Kluwer Academic Press, Boston, MA, 1999.

[2] P. M. Frank, Analytical and qualitative model-based fault diagnosis-A survey and some new results, European Journal of Control, vol. 2, pp. 6-28, 1996.

[3] J. Gertler, Residual generation in model-based fault diagnosis, Control-Theory and Advanced Technology, vol. 9, pp. 259-285, 1993.

[4] D C. Montgomery, J. B. Keats, G. C. Runger, and W. S. Messina "Integrating statistical process control and engineering process control", Journal of Quality Technology, vol 26, pp. 79-87, 1994.

[5] D. Q. Cai, M. Xie, and T. N. Goh, "SPC in an automated manufacturing environment", International Journal of Computer Integrated Manufacturing, vol 14, pp. 206-211, 2001.

[6] F. Tsung, and D. W. Apley, "The dynamic T^2 chart for monitoring feedback-controlled processes", IIE Transactions, vol 34, pp. 1043-1053, 2002.

[7] D. Shi, and F. Tsung "Modelling and diagnosis of feedback-controlled processes using dynamic PCA and neural networks", International Journal of Production Research, vol 41, pp. 365-379, 2003.

[8] G. Runger, M. C. Testik, and F. Tsung, "Relationships among control charts used with feedback control", Quality and Reliability Engineering International, vol 22, pp. 877-887, 2006.

[9] F. W. Faltin, and W. T. Tucker ,"On-line quality control for the factory of the 1990s and beyond" in Statistical Process Control in Manufacturing, J. B. Keats and D. C. Montomery, Eds. Marcel-Dekker, NY, 1991, pp. 331-354.

[10] W. S. Messina, D. C. Montgomery, J. B. Keats, and G. C. Runger, "Strategies for statistical monitoring of integral control for the continuous process industries" in Statistical Applications in Process Control, J. B. Keats and D. C. Montgomey, Eds. Marcel-Dekker, NY, 1996, pp. 193-215.

[11] E. L. Lehman, Testing Statistical Hypotheses, Wiley & Sons, NY, 1986.