Management of Oil/Gas Pipelines using Statistical Process Control

William V. Harper[1], David J. Stucki[1], Tony Alfano[1]
Thomas Yahner[1], Jim Ponder[2], Jeffrey Kobs[2]
[1]DNV GL USA, Inc., 5777 Frantz Road, Dublin, OH 43017
[2]Rose Rock Midstream, 6120 S. Yale Ave, Suite 1500, Tulsa, OK 74136

Abstract
One of the most critical aspects of oil/gas pipeline operation is the ability to assess portions of the lines that are higher risk. Doing this on a continual and timely basis is paramount. The challenges are many. Recent media discussion has focused on evolving data analytics; however, the application of long established statistical process control techniques provides a solid defensible visual identification of areas of potential concern.

Key Words: Statistical Process Control, SPC, Oil, Gas, Pipeline

1. Introduction

Risk assessment of a company's pipelines is commonly done on dynamic segments. A change in any of the chosen segmentation variables (e.g., pipe coating type; in-line inspection tool anomaly call, high consequence area) causes a new dynamic segment to be created for modeling purposes. The intent is to have portions of the line that should respond uniformly to pipeline threats and to be in the same consequence class in the case of product release. It is not unusual to end up with many thousands of such dynamic segments varying in length. The issue is how to easily and accurately identify the segments of potential concern.

The primary objective of a pipeline risk assessment program is to identify and manage the portions of the lines that are higher risk. Trusted and efficient tools are needed to meet this objective. One of these tools is Statistical Process Control (SPC).

SPC has been a mainstream tool in the automotive industry and general manufacturing. The visual tools are simple to apply resulting from work by Walter A. Shewhart in the 1920s and published in the book *Economic control of quality of manufactured product* (Shewhart, 1931 [1]). The role SPC, as applied to pipelines, is to identify those dynamic segments that have higher predicted threats resulting from higher probabilities of failure, release of product consequence, or the resultant combination risk.

2. The SPC Methodology

While fluctuations occur both temporally and spatially, it is important to identify the real signals (e.g., significant high-risk areas) from the inherent variability exhibited by the majority of the data. Separation of common typical variability from those unusual rarer events is the result of SPC. Here events are based on predicted frequencies of failure and resultant consequences for dynamic segments. For the current pipeline application, SPC is used to only identify segments on the high end (high expected failures or consequences). Such segments have a higher predicted frequency of failure (or expected loss/risk that combines frequency and cost) relative to the line in general. This can be done on a total system basis, on a specific line, across all failure modes, or focused on a specific failure mode such as external corrosion of the pipeline.

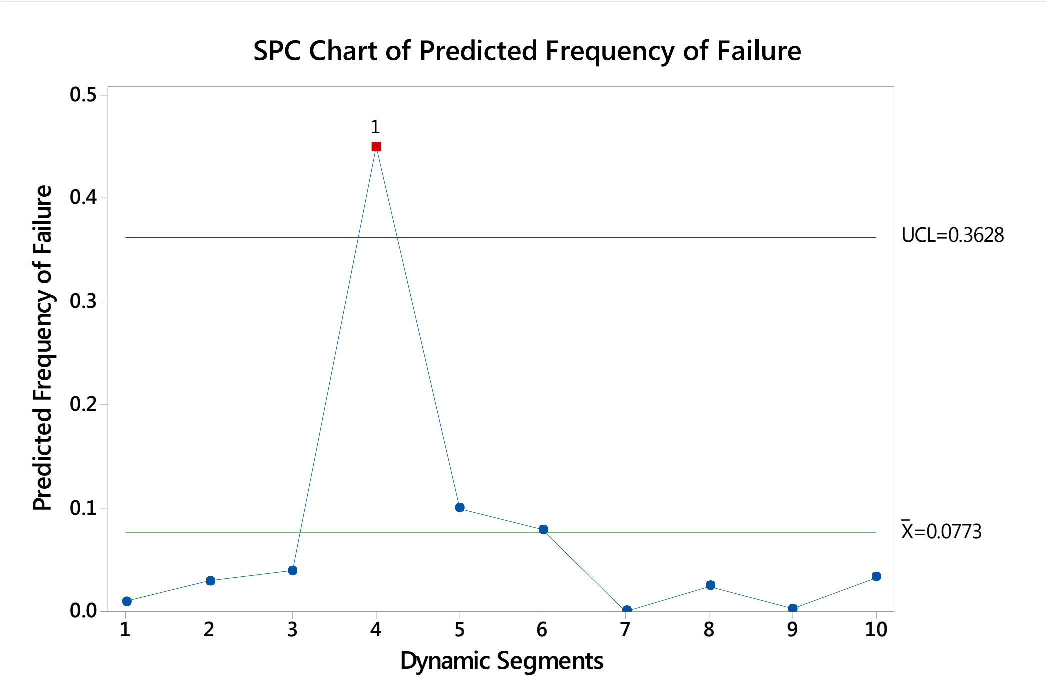

Figure 1: Example of SPC chart to identify dynamic segments of potential concern

A simple example is depicted in Figure 1 with hypothetical frequency of failure data for ten dynamic segments. Figure 1 provides the average $\overline{X}$ frequency of failure for the hypothetical dynamic segments as well as the upper control limit (UCL). The mathematical steps to compute the average line and the upper control limit (often called the three-sigma limit) are given below. In Figure 1 the 4th dynamic segment would be classified as a signal as mentioned in the prior paragraph having a high frequency of predicted failure compared to other dynamic segments.

Some background on SPC will help put things in perspective. The most common SPC charts taught are based on taking around 4-5 samples at some specified time interval such as every 20 minutes from a production line. Each subgroup is measured for the quantity of interest (e.g., the amount of time to complete a transaction) and the range and average of the resulting subgroup values are computed and plotted over time on $\overline{X}$ and Range (R) charts. Various criteria are applied in the steps to establish average, lower control limit (LCL), and upper control limit (UCL) for both charts. The control limits are typically plus and minus three standard deviations. These plots provide a way to separate common (typical) cause variation from special or assignable (unusual) cause variation. In most SPC analyses both the lower and upper limits are important whereas in pipeline risk management only the upper end is of importance.

For the application to pipelines, SPC is applied at the system level (e.g., all lines) as well as separately for individual pipelines. This allows the ability to isolate the impact of a common factor such as product type on a given line. It also provides a mechanism to examine between line characteristics such as a high flow versus low flow. In doing so it is possible to find the special cause variation high risk segments relative to other segments on that line.

The SPC approach used is called I-MR charts or Individuals and Moving Range charts. Some call the Individuals chart an X chart where the X here represents a single value (one dynamic segment's value for a particular metric); therefore I-MR is sometimes labelled XmR (for X and moving Range). Going back to the most commonly taught $\overline{\mathrm{X}}$ and R charts, one purpose of the R (range) chart is to estimate the within subgroup variability – such as the standard deviation. Estimation of standard deviations from a small number of samples (around 4-5) has stability problems and the use of the average range is common practice to quantify the within sample (within subgroup) standard deviation. For those used to analysis of variance (ANOVA) terminology, a within group variability is used to establish the bounds for the means or $\overline{\mathrm{X}}$ chart. When the number of samples per group is just 1 (pipeline case in which each dynamic segment stands on its own), it is not possible to compute a standard deviation. To address this aspect, a moving range using the adjacent values is used as explained below to quantify variability. Note that the moving aspect is only applied to the range. The estimated loss or frequency of failure is not averaged that represents the risk related value for each dynamic segment. Each dynamic segment stands on its own when plotted as frequency of failure or expected loss as an X value.

The formulas that follow provide the basic aspects of both the Individuals and Moving Range charts (Breyfogle, 2003 [2]). They do not include the control limits for the Moving Range chart as these are not directly used in pipeline risk assessment but they can be found in Wheeler and Chambers (1992 [3]). There are n dynamic segments where n varies based on the given pipeline examined. Every dynamic segment has multiple threat frequencies and consequences for which SPC is applied individually on each. Each moving range $\mathrm{MR_i}$ (i = 2, n) as shown in equation 1 is the difference between adjacent dynamic segments as $\mathrm{x_i} - \mathrm{x_{i-1}}$ where $\mathrm{x_i}, \mathrm{x_{i-1}}$ are the threat frequency or expected loss for the given dynamic segment "i" and the prior dynamic segment "i-1". Thus, x_i is used for two purposes – once for threat frequency and once for expected loss. Expected loss is multiplication of a given threat frequency and the corresponding consequence cost of an incidence (leak, rupture) from that threat on a particular dynamic segment. Since there may be multiple releases from a given line the expected cost must factor in the estimated number of occurrences or frequency. When x_i is used in equations 1-4 as an expected loss this multiplication has already been computed. From the notation, there are n-1 moving ranges since there is no $\mathrm{MR_1}$ as there is no 0th dynamic segment. $\overline{\mathrm{MR}}$ in equation 2 is the average of the moving ranges ($\mathrm{MR_i}$). $\bar{\mathrm{x}}$ in equation 3 is the average of all n of the dynamic segment $\mathrm{x_i}$. Since it is not an overall average of averages as would be the case for a typical $\overline{\mathrm{X}}$ plot in which each subgroup has multiple observations, it is common to use a lower case $\bar{\mathrm{x}}$ for the individuals mean.

Moving Range Equations:

1) $\mathrm{MR_i} = |\mathrm{x_i} - \mathrm{x_{i-1}}|,\ \mathrm{i} = 2, \mathrm{n}$

2) $\overline{\mathrm{MR}} = \frac{1}{\mathrm{n}-1} \sum_{\mathrm{i}=2}^{\mathrm{n}} \mathrm{MR_i}$

The lower and upper three sigma individual chart control limits (LCL, UCL) are given in equation 4; similar two sigma control limits replace the 3 in the equation with 2. The parameter d_2 (Duncan, 1965 [4]) in equation 4 is standard in SPC and has a value of 1.128 (for subgroups of size 2) and is used to convert a range to an estimate of the standard deviation as a function of how many observations are used to compute the range (2 for this moving range). Typical SPC applications use 3 standard deviations; however, some add 2 standard deviations for an early warning to address problems before they become more severe. In pipeline risk assessment, the focus from a SPC perspective is on those segments that exceed the upper control chart three sigma limit.

Individual SPC Equations:

3) $\bar{x} = \frac{1}{n}\sum_1^n x_i$

4) $(\mathrm{LCL}, \mathrm{UCL}) = \bar{x} \pm 3\,\frac{\overline{\mathrm{MR}}}{d_2}$

The moving ranges are used to estimate the variability and are converted to an estimate of the standard deviation, i.e., $\frac{\overline{\mathrm{MR}}}{d_2}$ is an estimate of the standard deviation. There are no moving averages for frequency of failure or expected loss. Thus, each dynamic segment is not averaged with an adjacent segment and stands on its own relative to its frequency of failure or expected loss.

Figure 2 is a sample output from a risk assessment. In this application, a Moving Range chart is not created though the equations above are computed as they are needed in the Individuals (or x) chart. Two sets of means (averages) and control limits are shown. There are two upper control limits for each – in addition to the three-standard deviation limit there is a two-standard deviation limit. One set of results is based on the complete system (labeled ‘Global Expected Loss’ in the legend) or network. The other is based on just the given line (labeled ‘Pipeline Expected Loss’ in the legend) under assessment. The three-standard deviation upper control limit has +3Sigma as part of the name in the legend. A two-standard deviation upper control limit (+2Sigma) is used to help differentiate dynamic segments that represent a second tier of segments for maintenance management planning. The segments falling above a given control limit have either higher relative frequencies of failure or higher relative expected cost.

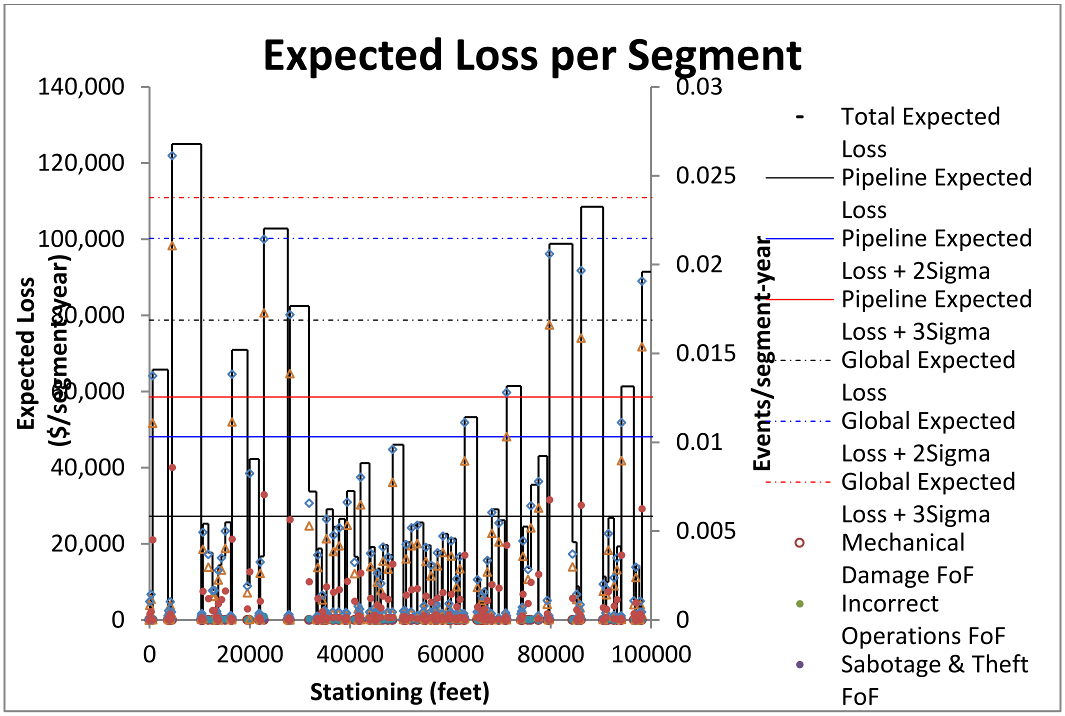

FIGURE 2: Sample expected loss output – likely rename as well as replace above

Providing both global and local (specific line) statistics helps management and engineers assess priorities. Figure 2 shows the engineer that many dynamic segments exceed the 3 Sigma threshold (solid red line "Pipeline Expected Loss + 3Sigma") for the given pipeline, but only one dynamic segment from this particular line is considered a high relative expected loss for the full pipeline system.

This SPC based system has proven to be extremely effective in the identification of segments that have significantly higher expected frequencies of failure as well as higher relative expected loss. This allows management to effectively plan maintenance budgets and prepare for audits. Additionally, the full system summarizes the frequencies of failure predictions along with the costs to comprehensively describe each line and the total system in addition to the use of SPC to highlight the areas of most immediate concern line by line and across the total system.

3. Myths about Statistical Process Control

Various myths about SPC have been addressed (Balestracci 2011 [5]; Wheeler and Chambers, 1992 [3]; Wheeler, 2010 [6]) with the major ones listed below. In this paper, only the first is examined in detail. The reader is referred to the references for in-depth explanations on the subsequent myths and their debunking.

1. Data must be normally distributed.
2. SPC works because of the central limit theorem.
3. Data must be "in control" (within $\pm 3\sigma$) before implementing.
4. Three standard deviation limits are too conservative.
5. Data must be independent, i.e., auto-correlated data will not work.

The top myth that the data must follow the normal distribution is the most common of the five above myths and the key one to be dispelled. This first myth ties closely with the second myth concerning the central limit theorem in which the sampling distribution of averages of data are well known to generally follow a normal distribution. Figure 3 is a normal distribution with +/- 1, 2, and 3 standard deviations marked and is discussed in more detail below.

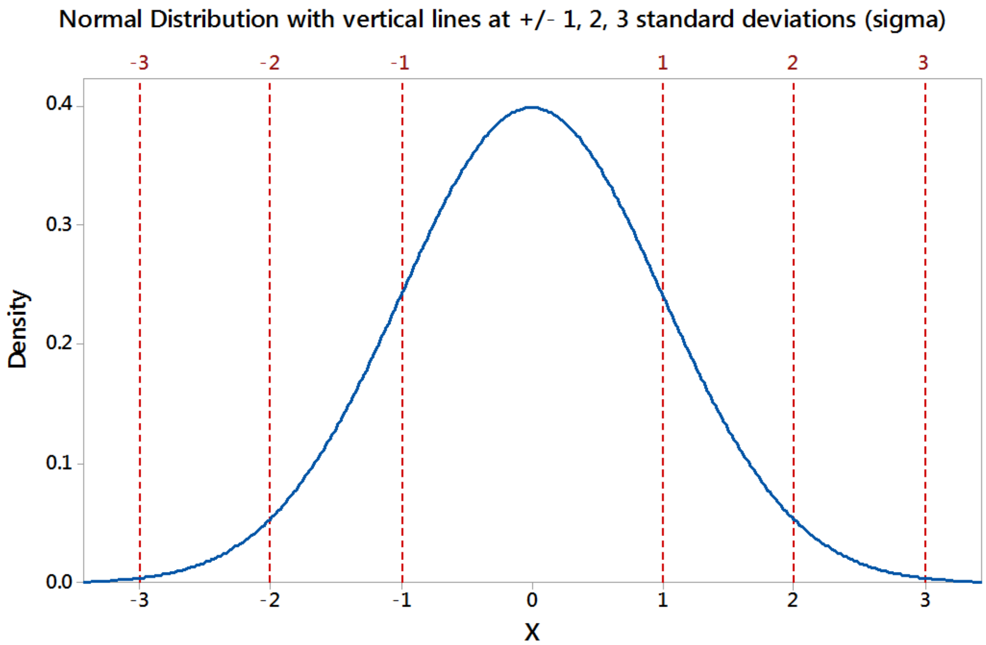

Figure 3: Plot of normal distribution with markings at +/- 1, 2, and 3 standard deviations

It can by shown that the SPC methodology can be applied when data is not normally distributed by comparing the normal distribution percentages based on the standard deviation vertical lines in Figure 3 to a wide range of distributions. An assessment of how the proportions in the ranges about the mean μ of $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ (σ is the standard deviation or sigma) differ across the distributions is given in Table 1 (based on Figures 4-4 through 4-6 in Wheeler and Chambers [3].)

A considerable variation for the six distributions being within $\pm 1\sigma$ of the mean is seen in Table 1. However, the percentages are close to the normal distribution probabilities for both $\pm 2\sigma$ and $\pm 3\sigma$ standard deviations of the mean. This is especially so for $\pm 3\sigma$ that is the basic foundation of most statistical process control charts. The risk based model focus is primarily on the $+3\sigma$ line above the mean with a secondary early warning screening available at $+2\sigma$ above the mean.

Table 1: Percentages within a number of standard deviations about the mean

Distribution	$\pm 1\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
Uniform	57.7	100	100
Right Triangular	62.9	96.2	100
Normal	**68.3**	**95.5**	**99.7**
Burr	72.6	95.2	99.1
Chi-Square	73.8	95.3	98.6
Exponential	86.5	95.0	98.2

Since the constants used in SPC are indeed based on the normal distribution (d_2 as shown in the earlier I-MR formulas), many have felt the data must be normal to apply SPC. The above should address this concern. In essence practical implications on non-normality are minimal. In addition to the references at the start of this section, Burr (1976 [7]) examined this in detail for various statistical distributions. In Shewhart's original (1931 [1]) book, he examined uniform distributions and right triangle distributions (both highly non-normal) as well as Chebyshev's inequality. Thus, from a practical view point, this issue should no longer be a concern; however, many text books still mislead practitioners. The reader is urged to investigate the remaining SPC myths with the references provided.

4. Marriage of SPC and Financial Return on Investments

The identification of high risk dynamic segments is a key step in integrity management; however, the operation of the pipeline system must have sufficient financial return. This involves assessment of the economic value of the pipeline risk mitigation alternatives including accounting for the time value of money. Cash flows (positive and negative) using the organization minimum attractive rate of return aid management decisions on when and how to address enterprise risks. Blending the ability for identifying critical dynamic segments using SPC from a safety perspective with risk-informed economic analysis oprovides a path forward during good and poor financial times.

Establishing a risk management system as further outlined in a related paper Alfano and Weichel (2016, [8]), utilizes this information allowing for a cost-effective safety and financial assessment of the enterprise. The resulting system can be used for both managing current assets and also for evaluating potential mergers and acquisitions. This allows an organization the opportunity to identify the potential additional aggregate risk which is transferred once they take ownership, and compare this to the potential financial return.

Since early 2016 the industry continues to face tremendous financial pressures with the reduction in oil prices. The long-term organization survival requires economically sound integrity and investment management. A blend of a statistically SPC driven identification opportunities with solid well established economic evaluations will help both financial managers and engineers create a defensible organizational plan to address stakeholder concerns and those of the regulatory bodies.

5. Summary

By design, pipeline risk assessments are meant to be used to prioritize preventative and mitigative measures. In order to do this effectively, there needs to be a clear link between risk related results and the chosen controls. While this may sound simple, this process can be anything but trivial. The risk management process can often be complex, and it will vary based on company policies and systems.

The use of statistical process control provides a sound way to identify the portions of the pipe that need more in-depth attention and possible replacement or repair. SPC is a long-proven methodology and one that aids management and engineers in quickly and thoroughly assessing risk-based priorities. The graphical SPC plots provide an easy to understand assessment exhibiting either expected loss or frequency of failure for any or all failure modes for the dynamic segments of the line. Combining SPC and enterprise financial analysis aids the development of both short and long term sustainability of the organization.

References

[1] Shewhart, Walter A., 1931, Economic Control of Quality of Manufactured Product, D. Van Nostrand Co., Inc., New York (50th anniversary commemorative reissue, American Society for Quality)

[2] Breyfogle, Forrest W. III, 2003, Implementing Six Sigma, 2nd ed., pp. 226-227, John Wiley & Sons, Inc., New Jersey

[3] Wheeler, Donald J. and David S. Chambers, 1992, Understanding Statistical Process Control, 2nd Ed., SPC Press, Knoxville, Tennessee

[4] Duncan, [4] Acheson J., 1965, Quality Control and Industrial Statistics, 3rd ed., p. 908, Richard D. Irwin, Inc., Homewood, IL

[5] Balestracci, Davis, 2011, Four Control Chart Myths from Foolish Experts, Quality Digest (March 30, 2011), available on the web at //www.qualitydigest.com/inside/qualityinsiderarticle/fourcontrolchartmythsfoolishexperts.html#

[6] Wheeler, Donald J., 2010, "Are You Sure We Don't Need Normally Distributed Data?", Quality Digest (November 1, 2010) also can be found at http://www.qualitydigest.com/inside/twitter-ed/are-you-sure-we-don-t-need-normally-distributed-data.html#

[7] Burr, Irving W., 1976, Statistical Quality Control Methods, Marcel Dekker Inc., New York

[8] T. Alfano, M. Weichel, "Ensuring Pipeline Risk Model Usability," 11th International Pipeline Conference, paper no. IPC2016-64025, (Calgary, AB: ASME, 2016)