

Monitoring Laboratory Methods to Ensure Data Reliability

Testing laboratories associated with manufacturing plants play a key role assisting production personnel in monitoring the manufacturing process. They are often chartered with providing customers, both internal and external, with proof that the material or products sold meet customer specifications. To ensure that the data they produce provides the users with the appropriate level of reliability, the laboratories need to monitor their own processes, the testing procedures themselves.

The methods used to monitor a process, track conformance to specifications, and evaluate the measurement process are known collectively as Statistical Process Control (SPC). SPC enables an organization to track and reduce process variability with tools such as control charts. Laboratories often refer to the use of SPC methods in their internal quality control program as Statistical Quality Control (SQC).

Laboratories use the terms precision and accuracy to characterize how well a method performs. Precision describes how well a measurement can be repeated. Accuracy refers to the closeness of the measured value to the true value. Figure 1 illustrates how accuracy and precision work together to define method performance.

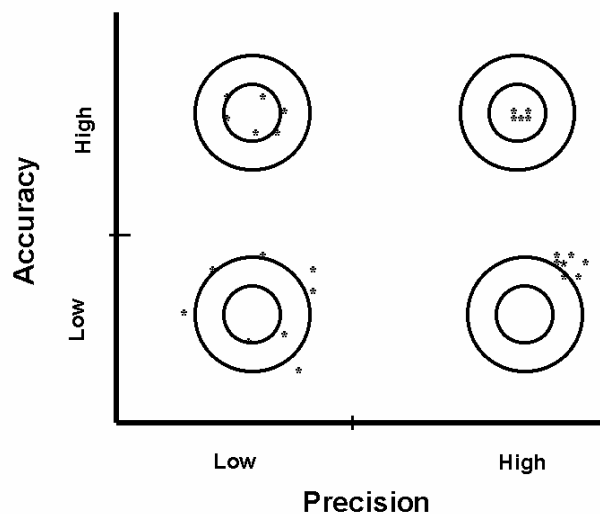


Figure 1 - Accuracy and Precision

Ideally, customers of testing data want their test results to be generated from a method that is both accurate and precise. In reality, to achieve a high level of reliability may be cost prohibitive or too time consuming. In order to determine if the data is fit for use, the laboratory and its customer must understand how the data will be used, what types of decisions will be made using the data, as well as industry standards and regulatory guidelines. Control charts measure the level of method variability as well as provide ongoing feedback to laboratory staff. In this paper we will look at how a laboratory can use the control charting capabilities of NWA Quality Analyst[®] to monitor the performance of its testing methods.

Types of Errors

To successfully use control charts to monitor and manage variation in analytical processes, the analyst must first understand the sources of variability. Four types of errors characterize the uncertainty of a test result:

1. Random errors
2. Variable systematic errors
3. Constant systematic errors
4. Gross errors

1. Random Errors

Random errors are uncontrollable fluctuations that will be seen on either side of the mean. They may be caused by environmental factors such as small changes in temperature, humidity, or barometric pressure. They can be caused by electrical fluctuations or differences in sample homogeneity. All glassware used to contain or deliver materials has published tolerances in which the true volume is found. Balances have published tolerances as well, and will add to variability of the test method. This type of error is often thought of as “noise” and is uncontrollable or indeterminate in nature. No matter how carefully a method is validated or performed, there is always some variability in the measurement due to random errors. Ideally, the analyst wants these errors to be small and consistent.

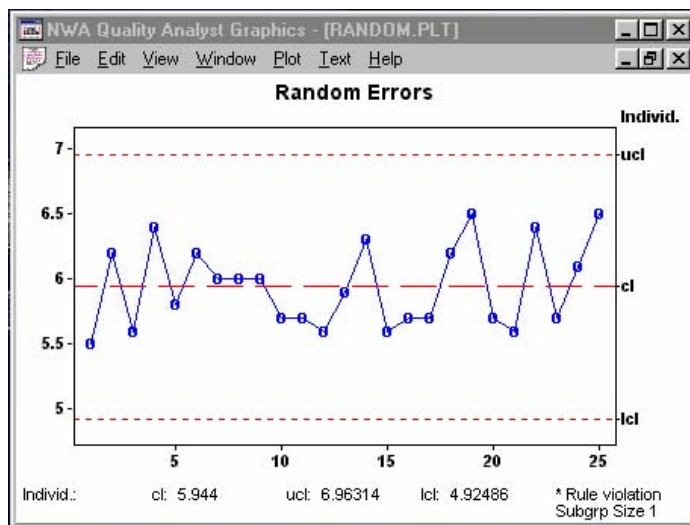


Figure 2 - Random error illustrated

2. Variable Systematic Errors

Systematic errors arise from the measurement system itself and can be classified as variable or constant. *Variable systematic errors*, like random errors, can appear on both sides of the mean, but are much larger in magnitude. These errors can be identified and removed or reduced if necessary. Some examples of factors that cause variable systematic errors are large environmental changes, differences between analysts, static build-up inside balance housing, and dirty glassware. The presence of random and systematic errors will affect the precision of a test method.

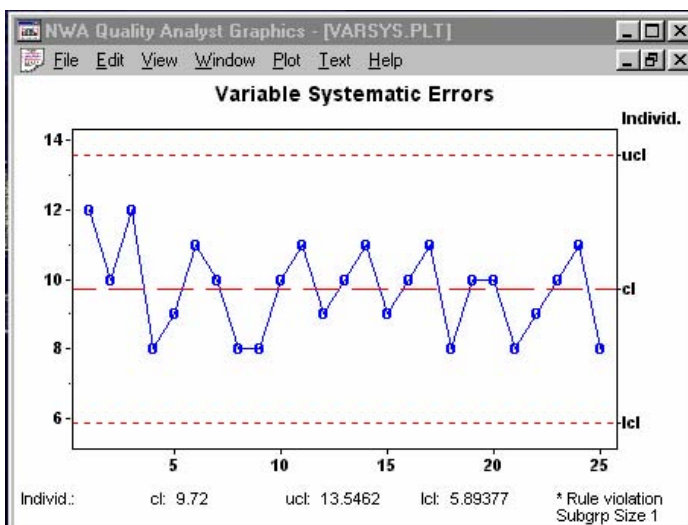


Figure 3 – Variable systematic error illustrated

3. Constant Systematic Errors

Constant systematic errors can be identified by a constant offset, either positive or negative, from a mean value or true value of a reference material. This type of error may be called bias and will affect the accuracy of a method. Constant systematic errors can be caused by improper instrument calibration, reagents used at the wrong concentration, or glassware used incorrectly. Both variable and systematic errors are determinate errors because the source of the error can be identified and corrected.

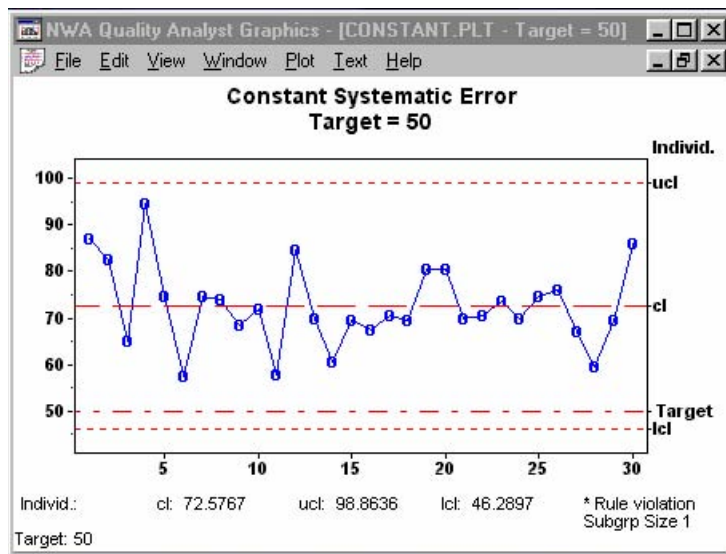


Figure 4 - Constant Systematic Error illustrated. Note target value = 50.

4. Gross Errors

Uncertainty can also be **introduced by gross errors**. These errors occur when material is spilled, calculations are performed incorrectly, etc. Usually, these errors are caught by the analyst, the mistake noted, and the test repeated.

Monitoring the Precision of Test Methods

The complexity of determining method precision depends on the nature of the test and the requirements surrounding the method. In its simplest form, the standard deviation can be calculated on repeat measurements made on a representative sample. With a very small sample set, the range can be used as a measure of precision. The more precise the method, the smaller the standard deviation or range. To determine the precision of a method in this manner can give a better picture of the method's inherent performance than what is seen on a day to day basis since it is likely that the measurements will be taken on the same day with the same analyst, using the same equipment and reagents.

A more rigorous method to determine method uncertainty involves conducting a Propagation of Errors determination which identifies all of the components of the method that contribute uncertainty, and measures their combined variability. Differences in analyst, equipment and environment can all contribute to method variability and should be included in the measurement of precision.

To monitor precision over a long time period, the laboratory needs a stable sample that contains the analyte in a matrix similar to production samples. Such a sample may be available commercially as a certified standard, or may be obtained in-house. When this material has been found, testing can be conducted over a period of time and the standard deviation can be calculated.

The next step is to plot the results on a control chart. Because testing can be time consuming or expensive, the most common laboratory control chart is the individuals chart where the each value is plotted on the chart rather than averaging repeat measurements to make an X-bar chart. The control chart gives the analyst a graphical representation of how the method performs over time. Trends and cycles can be easily seen and the control limits give a measure of method variability.

Control limits can be calculated a number of ways. Quality Analyst provides two ways of calculating control limits.

1. Use the average range to estimate the standard deviation of a process.
2. Calculate the standard deviation directly from the individual data points.

The first method calculates a more robust estimate of the standard deviation. The control limits will be narrower than control limits calculated based on the actual standard deviation. The second method is preferred by some regulatory agencies. It is important to indicate on the control chart which method is used. All control limits in this paper have been calculated using individual measurements.

The key to establishing this type of control chart is to have a control sample that is both *stable*, and *representative* of the process being monitored. There are times when these two criteria cannot be met due to the nature of the material. In this case, precision can be tracked by conducting multiple measurements (usually 2) on an actual sample. The difference between the two measurements (range) can be plotted on a control chart. This gives the laboratory staff the standard deviation of duplicate measurements over time as well as a graphical picture of how the method is performing.

Duplicate Chart

The colorimetric analysis of sulfide in wastewater by methylene blue is a fairly easy test to perform, but a stable control standard is not available. Standards prepared for calibration must be used within one hour, and should be prepared daily. Process samples are also unstable. This is a case where duplicate measurements on actual samples are a suitable technique to track the precision. Table 1 shows a subset of data that will be plotted using Quality Analyst. Note that in addition to recording the date and two data points for each sample, the analyst performing the test is also recorded. An additional comment field is included to capture other information that may affect the test result.

Row	DATE 1---(D)	DUP_1 2---(I)	DUP_2 3---(I)	DIFF 4---(I)	RPD 5---(I)	ANALYST 6---(A)	COMMENT 7---(A)
1	2/2/1999	0.37	0.37	0.00	0.00	BJB	*
2	2/5/1999	0.38	0.35	0.03	8.22	RLF	*
3	2/8/1999	0.38	0.33	0.05	14.08	JAL	*
4	2/11/1999	0.37	0.36	0.01	2.74	BJB	*
5	2/14/1999	0.41	0.39	0.02	5.00	BJB	*
6	2/17/1999	2.33	2.51	0.18	7.44	BFB	Unit Upset
7	2/20/1999	0.36	0.42	0.06	15.38	JAL	*
8	2/23/1999	0.36	0.41	0.05	12.99	JAL	*
9	2/26/1999	0.36	0.37	0.01	2.74	SHN	*
10	3/1/1999	0.65	0.74	0.09	12.95	SHN	*
11	3/4/1999	0.4	0.35	0.05	13.33	BJB	*
12	3/7/1999	0.37	0.39	0.02	5.26	BJB	*
13	3/10/1999	0.72	0.91	0.19	23.31	DCC	*
14	3/13/1999	0.35	0.38	0.03	8.22	DCC	*
15	3/16/1999	0.42	0.4	0.02	4.88	MRW	*
16	3/19/1999	0.39	0.37	0.02	5.26	MRW	*
17	3/22/1999	0.46	0.38	0.08	19.05	MRW	*
18	3/25/1999	0.39	0.33	0.06	16.67	MDB	*
19	3/28/1999	0.37	0.4	0.03	7.79	MDB	*

Table 1 - Subset of Sulfide data table

The absolute difference between duplicate value 1 and value 2 is easily computed using the calculation editor. The calculated values are shown in blue text in Table 1. The control chart (Figure 5) is easily constructed from the difference data.

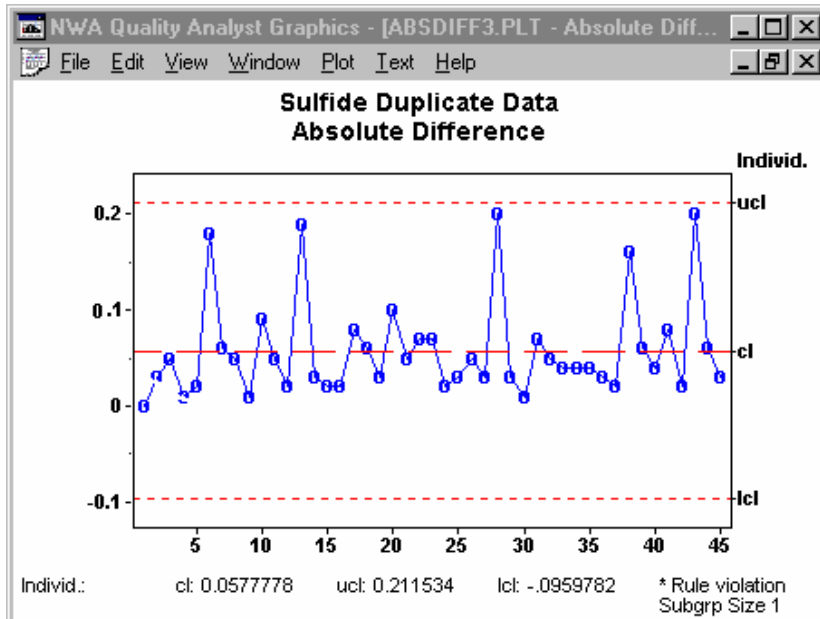


Figure 5 - Absolute Value Control chart

The difference calculated between each duplicate represents the precision associated with the analyst performing the test on a single day. The differences seen day to day represent variability due to different analysts, different environmental conditions, and other changes that can affect this method.

None of the data are outside of the control limits, even with the unit upset on 12/17/1999 (row 6). The plotted data represent differences between measurements and not the original measured values. Rather than tracking what is going on in the unit, the absolute difference gives us an indication of the method variability. There are, however, a number of points that show an absolute difference greater than 0.1. Figure 6 shows each of these data points marked with their corresponding row number.

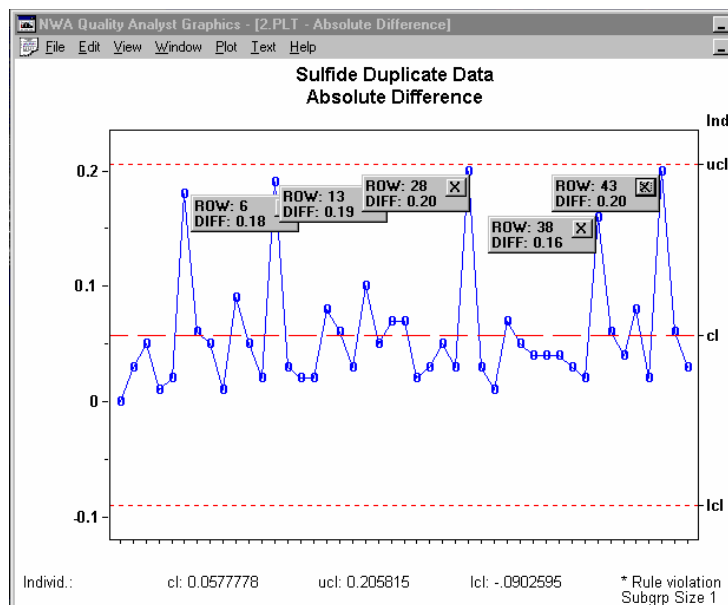


Figure 6 – Points with Difference >.1

Looking at the original data, many of these points correspond to higher levels of sulfides in the sample. It is not uncommon for the imprecision of the test to increase with increasing values of the analyte. Two control charts can be kept or the Relative Percent Difference (RPD) can be calculated and plotted on a single chart.

RPD is equal to the difference between the two duplicates, divided by the average value of the duplicates, and multiplied by 100.

$$RPD = [abs(Dup1 - Dup 2)/Avg (Dup1: Dup2)] * 100$$

This calculation is easily entered into Quality Analyst and the resulting chart is shown in Figure 7.

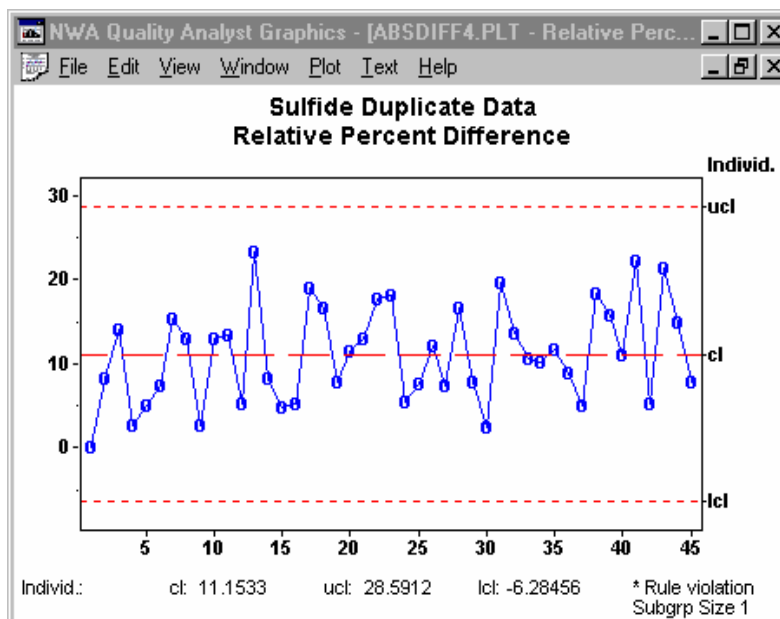


Figure 7 - Relative Percent Difference chart

There are no out-of-control points indicated on this chart and the data seem to be randomly distributed around the mean. The average RPD of 11% may be high depending on how the data will be used and may warrant improvement.

Monitoring Method Accuracy

Quantifying and monitoring method accuracy is more difficult than measuring precision. Accuracy is the closeness of the measured value to the true value. The true value is a quantity that can be estimated, but is unknown. Accuracy is affected by the amount of method imprecision as well as bias. To determine bias the analyst needs a reference material with a known mean value that is similar in composition to the material measured on a day to day basis.

While certified reference materials are available from standards organizations, they are quite expensive and often do not represent the material being tested. When used, they can give the laboratory the assurance that the method is being conducted properly with regards to the reference material. Sampling and matrix effects associated with actual samples make it difficult to translate the results obtained from a highly characterized standard to actual samples tested in the laboratory.

Spike recovery quality control samples can indicate the method bias under daily operating conditions. Because spike recovery is calculated from multiple tests conducted on the same sample, the random and variable systematic errors should be low. Usually, the sample is analyzed in duplicate and a known amount of the analyte is added to a third sample. Percent spike recovery (PSR) is calculated using the following equation:

$$PSR = 100 * \frac{[(\text{Measured spike concentration} - \text{average concentration of unspiked duplicates})]}{[\text{Amount of spike added}]}$$

If there were no interferences or matrix effects and the variability was low, one would expect the recovery to be close to 100%. Spike recoveries that are always low or always high indicate a method bias that should be investigated. If testing is being conducted under regulatory guidelines, spike recovery limits may be supplied by the agency. Spike recovery limits of 75% - 125% or 80% - 120% are common.

Spike Recovery Chart

Spike recovery data can be collected, calculated and charted easily using Quality Analyst. Table 2 below shows the typical data that would be collected for this chart for nickel tested using atomic absorption spectroscopy.

Row	DATE 1---(D)	ANALYST 2---(A)	DUP1 3---(I)	DUP2 4---(I)	SPIKE 5---(I)	RECOVERY 6---(I)	COMMENTS 7---(A)
22	03/21/99	SSB	24.0	22.5	28.1	97.0	*
23	03/22/99	SSB	12.5	14.5	18.5	100.0	*
24	03/23/99	KMT	27.5	28.1	31.2	*68.0	New hire
25	03/24/99	KMT	23.8	24.0	28.2	*86.0	*
26	03/25/99	KMT	19.6	17.4	22.6	*82.0	*
27	03/26/99	KMT	21.2	19.8	24.2	*74.0	*
28	03/27/99	KMT	22.0	20.7	25.3	*79.0	*
29	03/28/99	SMT	20.6	22.5	26.5	99.0	*
30	03/29/99	SMT	20.5	20.1	25.2	98.0	*
31	03/30/99	SMT	23.6	23.5	28.5	99.0	*
32	03/31/99	SMT	28.9	27.5	32.8	92.0	*
33	04/01/99	SMT	26.3	24.9	30.5	98.0	*

Table 2 - Raw spike recovery data

Using the calculation editor, the average of the unspiked duplicates and spike recovery can be calculated. The control chart of the percent spike recovery values is shown in Figure 8.

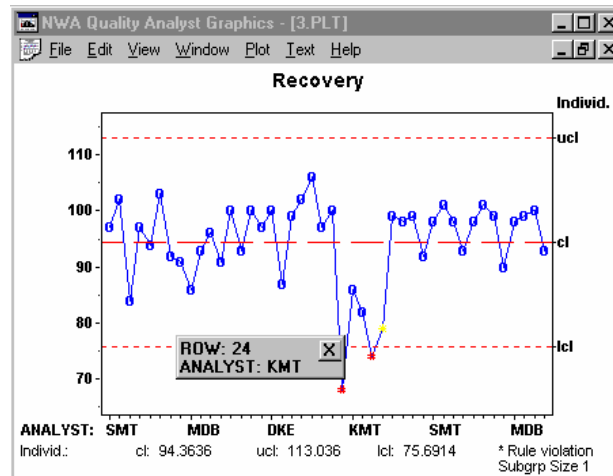


Figure 8 - Spike control chart

Looking at the control chart, there is a section of data that has fallen below the lower control limit and that is also showing some pattern rule violations. Going back to the data, we can see that there was a newly hired analyst performing the analysis. This person produces results that are consistently low and is introducing constant systematic error. All the data generated by this analyst is suspect and should be recalled and the samples rerun if possible. Continued training and closer supervision are warranted until improvements are seen.

Because there is an assignable cause associated with this data, it can be tagged or set aside and not used in subsequent data analysis. The recalculated control chart is shown in Figure 9.

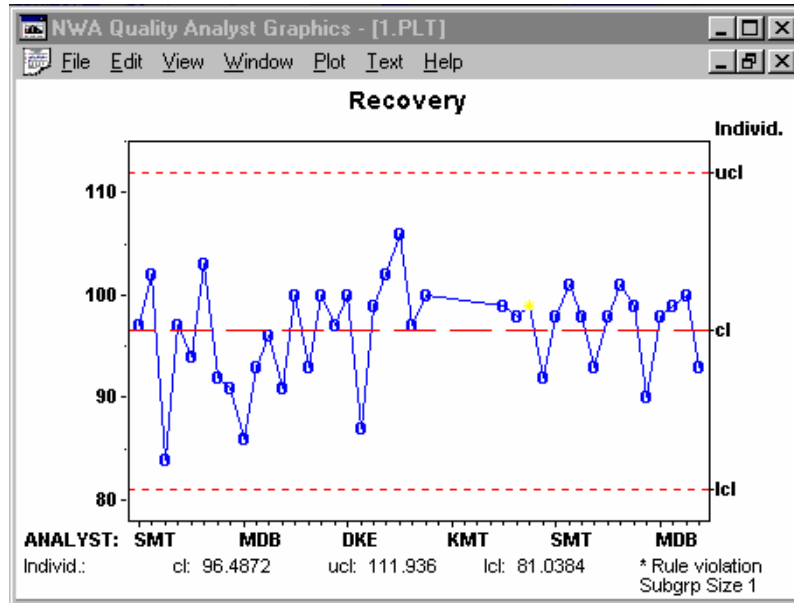


Figure 9 - Spike Recovery with New Hire Data Tagged

Capability Analysis

EPA guidelines suggest that spike recoveries fall in the range of 75% - 125% recovery if the result is to be considered valid.

The expected level of performance can be treated as specifications. Capability analysis looks at how wide the measured distribution is compared to the specification limits. Before performing such an analysis, however, the process must be stable and in statistical control. Some common measures of capability are the Cp and Cpk indices. The equations are given below.

$$Cp = \frac{\text{Upper specification limit} - \text{Lower specification limit}}$$

$$6 * \text{Standard deviation}$$

$$Cpk = \frac{\text{Average} - \text{Nearest specification limit}}$$

$$3 * \text{Standard deviation}$$

The higher the ratios, the more “capable” the method is of meeting specifications. Capability is represented graphically by plotting a histogram of the data. If the mean and specification limits are added to the chart, the user can see if method bias is present as well as seeing how well the method is performing with regards to the specification. The calculated values of the indices are also reported on the graph shown in Figure 10.

In this example, both Cp and Cpk are greater than one, which means that the process (in this case, the analytical method) is capable of consistently producing data within the spike recovery guidelines. However, there may be an

opportunity for method improvement by investigating the points that fall below 90% recovery and those that fall above 105% recovery.

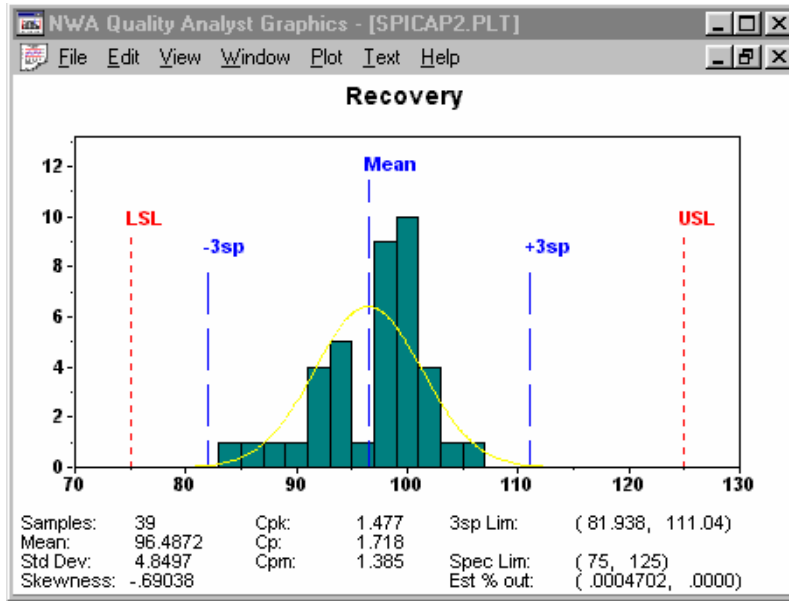


Figure 10 - Capability plot of Nickel Spike Recovery

Conclusion

Evaluating and monitoring the measurement process are key activities in any SPC program. Understanding the nature of measurement imprecision and being able to quantify testing imprecision will give the consumers of the data an indication of reliability.

Monitoring the testing process at some predefined frequency gives the analyst and laboratory the assurance that the test method is in control and that results can be released to production personnel with confidence. Having control standards and charts may be a requirement of regulatory agencies and are an indication that the data produced by the laboratory are defensible.

The proper tools such as NWA Quality Analyst®, which can store the data, perform calculations and produce a variety of control charts, simplify the monitoring program and reduce laboratory workload.

For More Information

Northwest Analytical
Phone: 503-224-7727
Toll-free: 888-692-7638
Fax: 503-224-5236
sales@nwasoft.com
www.nwasoft.com

Lisa Templeton has 20 years of experience as an analytical chemist in a number of different industries. She began working as a research chemist for the U.S. Air Force where she learned the power of statistical techniques for validation of analytical methods. After moving to private industry as a laboratory supervisor for a sulfuric acid plant, Ms. Templeton was asked to lead a plant wide quality initiative, which included teaching all plant personnel statistical process control techniques. She later moved back to R&D with a major oil company. Her group developed and taught classes company wide in laboratory Quality Assurance/Quality Control and Sampling. These classes were tailored to meet the needs of each client and included extensive follow up work to facilitate implementation. She moved back to an industrial position as the laboratory supervisor at a refinery. There, she implemented a laboratory information management system (LIMS) which included an integrated SPC program. Now a consultant in the field of laboratory QA/QC, she also teaches applied SPC classes for Northwest Analytical.

©2004 Northwest Analytical, Inc., all rights reserved. NWA Quality Analyst is a registered trademark of Northwest Analytical. *All other company and product names may be trademarks or registered trademarks of those respective companies.

Northwest Analytical, Inc. • Phone: 503-224-7727 • Toll-free: 888-692-7638 • Fax: 503-224-5236
sales@nwasoft.com • www.nwasoft.com