WHAT YOU CAN LEARN FROM THE DATA – DATA MINING (STATISTICAL PROCESS CONTROL)

David Hutcheson, PhD
Animal-Agricultural Consulting, Inc
Amarillo, TX

Introduction

Statistical Process Control (SPC) was introduced and pioneered in the early 1920's. SPS methods were used in the US during World War II to improve the quality in the manufacture of munitions and other strategically products. SPC is an effective method of monitoring a process through the use of control charts. Control charts enable objective criteria for distinguishing background variation from events of significance based on statistically techniques. SPC monitors the process and variation about the process. By collecting data from samples at various points within the process, variations that may affect the quality of the final product may be detected and correct, thus reduce wastage. In addition to reducing wastage, SPC can reduce production time to produce the product.

In manufacturing particularly feed manufacturing the quality is by observation prior to being fed. Nutrient analysis may be performed on a routine basis but data is after the fact. In contrast SPC uses statistical tools to observe the performance of the production process and predict significant deviations that may affect the final mixed feed from being outside the specifications.

There are two kinds of variations that occur in processes. First type is variation known as common causes such as temperature, moisture, moisture in the ingredients and the ingredients (feed mixing). The variations are usually small and near to an average value and are usually normally distributed. This would be considered controlled variation for SPC methods. The second type variation is a special case and is less likely to appear during the process than the first type variation, common cause. An example is in feed mixing, the ingredients that are added may vary from within the batch by being either less or more than called for in the formula. This may be to operator error or scales. The mixer may not be performing within tolerances of the manufacturer, either due to worn parts or not mixing properly. This type of variation would be considered as uncontrolled variation.

There are three phases to SPC:

1. Stabilization of the process by the identification and elimination of special causes.
2. Active improvement efforts on the process from common causes.

Monitoring the process to ensure the improvements are maintained, and incorporating additional improvements as the opportunity arises (control charts).

Control Charts

The fundamental of SPC is the control chart. The control chart is the charting of the data collected and recording measurements over time and analyzing the chart on a timely basis. The control chart is a graph used to study how a process changes over time. Data are plotted in time order. A control chart always has a central line for the average, an upper line for the upper control limit and a lower line for the lower control limit. These lines are determined from historical data. By comparing current data to these lines, you can draw conclusions about whether the process variation is consistent (in control) or is unpredictable (out of control, affected by special causes of variation).

The control chart should be used when:

1. Controlling ongoing processes by finding and correcting problems as they occur.
2. Predicting the expected range of outcomes from a process.
3. Determining whether a process is stable (in statistical control).

4. Analyzing patterns of process variation from special causes (non-routine events) or common causes (built into the process).
5. Determining whether your quality improvement project should aim to prevent specific problems or to make fundamental changes to the process.

All SPC control charts have the following properties:

1. The x-axis is sequential, usually a unit denoting time
2. The y axis is the statistic charted for each time
3. Limits are defined for the statistic that is being plotted. The statistical control limits are determined from previous and/or current process data, providing an indication of bounds of the expected process behavior.

In order to define control limits the following information is needed:

1. A good history of the process to define the common cause variation
2. The history will be a basis for determining how wide to set control limits

There needs to be enough sub group data to distinguish between common cause and special cause of the variation. Control limits are determined mathematically and the formula used for the computation is normal probability theory. The most common used computational formulas used are the mean or average and standard deviation. Curve fitting can be used when distribution is not known and upper and lower limits can be calculated. Standard deviations of the estimated can be calculated from regression analysis.

Types of Charts

There are several types of charts that can be used in SPC:

1. Pareto Charts

A Pareto Chart is a vertical bar graph showing problems in a prioritized order and therefore can be determined which problems should be considered first.

2. X-Bar / Range Charts

X-bar and Range Charts are for data that is both quantitative and continuous in measurement. The X-bar chart monitors the process over time, based on the average of a series of observations, called a subgroup. The Range chart monitors the variation between observations in the subgroup over time.

3. X-Bar / Sigma Charts

X-bar and Sigma Charts are for data that is both quantitative and continuous in measurement. The X-bar chart monitors the process over time, based on the average of a series of observations, called a subgroup. The Sigma chart monitors the variation between observations in the subgroup over time. The sigma chart is a standard deviation chart.

4. Individual-X Charts

Individual-X and Moving Range Charts are for data that is both quantitative and continuous in measurement. The Individual-X chart monitors the process location over time, based on the current subgroup, containing a single observation. The Moving Range chart monitors the variation between consecutive subgroups over time.

5. Histograms / Process Capability Analysis

The graphics presented with this feature include a histogram of the data, with a distribution curve overlaid. The statistics provided largely deal with the capability indices. Capability indices attempt to indicate, in a single number, whether a process can be consistent or not.

6. Scatter Diagrams

A Scatter Diagram examines the relationships between data collected for two different characteristics. Although the Scatter Diagram cannot determine the cause of such a relationship, it can show whether or not such a relationship exists.

7. Autocorrelation Chart

The Autocorrelation chart is for identifying dependence of current data on previous data points. It tests for autocorrelation between observations of a given characteristic in the data set.

8. Exponentially Weighted Moving-Average (EWMA) Chart

EWMA is for data that is both quantitative and continuous in measurement. It plots weighted moving average values. A weighting factor is chosen by the user to determine how older data points affect the mean value compared to more recent ones. Because the EWMA Chart uses information from all samples, it detects much smaller process shifts than a normal control chart.

9. CuSum Chart

A Cu Sum Chart is for data which plots the cumulative sum of the deviations from a target. Because each plotted point on the Cu Sum Chart uses information from all prior samples, it detects much smaller process shifts than a normal control chart would.

10. Moving Average / Range Chart

Moving Average / Range Charts is for data that is both quantitative and continuous in measurement. The Moving Average chart monitors the process over time, based on the average of the current subgroup and one or more prior subgroups. The Range chart monitors the process variation over time.

11. Moving Average / Sigma Chart

Moving Average / Sigma Charts is for data that is both quantitative and continuous in measurement. The Moving Average chart monitors the process over time, based on the average of the current subgroup and one or more prior subgroups. The Sigma chart monitors the process variation over time.

12. Multivariate Chart

Multivariate Charts are used to detect shifts in the mean or the relationship (covariance) between several related parameters.

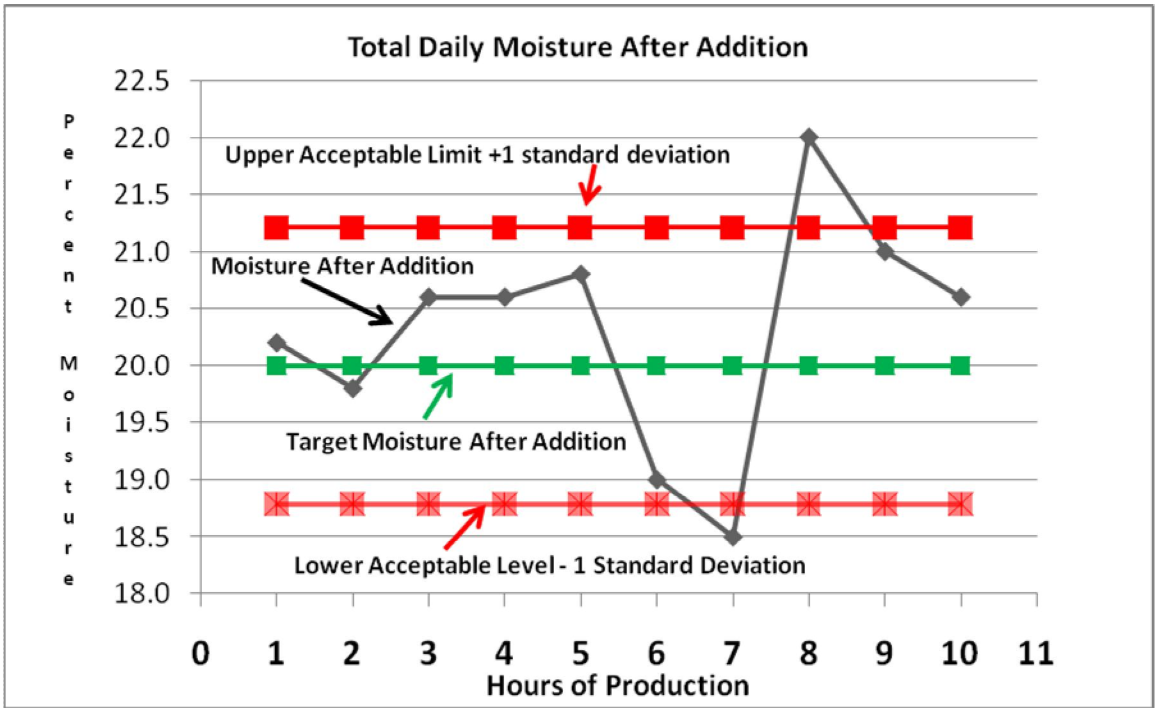
Charts can determine the variability in the process and is how consistently the process is performing. The charts allow the comparison of the performance to the required performance (Process Capability Index). The capability index may lead to design changes that improve the process.

Basic steps to Statistical Process Control

1. Choose the appropriate control chart for your data.
2. Determine the appropriate time period for collecting and plotting data.
3. Collect data, construct your chart and analyze the data.
4. Look for "out-of-control signals" on the control chart. When one is identified, mark it on the chart and investigate the cause. Document how you investigated, what you learned, the cause and how it was corrected.

Monitoring moisture additions during steam flaking is an example of a control chart/ sigma. In this example, the moisture is measured in the incoming grain and automatically water is added prior to the grain entering the steam chamber. This moisture is recorded automatically in the computer. The standard deviation of daily samples is determined and a daily chart is developed. Figure 1 is an example chart from moisture monitoring during steam flaking.

Figure 1 Moisture monitoring steam flaking



The upper and lower limits are plus or minus 1 standard deviation from the means of previous moisture data points recorded. At the seventh hour and the 8 hour the process was out of control both below the limit and above the limit. At the 7 hour the operator detected that not enough moisture was on the grain as it entered the steam flaker. The operator checked the system and found that nozzles that were spaying water on grain had become plugged. After unplugging the nozzles the extra water in the system was discharged and was above the upper control level. Continue to monitor the chart and as data points are charted check for new out-of-control signals. When a new control chart is started, the process may be out of control. When at least 20 sequential points from a period when the process is operating in control, recommend recalculate control limits.

Developing the control chart is fundamental to this process. Steps necessary to development control charts:

1. Define the purpose of the chart
2. State the aims of the chart
3. Define how data are collected
4. Ensure support is available during data collection and interpretation

Example of SPC in feed mixing of a complete ration

1. The purpose of the chart is to monitor the ingredients of the ration for opportunity to control feed cost.
2. The aims of the chart are to determine variability of loading the mixer wagon using a front end loader with scales on the mixer wagon.
3. The data will be collected after each ingredient is added to the batch.

The data is collected daily for each batch that is mixed. The data can be entered into the computer program or a data logger and can be used to transfer data to the computer for analyzing. Means and standard deviations are calculated accumulative for developing the SPS charts. The charts are developed daily with the control limits Plus or minus 1 standard deviation. After 14 days of data points the variability around the addition of the ingredients allow the charts to be developed and the control charts are calculated and plotted. The charts determine the capability of the mixing system.

The example is the result of weighing individual ingredients into a mixer wagon and the associated variability of weighting ingredients with a front end loader. The ration and batch mix is shown in Table 1.

Table 1 Ration and Batch Mix.

Ingredient	Percent Ingredient	Weight of Ingredient,	Weight of

		lbs	Batch, lbs
Corn Silage	25%	2000	2000
Hay	30%	2400	4400
Corn	30%	2400	6800
Supplement	15%	1200	8000

Table 2 shows the data that was collected over 14 days. This was a batch mix of 8000 pounds.

Table 2. Batch weights as recorded from mixer

Day	Silage	Hay	Corn	Supplement
1	2020	4420	6800	8020
2	2000	4480	6780	8020
3	1960	4400	6820	7980
4	2000	4440	6780	8020
5	2020	4380	6820	8020
6	2040	4420	6840	8000
7	1980	4480	6820	8040
8	2020	4440	6820	8020
9	2040	4380	6800	8000
10	2040	4480	6840	8020
11	2080	4380	6780	8000
12	2000	4420	6840	8020
13	1940	4380	6840	8000
14	2020	4420	6780	8020
Mean	2011	4423	6811	8013
Standard Deviation	36	38	24	15

The following is the silage chart developed for the batch mixing to establish the capability of the mixing process (Figure 2). The process continually added more silage than was called for by the batch. The silage was the first ingredient added and apparently was difficult to achieve the target and the accumulated mean illustrates from the beginning was more than called for by the batch. The apparent difficulty is probably the loader operator not able to control the amount going into the mixer due to the type of ingredient.

Figure 2. Silage Chart

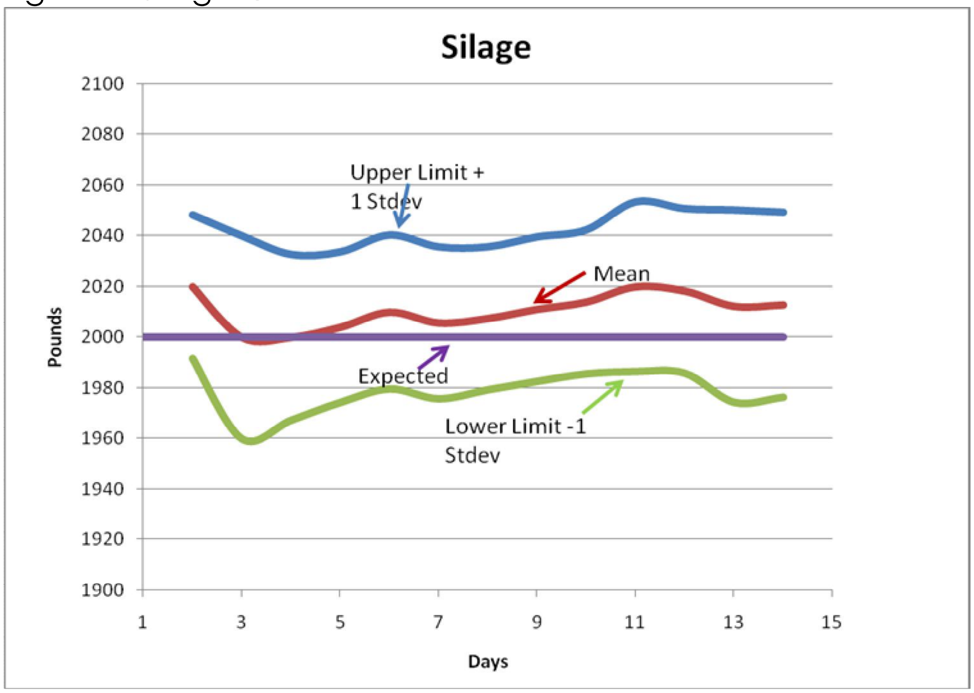
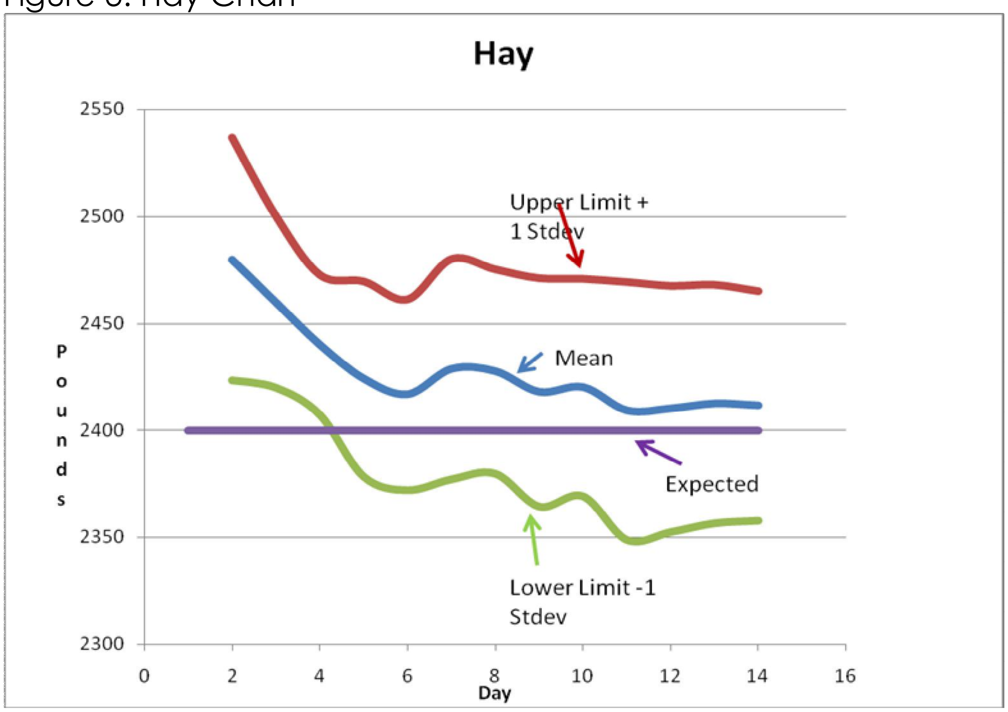


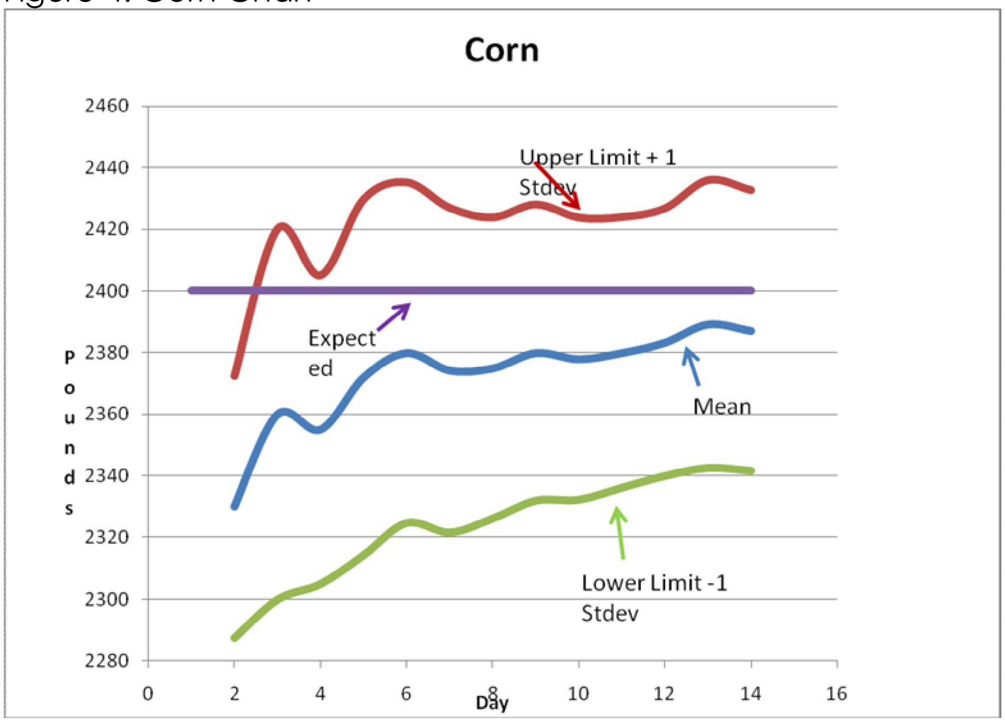
Figure 3. Hay Chart



The Hay chart, Figure 3 indicates that it was over the batch amounts and was probably due to operator difficulty of delivering the hay to the mixer.

The Corn chart is presented in Figure 4. The corn was delivered less than expected. The operator was able to deliver the corn to the target amount, however since the silage and hay was over the target amount the corn was below target amount.

Figure 4. Corn Chart



The supplement was delivered last and was a pelleted supplement and was delivered close to expected. (Figure 5). Since corn was delivered to close to target weight for the batch the supplement was close to targeted amount.

Note using accumulated means and standard deviations, after 14 days the process become somewhat stable. However the variation and means in the beginning was sub groups of small number of data points.

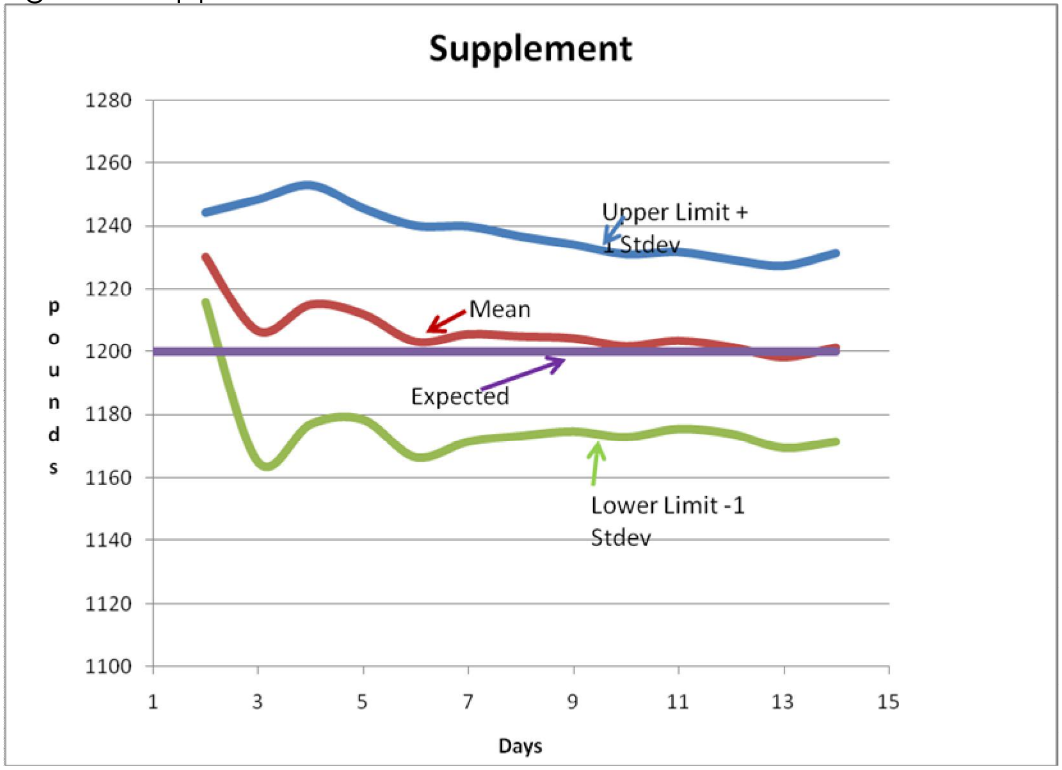
The variance of the 4 ingredients were

1. Silage ± 36 pounds and coefficient of variation is 25%

2. Hay ± 54 pounds and coefficient of variation is 30%
3. Corn ± 45 pounds and coefficient of variation is 30%
4. Supplement ± 30 pounds and coefficient of variation is 15%

Therefore, this mix system with supplement added last has the least variance and is the most expensive ingredient.

Figure 5. Supplement Chart.



The mixer charts illustrate

1. The difficulty of some ingredients to be added to the mixing process
2. When the ingredients added are more than targeted one of the other ingredients will be less than targeted
3. The last ingredient add, if easy to control will be close to target or less than expected.
4. The most expensive ingredient has the least amount of variance.

The charts allow statistical analysis of the process to determine capability of this process. The next step would be to establish charts of ingredients into the mixer with the accumulated standard deviations and chart the daily ingredients into the mixer; this would allow the monitoring of the process to determine if it is in control or out of control.

The variability of nutrients as presented in previous papers could be added to this variation and control charts for nutrients could be developed. This would determine the capability of deliver nutrients to the cattle and the expected upper and lower limits.

Summary

Statistical Process Control is easy to do. Although it involves complex mathematics.

Selected Examples of SPC

1. de Vires and Conlin, 2003a, developed SPC charts to monitor estrous detection efficiency in dairy cow herds. The SPC charts allowed for timely signaling of changes in estrus detection ratios to determine the average time to signal estrus.
2. de Vires and Conlin, 2003b.showed the economic value of estrus decreased detection efficiency.
3. Lukas, J.M., et.al. 2008a, SPC developed charts to assess whether corrective action needed and was corrective action successful in decreasing the mean variation in bulk tank somatic cell counts
4. Lukas, J.M., et. al. 2008b. Monitoring means and sigma of bulk tank somatic cells count has the potential of warning the producer of an upcoming violation.
5. Lukas, J.M., et. Al. 2005, indicated SPC charts monitoring individual bulk tank somatic cells counts and to monitor sub clinical mastitis.

6. Niza-Ribeiro, J., et. al. showed that the capability index could be used to determine the ability for compliance of bulk tank somatic cell counts among dairy herds.

References for Selected Examples

de Viries, A. and B. J. Conlin. 2003a. Design and performance of statistical process control charts applied to estrus detection efficiency. J. Dairy Sci. 86:1970-1984.

de Viries, A. and B. J. Conlin. 2003b. Economic value of timely determination of unexpected decrease in detection of estrus using control charts. J Dairy Sci. 86:3516-3526.

Lukas, J.M., J.K. Reneau, C. Munoz-Zant and M.L. Kinsel. 2008a. Predicting somatic cell count standard violations based on herd's bulk tank somatic cell count. Part II Consistency Index. J. Dairy Sci. 91:433-441.

Lukas, J.M., J.K. Reneau and M.L. Kinsel. 2008 b. Predicting somatic cell count standard violations based on herd's bulk tank somatic cell count. Part I Analyzing variations.

Lukas, J.M., D.M. Hawkins, M.L. Kinsel and J.K Reneau. 2005. Bulk tank somatic cell counts analyzed by statistical process control tools to indentify and monitor subclinical mastitis incidence. J. Dairy Sci. 88:3944-3952.

Niza-Ribeiro, J., J.P.T.M. Noordhuizen and J.C. Menezes. 2004. Capability Index-a statistical process control tool to aid in udder health control in dairy herds. J. Dairy Sci. 87:2459-2467.