Analysis of Highland Lakes Inflows Using Statistical Control Charts

Dr. William McNeese, Ph.D. - June 22, 2013
www.spcforexcel.com

Summary

Data for the yearly water inflows have been analyzed using statistical control charts. One purpose of a control chart is to determine if a process has changed. This is done by plotting the data over time, calculating the average and process limits using well-established statistical techniques and then interpreting the control chart. The presence of outliers or certain patterns are indications that the process has changed. The data for this analysis were supplied by David Lindsay, who obtained the actual annual inflow data from LCRA.

Data from 1942 – 2012 were analyzed. This statistical analysis provides the answer to the following question:

"Has the water inflow to the Highland Lakes changed significantly in this time period?"

Based on this analysis, the answer to this question is a definite **yes**. The analysis shows that the water inflow to the lakes has decreased significantly in recent years. This decrease is statistically significant – and real. The analysis does not address the reason(s) for the decrease – only that the decrease has occurred. The details of the analysis are given below.

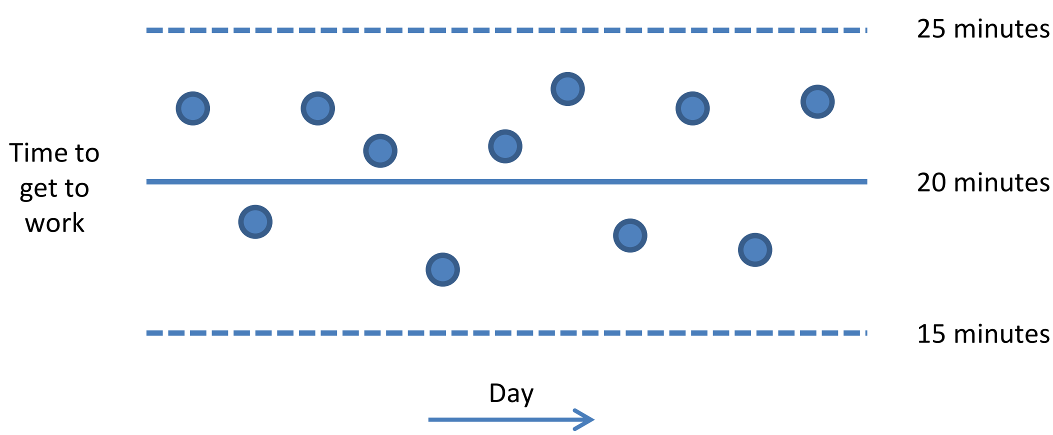
Analysis Methodology

Control charts were used to analyze the data. One purpose of control charts is to monitor a variable over time to see if anything has significantly changed. This type of chart focuses on two different types of variation: common causes and special causes. The following everyday example helps explain how control charts work.

Think about how long it takes you to get to work each day. Does it take the same time each day? No, of course not. But there is a certain average time it takes. Assume that this average time is 20 minutes. It does not take exactly 20 minutes each day. There is a range of time that you consider "normal" to your process of getting to work. Maybe one day it takes 18 minutes; another day it takes 23 minutes. But as long as the time is within a "normal" range, it does not concern you. Suppose this normal range is 15 to 25 minutes. You don't how long it will take you to get work tomorrow, but as long as things are "normal", the time will be from 15 to 25 minutes. The differences from day to day are simply due to the amount of traffic, the speed you drive, etc. This is called common causes of variation – it is the normal variation in the process. You might call it the baseline data to judge the future against.

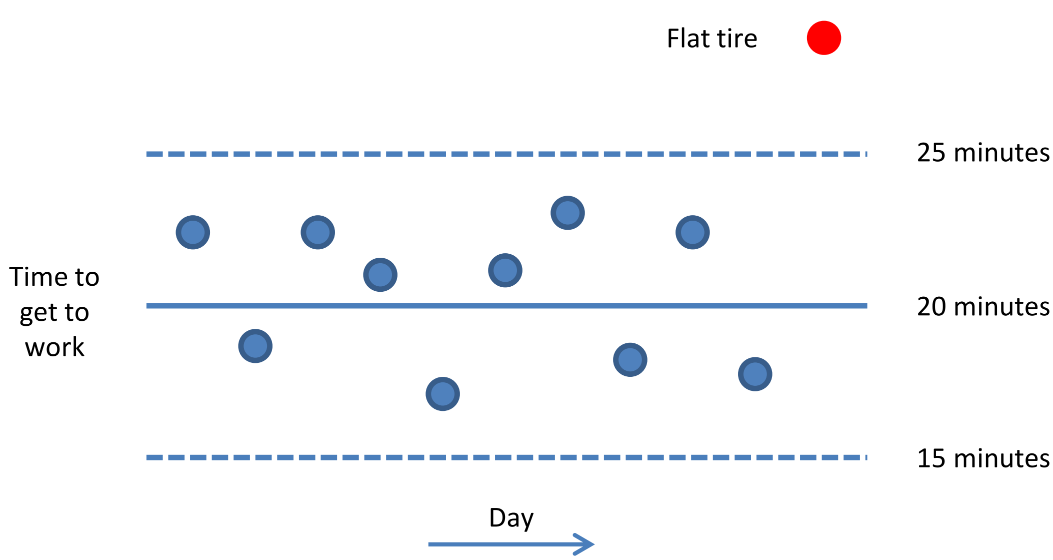
Figure 1 is an example of a control chart for your process of getting to work. The data are plotted over time. The top dotted line is called the upper control limit. It is the largest value you would expect from the process with just common causes (normal variation) present. The lower dotted line is called the lower control limit. It is the smallest value you would expect form the process with just common causes of variation present. As long as there are no points beyond the control limits and no patterns present, the process is said to be in "statistical control". Only common causes of variation are present and you can predict what will happen in the near future. This prediction is the key. You don't know how long it will take you to get to work tomorrow, but you know it will be between 15 and 25 minutes, with an average of 20 minutes.

Figure 1: Control Chart for the Time to Get to Work



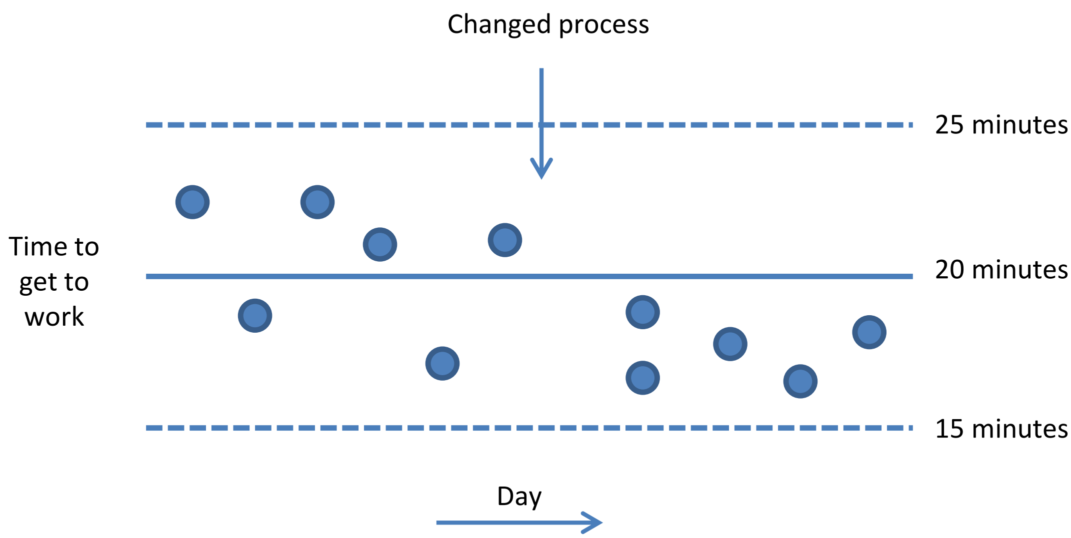
Now, suppose you have a flat tire when driving to work. How long will it take you to get to work? Definitely longer than the 15 to 25 minutes in your "normal" variation. Maybe it takes you 50 minutes to get to work. This is a special cause of variation. Something happened that has caused a change. It is not part of the normal process. These types of special causes are not predictable and are sporadic in nature. Figure 2 is an example of the impact of a "flat tire" on getting to work.

Figure 2: Control Chart with Special Cause (Flat Tire)



Once you fix the flat tire, the process will come back into statistical control. But there are special causes that represent significant shifts in the process average. For example, suppose you wanted to decrease the amount of time you get to work. So, you change your process. You get up earlier and drive a different route. This represents a fundamental change in the process. The control chart will show the impact of that change. Figure 3 shows such a process.

Figure 3: Control Chart with Changed Process



This type of shift will show up in a control chart as a special cause of variation. It is clear from the chart that the average has shifted downward. The new average for the process of getting to work can be estimated from the data. This is the type of shift that occurred with the inflow data from the lakes.

Inflow Data Analysis

The data used in the analysis are shown in Appendix A. The control chart used in the analysis was an individuals (X-mR) control chart. To use this type of control chart, the data must follow somewhat of a normal distribution. The inflow data are not normally distributed. The X-mR chart is very robust to non-normal data, but some may feel it is not a valid application for this type of chart. The histogram for the inflow raw data is shown in Figure 4.

The distribution is skewed towards the left. Inflow data are often transformed using the natural logarithm for analysis purposes. The data were transformed and the resulting histogram is shown in Figure 5. The data appear normally distributed. Figure 6 is a normal probability plot of the transformed data. This confirms that the transformed data are normally distributed. The p value for the normal probability plot is 0.467.

The X-mR control chart technique can now be applied to the transformed data. The X chart for the transformed data is shown in Figure 7.

The points plotted are the yearly inflow. There are additional lines plotted on the chart. One line is the average, which is 13.85. This is the average of the natural logarithms of the inflow data. The top line is the upper control limit (UCL). This represents the largest value one would expect if the process has not changed. The two lines between the average and UCL represent one and two standard deviations above the average. The bottom line is the lower control (LCL). This is the smallest value one would expect if the process has not changed. The two lines between the average and the LCL represent one and two standard deviations below the average. The lines between the average and the control limits are used to detect patterns that indicate a process has changed.

Figure 4: Histogram for Raw Inflow Data

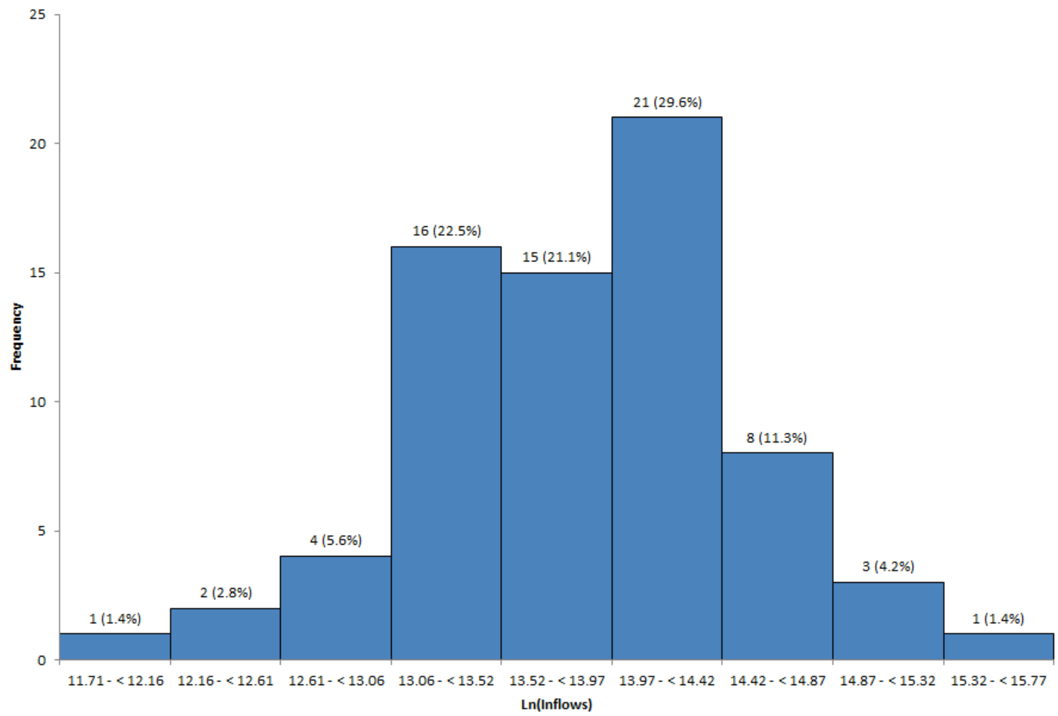


Figure 5: Histogram of Transformed Inflow Data

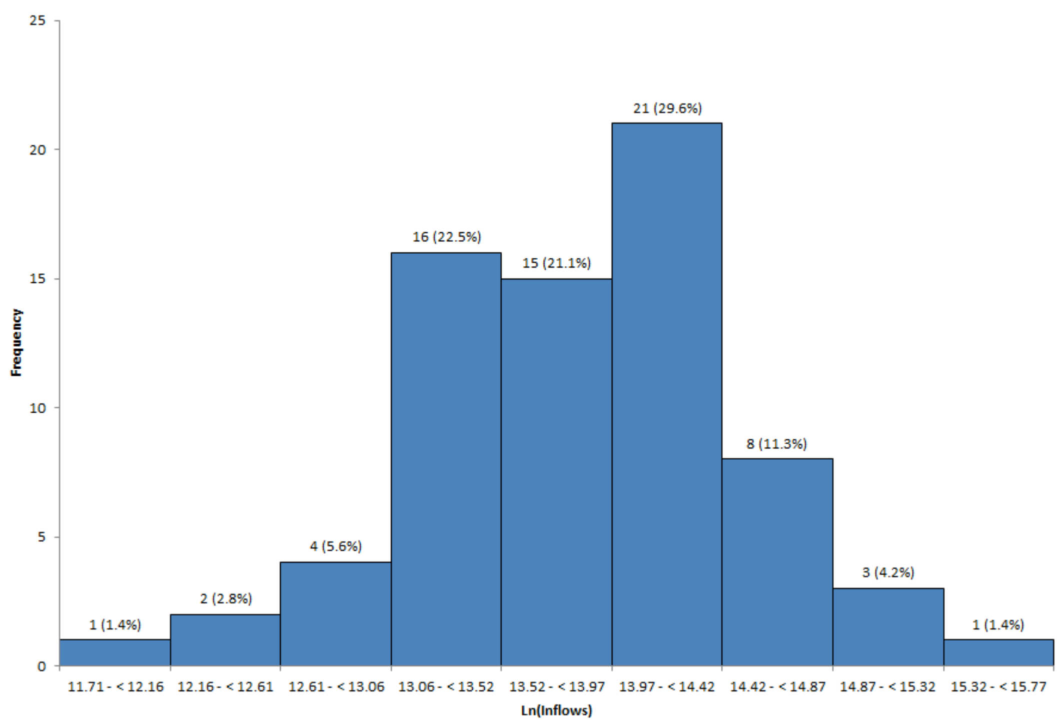


Figure 6: Normal Probability Plot of Transformed Inflow Data

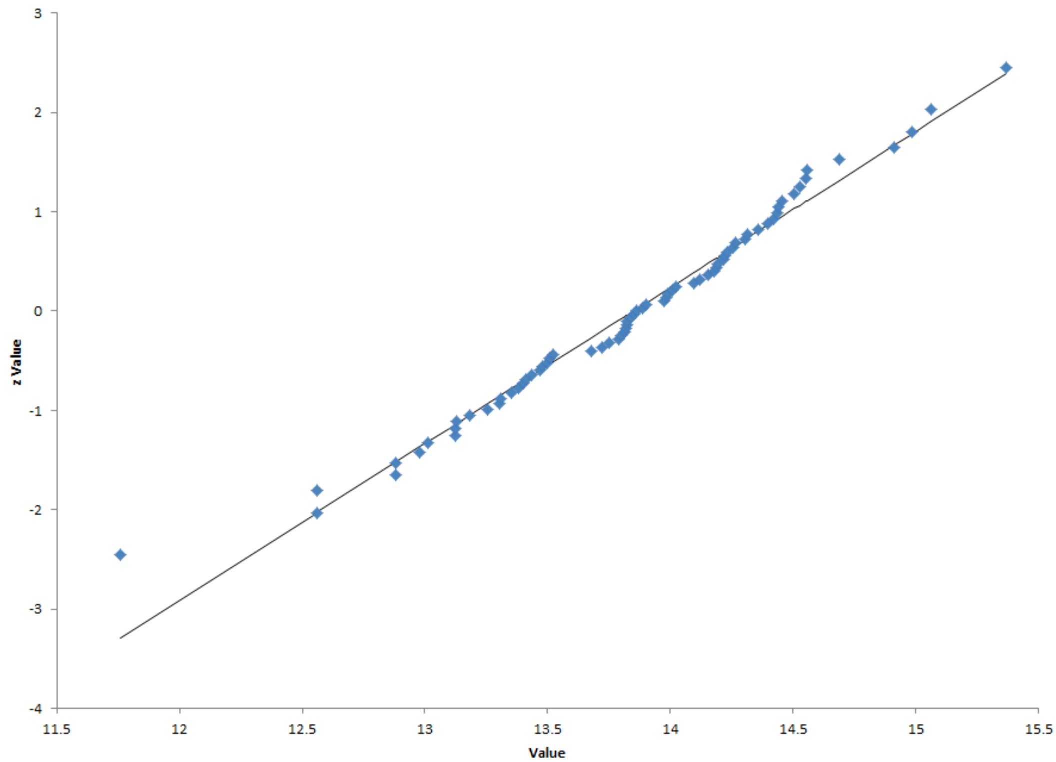
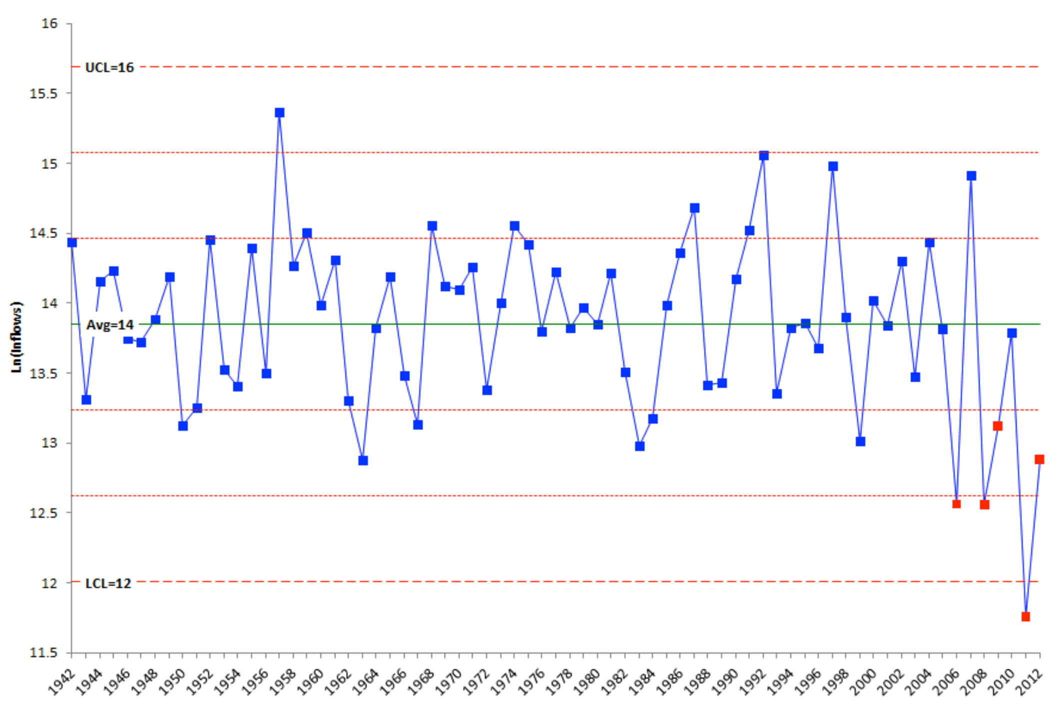


Figure 7: Control Chart for Transformed Inflow Data

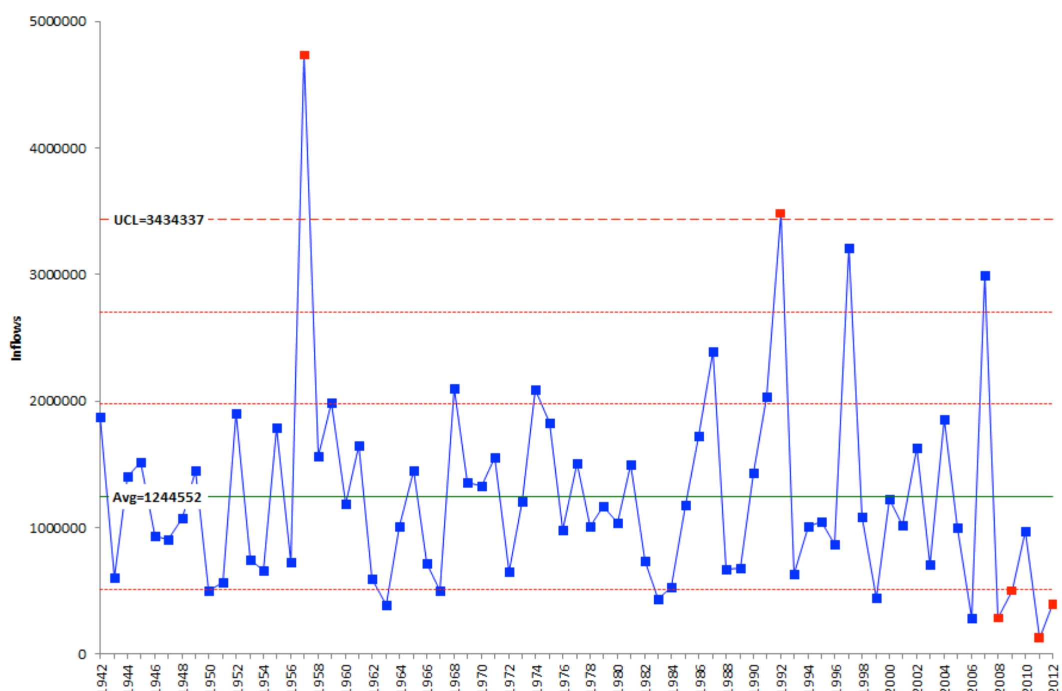


There are five red points at the right-hand side of the control chart. These red points represent “out of control” conditions – something has happened during that time frame to change the process. There are three out of control conditions beginning in 2006:

- 2006 to 2008: two out of three consecutive points are beyond two standard deviations below the average; 2008 is the confirmation year that the process has changed.
- 2006 to 2012: four out of five consecutive points are beyond one standard deviation below the average; 2011 and 2012 are the confirmation year that the process has changed
- 2011: point is beyond the lower control limit

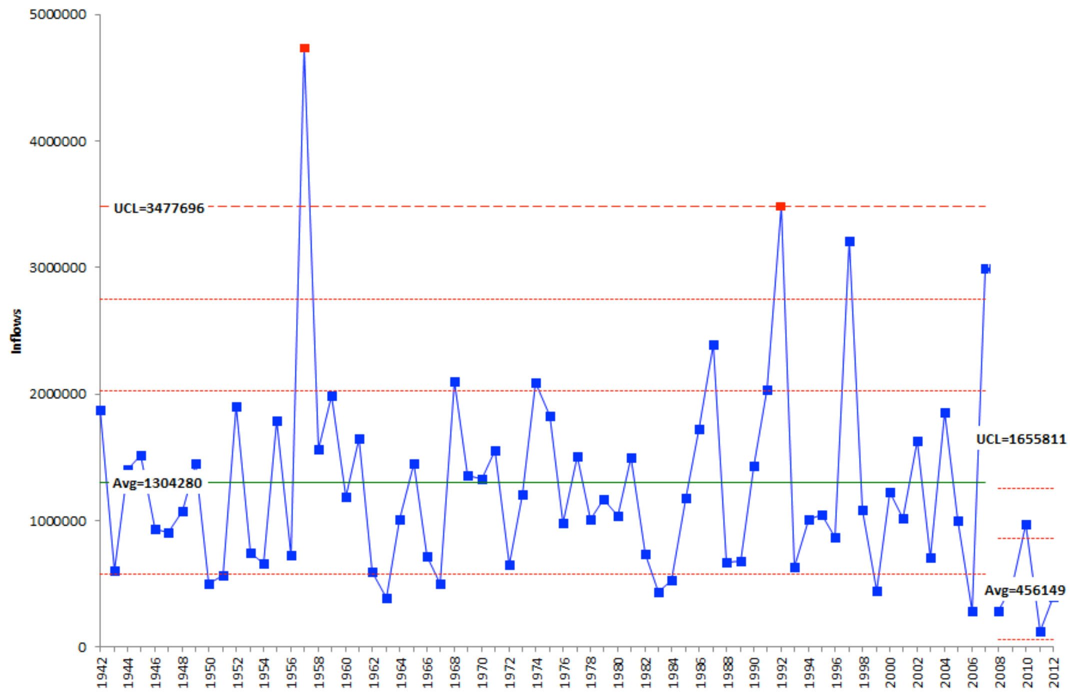
These conditions indicate that the process has indeed changed. The inflow data are no longer “consistent” over the years. There is statistical evidence that the process shifted downward starting in 2006. Figure8 shows in the X-mR control chart for the raw data. Note that this chart shows that the process is out of control in the last few years as well – although it indicates the process changed in 2008.

Figure 8: Control Chart Based on Raw Inflow Data



In conclusion, the control chart approach shows that there has been a shift in the inflows to the lakes over the past few years. The average inflow has decreased significantly as indicated by the control charts. The control chart shows the change – not the reason(s) for the change. So, what is the new process? Figure 8 shows that the shift occurred in 2008. Figure 9 shows the control chart with the chart split in 2008. The average inflow from 1942 to 2007 was 1,304,361. The average inflow in the new process is 456,149.

Figure 9: Split Control Chart to Estimate New Inflows



Appendix A: Raw Inflow Data

Year	Yearly Inflow		Year	Yearly Inflow		Year	Yearly Inflow
1942	1869518		1966	713993.3		1990	1435134
1943	602151.5		1967	503571.6		1991	2035664
1944	1404905		1968	2098909		1992	3482690
1945	1512085		1969	1357418		1993	629759.4
1946	935782.5		1970	1324530		1994	1006258
1947	908242.3		1971	1551080		1995	1047405
1948	1072715		1972	647921		1996	870509.6
1949	1455462		1973	1209395		1997	3212723
1950	501926.3		1974	2091525		1998	1088462
1951	570254.8		1975	1829737		1999	448161.5
1952	1897714		1976	983240.6		2000	1227130
1953	746945.6		1977	1502871		2001	1021712
1954	661556.9		1978	1004812		2002	1630324
1955	1789597		1979	1169847		2003	708077.3
1956	729079.9		1980	1036941		2004	1859272
1957	4732816		1981	1495960		2005	999540.7
1958	1566071		1982	734603.9		2006	285229.1
1959	1991513		1983	433312.1		2007	2996572
1960	1188341		1984	529697.5		2008	284462.3
1961	1645561		1985	1181458		2009	499731.7
1962	598867.6		1986	1723391		2010	975321.8
1963	392589.4		1987	2389690		2011	127802
1964	1007825		1988	667395		2012	393426
1965	1452809		1989	682212.6			