Analysis of Highland Lakes Inflows Using Process Behavior Charts

Dr. William McNeese, Ph.D.

Revised: Sept. 4, 2018

www.spcforexcel.com

Author's Note: This document has been revised to include the latest data for 2017. The additional data reinforces the conclusions made previously.

Summary

Data for the yearly actual water inflows and naturalized flows have been analyzed using process behavior charts. One purpose of a process behavior chart is to determine if a process has changed. This is done by plotting the data over time, calculating the average and the natural process limits using well-established statistical techniques and then interpreting the process behavior 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.

Data from 1942 – 2017 were analyzed for the yearly actual inflows and from 1942 to 2016 for the naturalized flows. Note that the data are historical. We are trying to analyze the historical data to determine how to use this historical data to make a prediction. The problem is determining which data is historically the "same" so it can be used to make a prediction about the future. That data must be homogeneous if it is to be useful in making predictions.

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 – in particular since 2008. *This decrease is statistically significant – and real – and occurs both for the actual flows and the naturalized flows.* The analysis does not address the reason(s) for the decrease – only that the decrease has occurred. The major conclusion is:

"The process changed significantly in 2008. A new process now exists that covers the period from 2008 to 2017. Only data from this period should be used to predict the process in the future. The data from 2007 and before are no longer valid to make a prediction about the future."

The details of the analysis are given below.

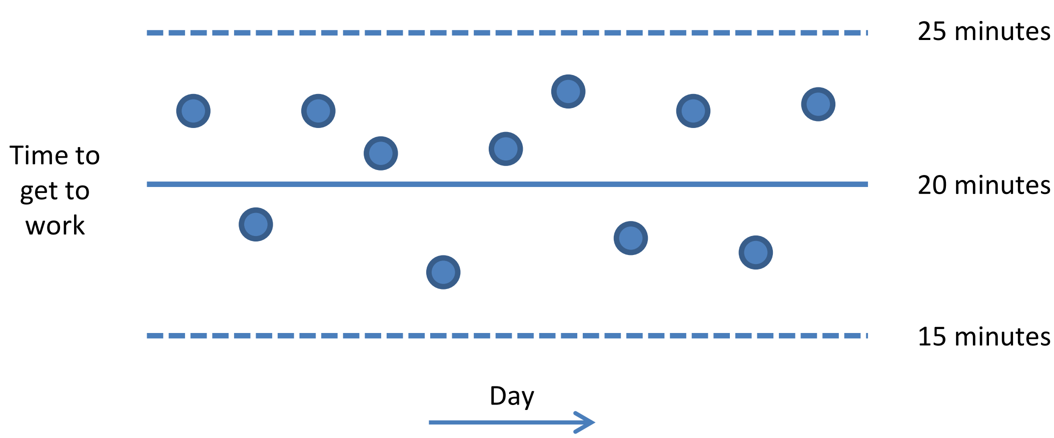
Analysis Methodology

Process behavior charts were used to analyze the data. One purpose of process behavior 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 process behavior 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.

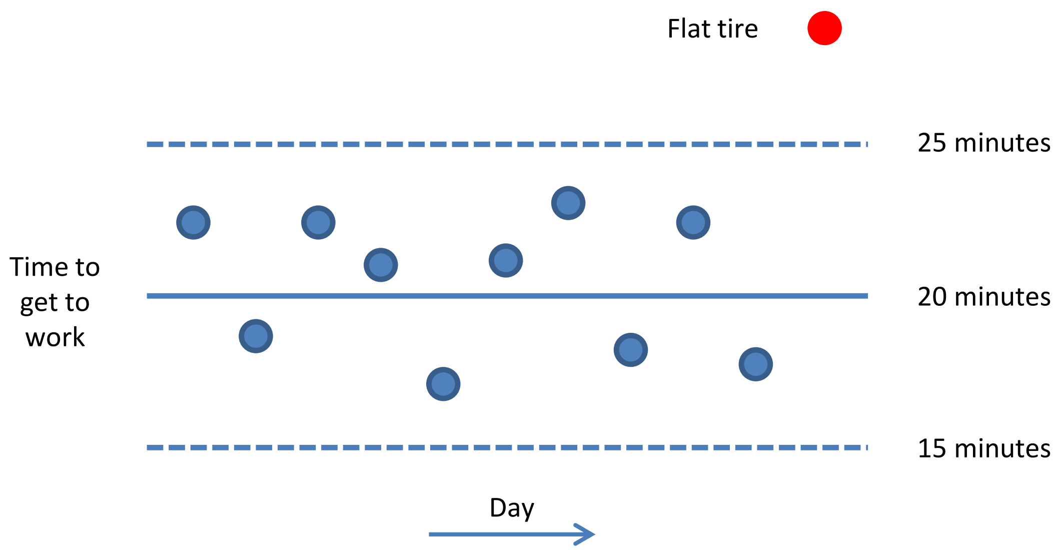
Figure 1 is an example of a process behavior chart for your process of getting to work. The data are plotted over time. The top dotted line is called the upper natural process limit. It is the largest value you would expect from the process with just common causes (normal variation) present. The lower dotted line is called the lower natural process 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 natural process 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: Process Behavior Chart for the Time to Get to Work



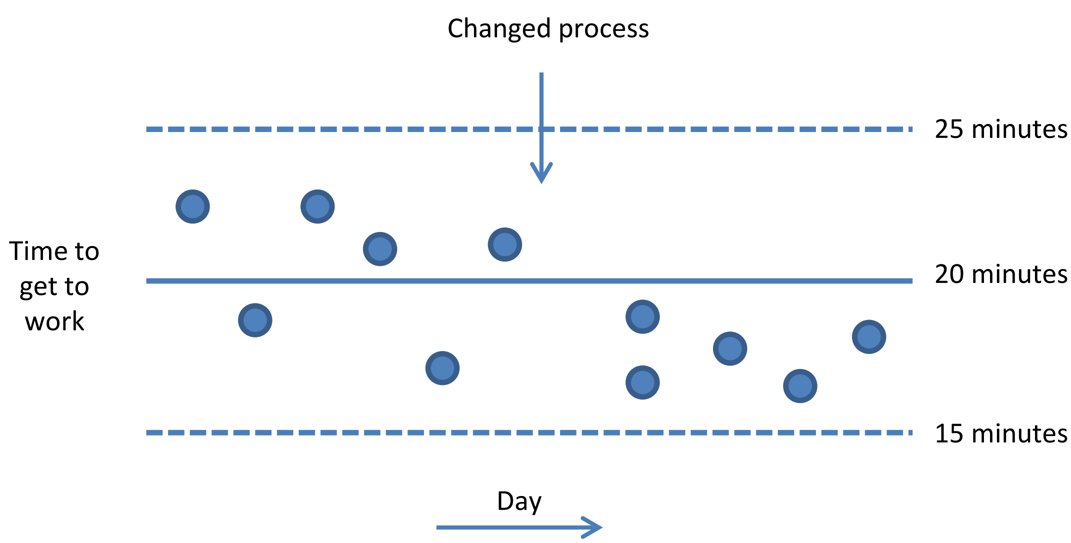
Now, suppose you have a flat tire when driving to work. How long will it take you to get to work? Definitely longer than the 15 to 25 minutes in your "normal" variation. Maybe it takes you 50 minutes to get to work. This is a special cause of variation. Something happened that has caused a change. It 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: Process behavior 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 process behavior chart will show the impact of that change. Figure 3 shows such a process.

Figure 3: Process behavior chart with Changed Process



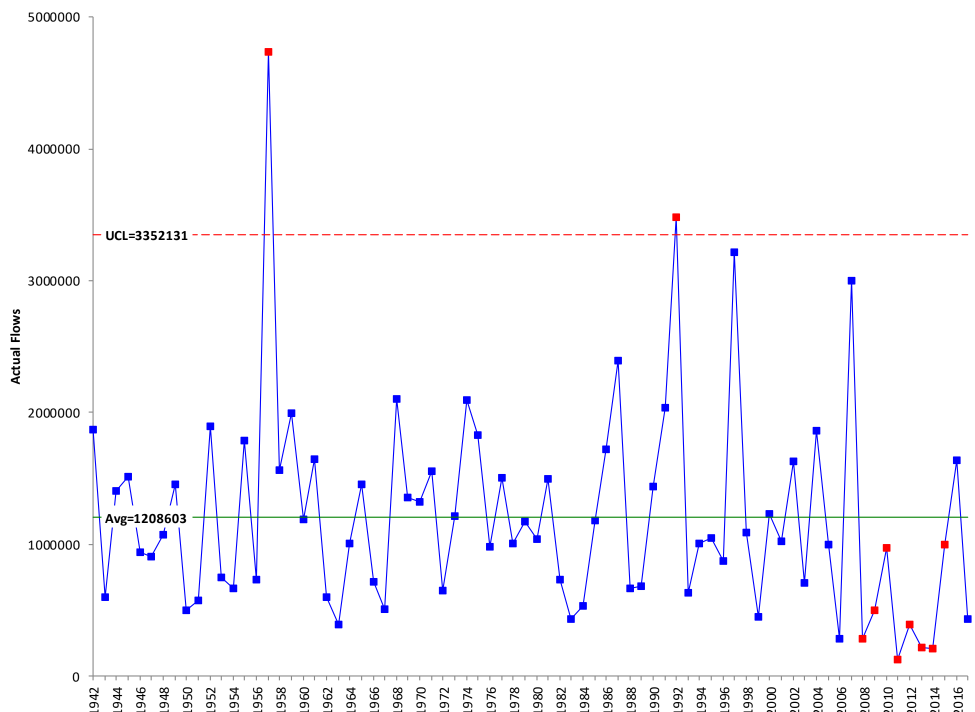
This type of shift will show up in a process behavior chart as a special cause of variation. It is clear from the chart that the average has shifted downward. The new average for the process of getting to work can be estimated from the data.

This is the type of shift that occurred with the inflow data from the lakes. The data prior to the process shift are no longer valid – that data does not represent the process anymore. Only the data since the process shift represents the process now!

Inflow Data Analysis

The data used in the analysis are shown in Appendix A and Appendix B. The process behavior chart used in the analysis was an individuals (X-mR) process behavior chart. Figure 4 is the process behavior chart for the actual inflows. There are two points beyond the upper natural process limit, the last occurring in 1992. But overall, the chart shows random variation until 2008. Starting in 2008, there are 8 points in a row below the average. This is an indication that the process has changes. In fact, 9 out of the last 10 points are below the average.

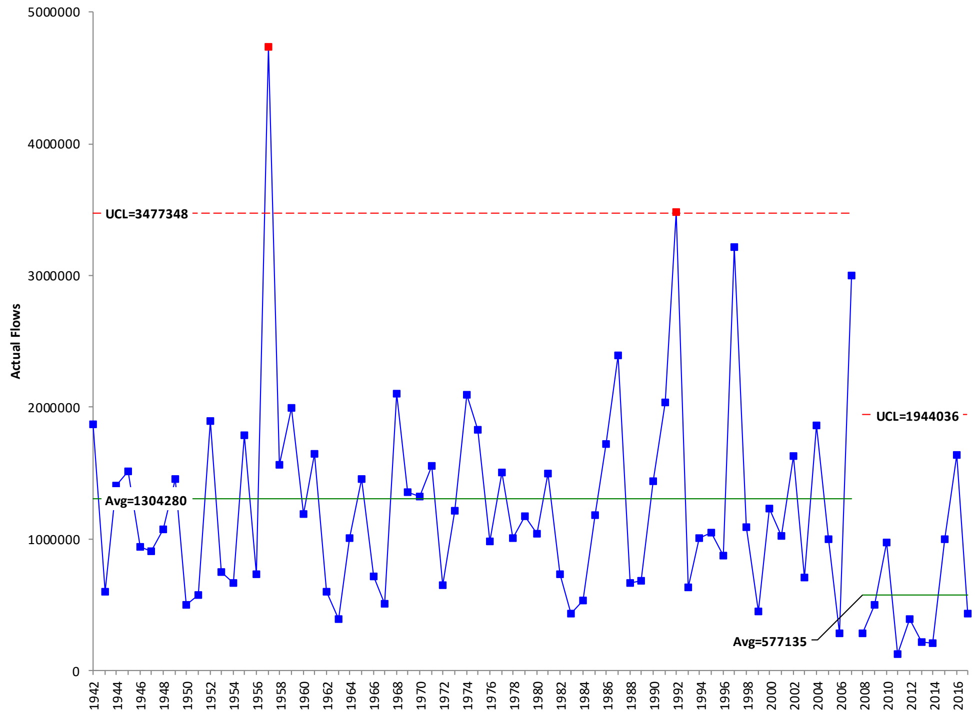
Figure 4: Process Behavior Chart for Actual Inflow Data



This condition indicates 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 2008.

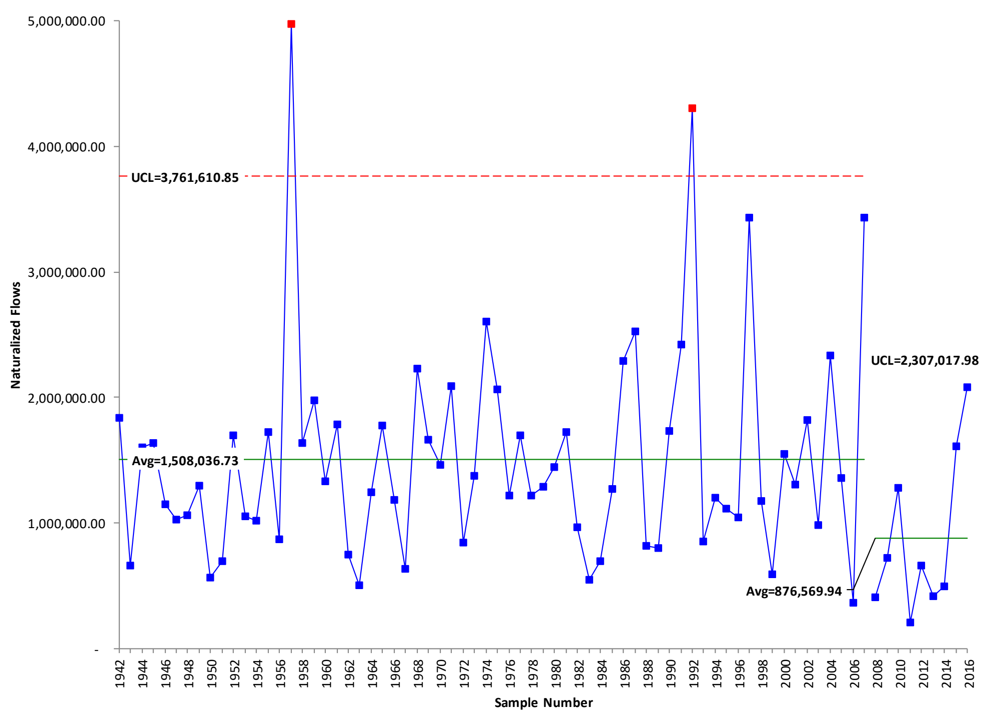
Figure 5 shows the process behavior chart with the new process starting in 2008. The new average is 571,135, down from 1,304,280. The analysis of historical data shows that the new process begins in year 2008. There is no reason to use the data prior to 2007.

Figure 5: Process Behavior Chart with New Process in 2008



There is a strong correlation between the actual flow and naturalized flows. The same pattern occurs with the naturalized flow as shown in Figure 6 and reinforces the new process begins in 2008.

Figure 6: Process Behavior Chart for Naturalized Flow



Appendix A: Raw Inflow Data

Year	Inflows		Year	Inflows		Year	Inflows
1942	1869518		1968	2098909		1994	1006258
1943	602152		1969	1357418		1995	1047405
1944	1404905		1970	1324530		1996	870510
1945	1512085		1971	1551080		1997	3212723
1946	935782		1972	647921		1998	1088462
1947	908242		1973	1209395		1999	448162
1948	1072715		1974	2091525		2000	1227130
1949	1455462		1975	1829737		2001	1021712
1950	501926		1976	983241		2002	1630324
1951	570255		1977	1502871		2003	708077
1952	1897714		1978	1004812		2004	1859272
1953	746946		1979	1169847		2005	999541
1954	661557		1980	1036941		2006	285229
1955	1789597		1981	1495960		2007	2996572
1956	729080		1982	734604		2008	284462
1957	4732816		1983	433312		2009	499732
1958	1566071		1984	529698		2010	975322
1959	1991513		1985	1181458		2011	127802
1960	1188341		1986	1723391		2012	393426
1961	1645561		1987	2389690		2013	216253
1962	598868		1988	667395		2014	207642
1963	392589		1989	682213		2015	1000054
1964	1007825		1990	1435134		2016	1636702
1965	1452809		1991	2035664		2017	429959
1966	713993		1992	3482690			
1967	503572		1993	629759			

Appendix B: Naturalized Flow Data

Year	Inflows		Year	Inflows		Year	Inflows
1942	1841450		1968	2232061		1994	1197138
1943	659076		1969	1663666		1995	1110010
1944	1605507		1970	1460815		1996	1042864
1945	1632951		1971	2088459		1997	3435896
1946	1152287		1972	843306		1998	1177695
1947	1026573		1973	1372947		1999	590008
1948	1059026		1974	2601923		2000	1547725
1949	1299346		1975	2062327		2001	1302834
1950	566910		1976	1219131		2002	1819933
1951	692098		1977	1701663		2003	985034
1952	1694065		1978	1216601		2004	2335730
1953	1053752		1979	1287824		2005	1357842
1954	1020516		1980	1449083		2006	362463
1955	1726129		1981	1720349		2007	3429934
1956	870704		1982	965262		2008	404999
1957	4972358		1983	546084		2009	717705
1958	1637498		1984	692485		2010	1282532
1959	1981323		1985	1274414		2011	210240
1960	1331240		1986	2286834		2012	659541
1961	1783773		1987	2527704		2013	419208
1962	749467		1988	814782		2014	497824
1963	499566		1989	804010		2015	1614544
1964	1246301		1990	1735715		2016	2082538
1965	1775350		1991	2418850			
1966	1183698		1992	4301203			
1967	634838		1993	852026			