CWIC
Colorado Water Issues Committee of the Texas Rice Producers Legislative Group

2511 San Bernard Drive
East Bernard, TX 77435
979-758-4670

The Purpose of CWIC:

Facilitate the availability of Colorado River water supplies for rice production in the four major irrigation operations on the Colorado River.

September 10, 2018

Dear LCRA WMP Revision Staff,

This letter is intended as further formal public comment on the WMP revision in follow-up to the September 6 meeting.

1. **Modelling Capabilities -** CWIC's modelling capabilities have only recently reached the point of being able to reproduce the results presented at the July meeting. As such CWIC is now able to model scenarios that could offer some hope of meeting our desire to reduce partial curtailments of first and second crops while remaining within the TCEQ framework.
2. **Trigger Modifications within Framework** -After considerable examination of the tabular output data from LCRA's original model run, CWIC has concluded that there are modifications to drought and curtailment triggers that would increase the reliability of both crops for us within the TCEQ framework while also having no significant effect on other critical outputs important to our fellow stakeholders. CWIC has made a preliminary model run to confirm this. CWIC is in the process of further reviewing and assessing these results and will be sharing those as soon as possible.
3. **Near-term Irrigation Demands -** CWIC maintains that for the foreseeable future there is a high likelihood that irrigation demands will come in lower than the weather-varied demands used in the modeling and even more certainly lower than the maximum allocation of 178,000 ac-ft. for first crop and the 66,000 ac-ft for second crop. By modeling higher demands than will occur, drought and curtailment triggers will be set higher than necessary. However, CWIC firmly believes that permanently reducing those modeled demands from this point forward could prove shortsighted for the future.
4. **Risk Aversion Acreage/Demand Reductions –** Irrigator risk aversion has the unfortunate result of reducing planted rice acreage below that acreage upon which projected demands and modeling efforts have been based. To be clear, the caps of 178,000 and 66,000 on first and second crop stored water usage respectively inadvertently bring about considerably lower demands resulting from lower planted acres that result from risk-aversion. In some cases, there will even be creditors dictating the acreage their customers can plant to limit their water-related lending risks.
5. **Utilization of Unused, Allocated First Crop Water –** Considering the likelihood of demands falling considerably under modelled demands as previously discussed (see comments 3, 4 and 5), CWIC is requesting

CWIC
Colorado Water Issues Committee of the Texas Rice Producers Legislative Group

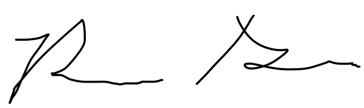
2511 San Bernard Drive
East Bernard, TX 77435
979-758-4670

that a portion of allocated-but-not-released first crop stored water be able to be carried over for use in second crop up to the maximum amount necessary to make a full second crop allocation. CWIC will be attempting this model scenario with a host of assumptions. CWIC is requesting technical assistance from LCRA that may be necessary for our modeling consultant (Joe Trungale) to accomplish this in an efficient and timely manner.

6. **Time Constraints** – CWIC is concerned that the time constraints being imposed on this revision process could lead to the submission of a less-than-finished product. While CWIC appreciates the need for such an abbreviated timetable, we wish to be on record as indicating that it may not provide enough time and opportunity for adequate consideration and analysis of salient proposals currently being developed. CWIC intends to expedite our proposals as much as possible, and does not currently foresee proposals beyond those that have been roughly laid out in these comments.

Thank you for the ongoing opportunity to comment and participate in this critical process. The livelihoods and financial security of many Texas families and businesses depend on LCRA producing the best and fairest plan possible.

Sincerely,



Ronald Gertson

Chair, Colorado Water Issues Committee of the Texas Rice Producers Legislative Group

LREWater,LLC
a Leonard Rice Engineers Company

COMMENT – LCRA WMP REVISION & ACCURACY OF NATURALIZED FLOWS

To:	John Hofmann, VP Water Operations Lower Colorado River Authority
From:	Jordan Furnans, PhD, PE, PG LRE Water, LLC
Date:	September 12, 2018

Dear LCRA,

The entire LCRA WMP revision process is based upon the notion that the planned actions resulting from WMP implementation would not have caused the Highland Lakes to drop below a combined storage of 600,000 acre-ft. To show this, you use the WAM model over the period of record 1940-2016. We all assume that the WAM model is accurately setup and has been fully vetted, is error-free, and accurately represents the period of record conditions.

In reviewing the WAM model, I studied the Excel worksheets used by Kirk Kennedy when updating the naturalized flows. I did not review them extensively, but focused on the water balance calculations involving Lake Travis. I had previously noticed that my independent calculations resulted in significantly different monthly inflows than are being used in the WAM, and I wanted to discover why. Upon reviewing the spreadsheets from Kirk Kennedy (obtained from Dr. Kathy Alexander of TCEQ), I found that the end-of-month lake storage is based on the lake level measured at midnight on the last day of the month (i.e. 00:00 hrs on say 10/31). In my opinion, this is incorrect as the month should end at midnight on the first day of the next month (i.e. 00:00 hrs on 11/1). In the WAM currently used in the WMP process, the monthly inflows from October are determined based on the time period from 00:00 hrs on 9/30 through 00:00 on 10/31. Thus the timeframe for the calculation is shifted one day backward from where I believe it should be.

This may or may not have significant consequences for regarding the WAM modeling results, especially with regard to the timing of when certain diversion actions or limits are triggered in the model. For example, on 10/31/2015, approximately 18,000 acre-ft of inflow entered Lake Travis. In the current WAM, this inflow is included in the naturalized flows for NOVEMBER, even though they actually occurred during OCTOBER. I am concerned that the naturalized flows being used in this WAM modeling effort for the WMP are not properly reflecting the timing of when historical flows actually occurred.

The impact of the "end-of-month" date discrepancy on WAM calculations cannot be known until the naturalized flows are adjusted and re-run in the WAM. I suspect that the impact will be to shift when water is available for storage in the lakes and for interruptible releases, and this could change the parameters of the WMP modeling, as well as the model results. It is possible that this change could cause periodic increases or decreases in interruptible releases, and in environmental flow criteria attainment.

As a result of the brief investigation I made into the naturalized flow calculations, I found what may (or may not) be an important methodological error which calls into question the validity of the WMP WAM modeling results presented so far. I have to wonder whether there are other "errors" incorporated into the naturalized flow calculations that would lead to different model results if the errors were identified and corrected. I request that LCRA provide documentation on the QA/QC measures undertaken when developing the WAM hydrology, including describing how the resulting naturalized flow numbers were reviewed and approved by TCEQ.

Until the naturalized flow numbers used in the WAM are fully vetted and accepted as the best, most accurate representation of historical natural flow conditions, any results derived from WAM modeling must be considered suspect. I do not believe further revisions to the WMP can be considered without first verifying the accuracy of the naturalized flows used as the driving factor within the WAM modeling used to justify WMP provisions.

I notified both Kirk Kennedy (Kennedy Resource Company) and Dr. Kathy Alexander (TCEQ) of my discovery and concerns as detailed in this comment. They have yet to respond to the email I sent them earlier today.

Sincerely,

Jordan Furnans

1101 Satellite View | Suite #301 | Round Rock, Texas 78665 | 512-736-6485
Jordan.Furnans@LREWater.com

Subject: Comment about WMP Revision schedule and request for additional meeting

*****From an external source – Think before you click. *****

This comment is submitted on behalf of the National Wildlife Federation, Texas Parks and Wildlife Department, the Texas Living Waters Project, and Environmental Stewardship. As mentioned during the last Participant Meeting, the currently envisioned schedule does not appear to allow for adequate time for informed comment on modeling results. For this current round of WMP revisions, the modeling necessarily has become even more complex than for previous versions because of the need to incorporate, for purposes of predicting flows, the impacts of upstream diversions under the amended Garwood water right and of the operation of the Arbuckle Reservoir. In addition, changes to the underlying WRAP code have added to the workload and delayed the availability of modeling results.

As a result of those factors and others, there simply does not appear to be sufficient time left in the proposed schedule, as we understand it, to allow for informed discussion of the implications of proposed revisions of the WMP, as reflected in modeling. Proposed changes that the above-referenced organizations (conservation representatives) have submitted for consideration have not yet been modeled, or at least modeling results are not available. In addition, as noted at the most recent meeting, the irrigator representatives will be submitting proposed changes that also likely will need to be incorporated into modeling. The most recently announced schedule contemplates only two additional Participant Meetings (Oct. 4 and a second meeting date TBD).

We are concerned that the two currently proposed meetings will not allow sufficient time for informed discussion of modeling results. It seems highly likely that the next meeting will be focused on an initial review of modeling results reflecting the implications of the changes requested by the conservation representatives and the irrigation representatives, and any others that may be submitted in the same timeframe, and any variations on those changes developed by LCRA staff. History demonstrates that, because of the complexity of the interacting factors, participants need time after modeling results are presented and discussed to digest and understand the implications of those changes. Again, based on past experience, the likely result will be that, after the next Participant Meeting, variations on the modeled scenarios will be recommended and that at least some of those variations, with some fine-tuning, may improve the overall balancing of interests. However, the modeling results for those variations will only be available for discussion at the last of the currently scheduled Participant Meetings, resulting in the loss of any opportunity to fine-tune scenarios based on participant input and discussion.

We very much appreciate the willingness of LCRA staff to adjust the process so far to facilitate more informed participation and respectfully request that an additional Participant Meeting be added to the schedule because of the potential for an improved outcome and an improved understanding of the implications of the proposed revisions.

Myron J. Hess
Tributary Consulting LLC
1705 Margaret Street
Austin, TX 78704
Myron@myronhess.com
Ph: 512-576-3948

Sent: Thursday, September 13, 2018 5:50 PM
To: LCRAWMP
Subject: Submittal for 2019 WMP

Since Arbuckle should pick up a very high percentage of ordered but not diverted water (OND) and the overwhelming majority of water from Arbuckle will be used for interruptible Ag use, I recommend OND water whether it be from storage or storable inflows (ROR above Mansfield) should be included in the total allocation of stored water allowed for that crop season as to not artificially increase the amount of interruptible water available.

COMMENTS OF THE CENTRAL TEXAS WATER COALITION RELATING TO LCRA'S PROPOSED UPDATES TO ITS 2015 WATER MANAGEMENT PLAN

SUBMITTED VIA EMAIL TO LCRAWMP@lcra.org
September 13, 2018

The Central Texas Water Coalition (CTWC) appreciates the continuing opportunity to submit comments, questions, and items for discussion with respect to LCRA's ongoing efforts to develop an updated Water Management Plan (WMP) for the operation of Lakes Buchanan and Travis. These comments include responses to matters raised during or after the most recent LCRA-hosted informational meeting on September 6, 2018.

Proposed Timeline of Activities. As several attendees noted in the WMP Participant Meeting on September 6th, the LCRA's presentation of information on the Firm Yield modeling and calculations is a very important aspect of the WMP update, and we join the others in requesting information on Firm Yield at the October 4 Participant Meeting (rather than at a November meeting). Under LCRA's timeline, if the proposed new WMP goes before the LCRA Board in December, there will be little time to evaluate the WMP if the key components are only provided a few weeks before the Board meeting. We are hopeful that these Firm Yield calculations will be revealed as soon as possible.

Determination of Naturalized Flows. We would appreciate further explanation of how LCRA calculates "naturalized flows" and how those calculations utilize stream gauges to make those calculations. More specifically, there seems to be a large deviation between LCRA's reported gauged inflows for 2015 and 2016 and the naturalized flows that are being used in the latest WMP modeling. Please explain the process for translating gauged inflow data into naturalized flows, including the major factors contributing to the differences between the naturalized and gauged flows. In addition, please explain how LCRA accounts for evaporation from the Highland Lakes in its development of naturalized flows and explain whether lake evaporation is one of the factors causing the deviations between the gauged and naturalized flows.

Developing a WMP that Assures Satisfaction of Firm Demands. As we have discussed, recent hydrology and the statistical evaluation of the historical data have demonstrated that the inflows to the Highland Lakes have declined so much that there is a "new normal." A noted statistical expert has determined that this shift in data is so substantial that inflow data prior to and including 2007 are no longer useful or valid as predictors for the future.[1] Instead, data from 2008 to the present is the most relevant and valuable dataset. Another example of the use of this statistical analysis to examine a natural system is presented in the work of Dr. Don Wheeler, who applied these methods to better understand major floods. Copies of these references are attached.

[1] Conversation with Dr. William McNeese on Sept. 4, 2018.

By including the 2008 to 2015 time period as the Drought of Record in its WMP modeling of Firm Yield, LCRA is acknowledging some aspects of this shift to a new normal. The obvious declines in stormwater runoff to the lakes, apparently associated with changes in land use, warmer temperatures, extended drought periods, the pumping of alluvial wells, the depletion of groundwater resources, and the proliferation of small upstream ponds, are all factors that justify more cautious decision making in the development of the next WMP. We should not gamble on the ability of the Highland Lakes to refill or to quickly regain storage volumes when the drinking water supplies for millions of residents is at stake. CTWC urges LCRA to choose the most protective and conservative options throughout its development of the new WMP.

<u>Addressing Released but Unused Orders for Interruptible Stored Water</u>. CTWC believes that the financial value of the stored water released for LCRA's Interruptible customers should be recovered by LCRA, and LCRA's determinations of available stored water for use by Interruptible customers should not identify that "ordered but not diverted" water as a new supply merely because it has moved from the Highland Lakes to the Arbuckle Reservoir. LCRA should assess fees for the water released for downstream irrigation customers, and LCRA should benefit from the release of stored water for Interruptible customers, whether or not the downstream customer diverts the ordered water. It seems appropriate to assess a fee to those customers who choose not to take the water released for them (similar to LCRA's assessment of reservation fees on its Firm customers) AND to recalculate the allocation of stored water available for such customers by deducting the volumes of water that were released but not diverted by that customer. Please explain how Interruptible stored water that is released but not diverted by a downstream Interruptible customer will bring revenue to LCRA. Also, please explain how LCRA plans to quantify such water for purposes of meeting but not exceeding the Interruptible water allocations included in the WMP.

<u>Modeling to Include the Operations of the New Arbuckle Reservoir</u>. In view of the likelihood that water stored in the Arbuckle Reservoir will be used to capture stored water from Lakes Buchanan and Travis that was ordered by an Interruptible customer **but not used**, CTWC suggests the development of a water accounting system that includes the volumes of water that were released from storage and captured in the downstream Arbuckle Reservoir as part of the total volume of Interruptible Stored water available for LCRA's Interruptible customers. In other words, water stored and released from the Highland Lakes and then re-captured in the Arbuckle Reservoir would continue to be recognized as "Highland Lakes stored water" for potential purchase by Interruptible customers, but the water would not be counted as a new supply upon its arrival in the Arbuckle Reservoir. Water that is released from Lakes Buchanan or Travis **but not used** to meet an Interruptible customer's demands would be counted against the total amount of stored water that is allocated for that customer for that season.

<u>Clarifications on Water Used in Hydroelectric Generation Operations</u>. The adjudication decree and LCRA's water rights contain specific provisions regarding the use of water in hydroelectric generation activities. More specifically, LCRA's rights to use water for hydroelectric generation "should include conditions that generally subordinate such rights to all present and future upstream rights to use the waters of the Colorado River and its tributaries for municipal, domestic, irrigation or industrial purposes." These conditions "should specifically prohibit the release of water through its dams solely for the purpose of hydroelectric generation, except

during emergency shortages of electricity, and during other times to the extent that such releases will not impair LCRA's ability to satisfy all existing and projected demands for water from Lakes Travis and Buchanan … pursuant to all firm, uninterruptible commitments and all non-firm, interruptible commitments." [2]

The existing WMP includes a discussion on Hydroelectric Power Generation in the chapter on Demands – in Section 2.5. This discussion indicates that LCRA employs a definition of "emergency shortage of electricity" that allows LCRA to release water for hydroelectric generation absent a downstream water demand in four situations. One listed way is Responsive Reserve Service (RRS), which is not recognized by ERCOT as "emergency", but rather an "Ancillary Service". Furthermore, RRS releases by LCRA are authorized ahead of time, because LCRA makes this capacity available in ERCOT's Day Ahead Market. (Please explain how these situations satisfy the criteria in LCRA's water rights regarding the limitations on releases of stored water except in the case of emergencies.

Recent Releases from Storage. It appears that LCRA plans to continue its current daily operating philosophy in the next version of the WMP. This is a concern, since this operating philosophy seems to allow massive discharges of stored water from the lakes even when the entire river basin is experiencing significant rains. Please explain the basis for this philosophy and for the recent decision to release water from lake storage rather than use it to refill Lakes Buchanan and Travis.

Accounting for Emergency Hydroelectric Releases. In a prior comment, CTWC asked how LCRA accounts for potential Emergency Hydroelectric Releases in its water modeling. In an online response to this question, LCRA stated that "Emergency hydroelectric releases are factored into the model as part of a simulated demand that represents conveyance adjustments and other releases amounting to an average of about 30 cubic feet per second on a daily basis." (Note: 30 cfs daily average represents about 21,700 acre feet per year.) CTWC does not believe in having two sets of units in the WMP. Acre-feet is the preferred unit of capacity, and acre-feet / year the preferred unit for rate.

Releases made for emergency hydroelectric generation are reported in LCRA's annual water use summaries. "Over the past five years, releases have averaged 164 acre-feet per year. The maximum release in one year was 490 acre-feet in 2014." Please explain the meaning of "conveyance adjustments and other releases" in the response. Also explain where the daily 30 cfs release (or 21,700 acre-feet per year) appears in the water modeling. Since a 30 cfs daily flow is a significant quantity (21,700 acre feet per year), we would like to understand where it is shown and accounted for within LCRA's water modeling and water availability calculations.

Attachments: "Analysis of Highland Lakes Inflows" by Dr. William McNeese (September 2018)
"Why We Keep Having 100-Year Floods" by Dr. Don Wheeler (June 2013)

---

[2] Attachment No. 2 to Final Judgment and Decree; Adjudication of Water Rights in the Lower Colorado River; April 1988

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.

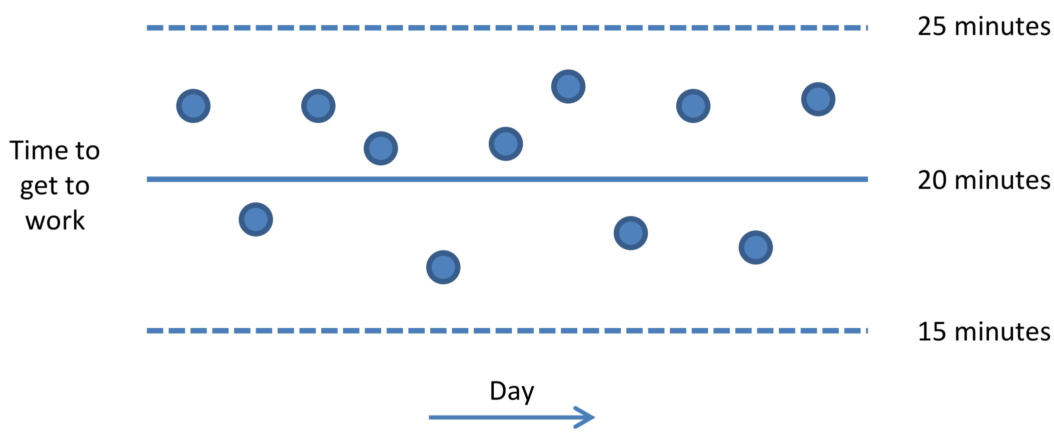
Analysis Methodology

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.

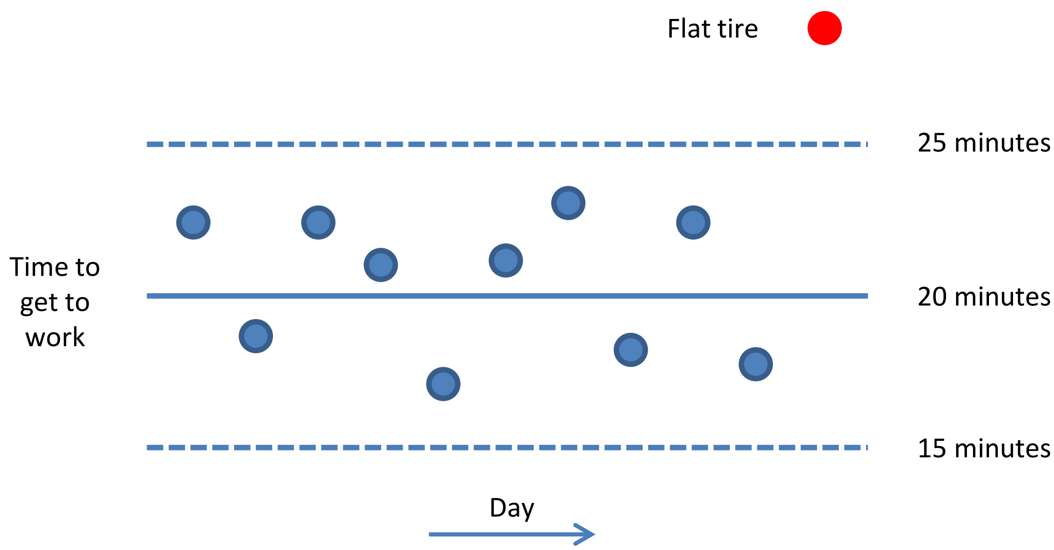
Figure 1 is an example of a process behavior chart for your process of getting to work. The data are plotted over time. The top dotted line is called the upper natural process limit. It is the largest value you would expect from the process with just common causes (normal variation) present. The lower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



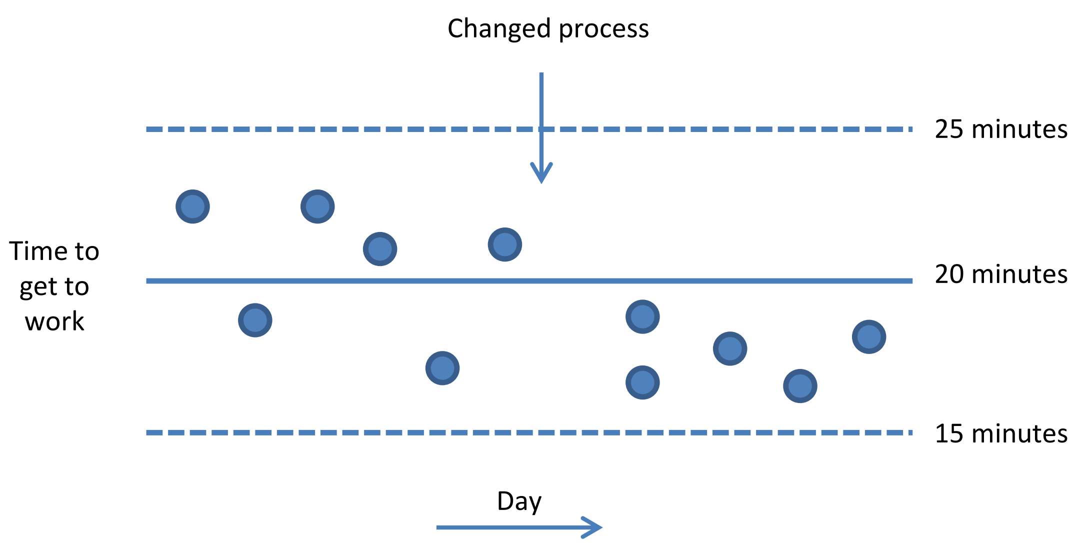
Now, suppose you have a flat tire when driving to work. How long will it take you to get to work? Definitely longer than the 15 to 25 minutes in your "normal" variation. Maybe it takes you 50 minutes to get to work. This is a special cause of variation. Something happened that has caused a change. It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



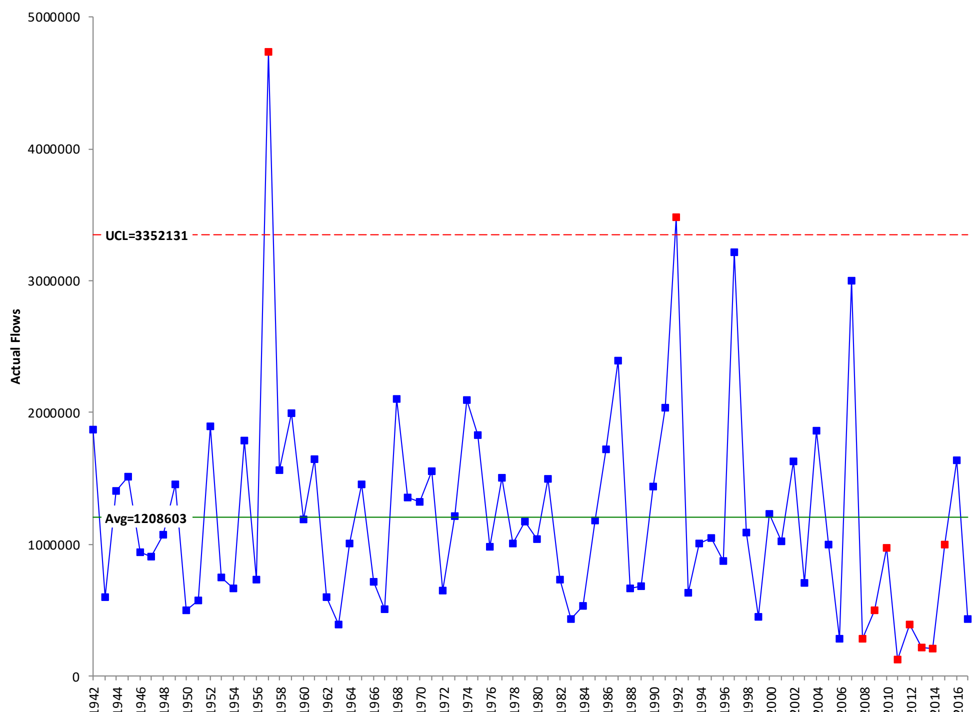
This type of shift will show up in a process behavior chart as a special cause of variation. It is clear from the chart that the average has shifted downward. The new average for the process of getting to work can be estimated from the data.

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.

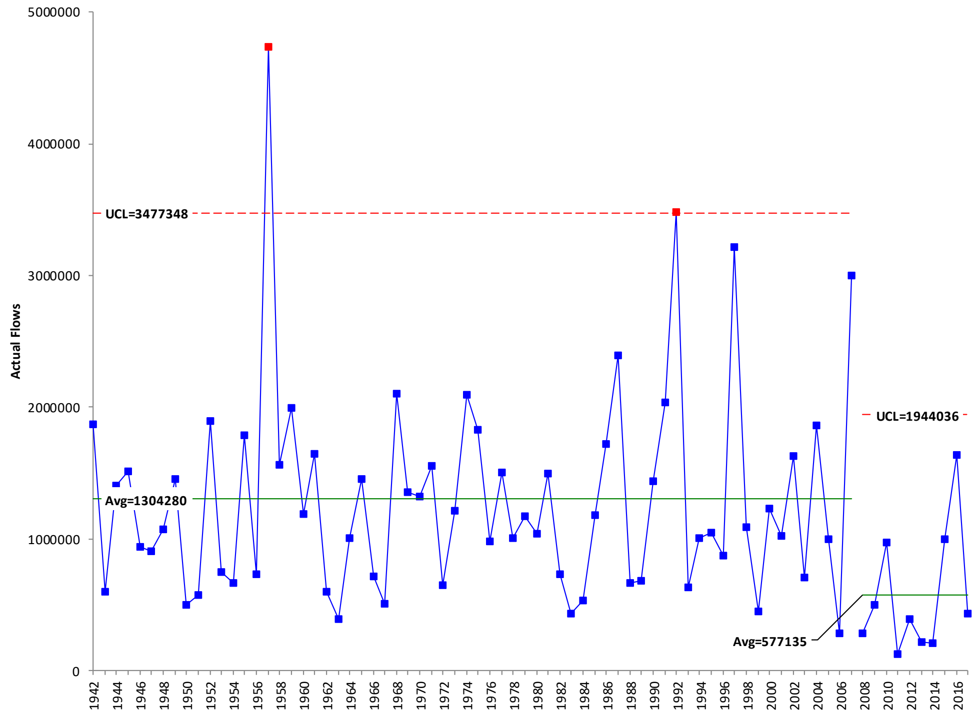
Figure 4: Process Behavior Chart for Actual Inflow Data



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.

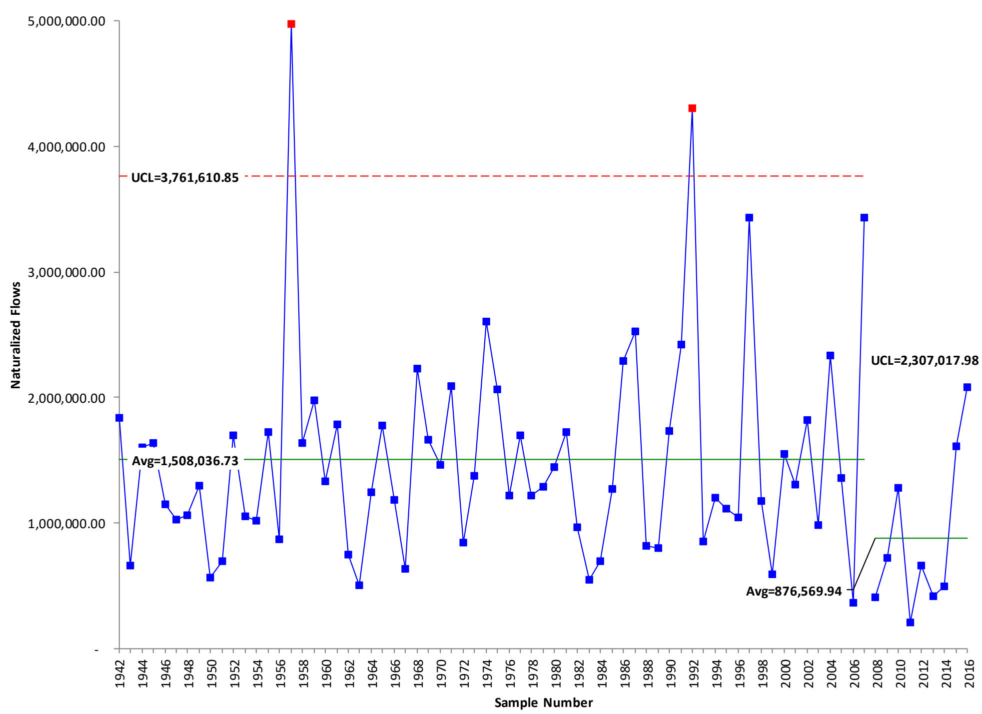
Figure 5 shows the process behavior chart with the new process starting in 2008. The new average is 571,135, down from 1,304,280. The analysis of historical data shows that the new process begins in year 2008. There is no reason to use the data prior to 2007.

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			

QUALITYDIGEST
Published on *Quality Digest* (http://www.qualitydigest.com)

Why We Keep Having 100-Year Floods

Making predictions using historical data

Donald J. Wheeler

Published: 06/04/2013

All data are historical. All analyses of data are historical. Yet all of the interesting questions about our data have to do with using the past to predict the future. In this article I shall look at the way this is commonly done and examine the assumptions behind this approach.

A few years ago a correspondent sent me the data for the number of major North Atlantic hurricanes for a 65-year period. (Major hurricanes are those that make it up to category 3 or larger.) I have updated this data set to include the number of major hurricanes through 2012. The counts of these major hurricanes are shown in a histogram in figure 1.

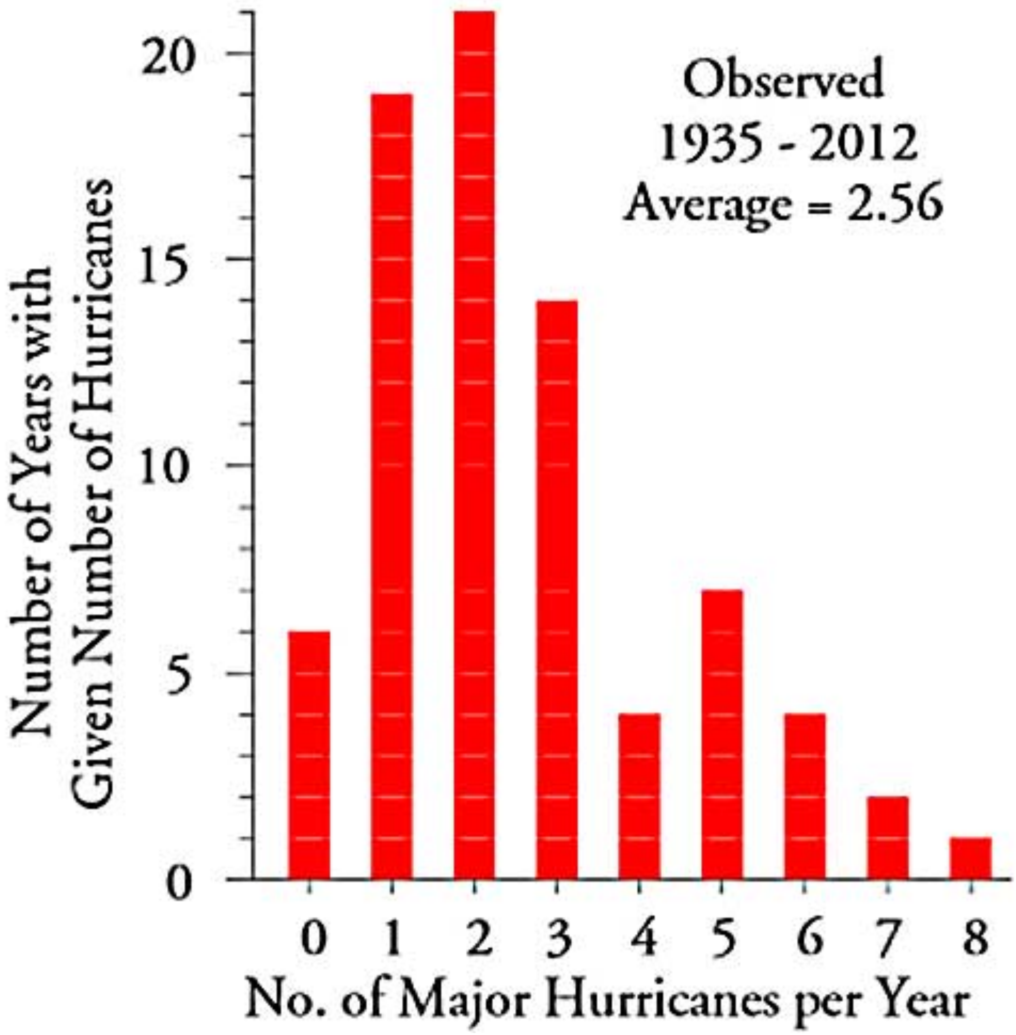


Figure 1: Histogram for the number of major North Atlantic hurricanes 1935–2012

A Poisson probability distribution is commonly used to model the number of hurricanes. Since the Poisson model only has one parameter, we only need to compute the average number of hurricanes per year to fit a Poisson distribution to the histogram of figure 1. The histogram shows an average of 2.56 major hurricanes per year. A Poisson distribution having an average of 2.56 is shown in figure 2.

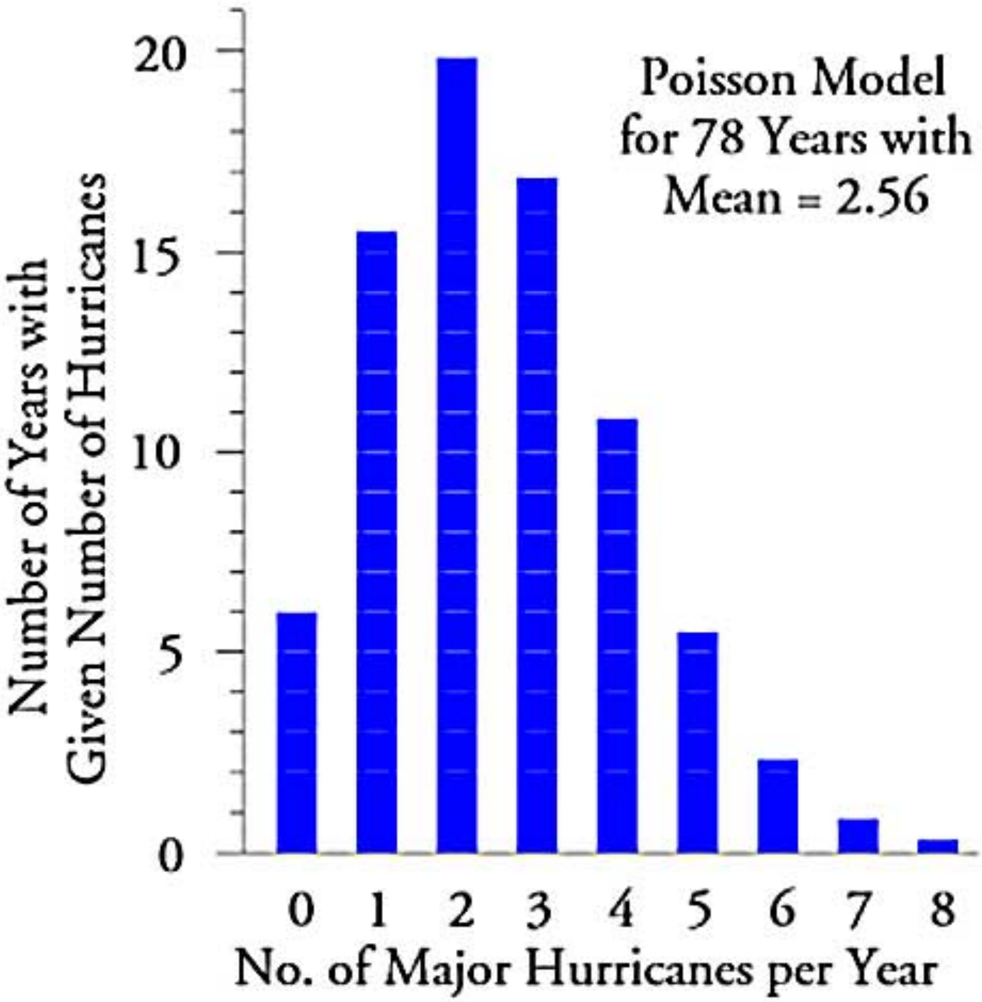


Figure 2: The Poisson probability model with mean of 2.56

Using any one of several different test procedures, we can check for a lack-of-fit between the model in figure 2 and the histogram in figure 1. Skipping over the details, the result of these tests is that there is “no detectable lack of fit between the histogram and a Poisson model with a mean of 2.56.” Thus, it would seem to be reasonable to use the model in figure 2 to make estimates and predictions about the frequency of major hurricanes.

For example, we might decide to obtain an interval estimate for the average number of major hurricanes per year. In this case, an approximate 95-percent interval estimate for the mean number of major hurricanes per year is 2.2 to 3.0.

Also, based on the model in figure 2, the probability of getting seven or more major hurricanes in any single year is 1.59 percent. This means that in any 78-year period we should expect to get about 78 × 0.0159 = 1.24 years with seven or more major hurricanes. Putting it another way, dividing 78 by 1.24, we find that a year with seven or more hurricanes should happen only about once every 63 years on the average. However, a glance at figure 1 shows that between 1935 and 2012 there have been three years with seven or eight major hurricanes. This is once every 26 years!

Well, perhaps we were simply too far out in the tail. Let’s look at how many years should have six or more major hurricanes. The Poisson model in figure 2 gives the probability of getting six or more major hurricanes in any single year as 4.61 percent. This means that in any 78-year period, we should expect to find about 78 × 0.0461 = 3.60 years with six or more major hurricanes. If we divide 78 by 3.60, we find that the model suggests we should have one year with six or more major hurricanes every 22 years on the average.

In contrast to this, figure 1 shows that there have been seven years between 1935 and 2012 with six or more major hurricanes. Dividing 78 by 7, we find that we have actually had one year with six or more major hurricanes every 11 years on the average. Thus, we are having extreme years at twice the predicted frequency. Sound familiar?

So what is happening? Why do these data keep misbehaving?

The first clue as to what is happening with these data can be found by simply plotting the data in a running record. When we do this, we immediately see that there are two different patterns of hurricane activity. In the periods before 1948 and from 1970 to 1994, there was an average of 1.49 major hurricanes per year. In the periods from 1948 to 1969 and after 1995, there was an average of 3.55 major hurricanes per year. During the periods of low activity, there could be up to four major hurricanes in a single year. During the periods of high activity, there could be up to 10 major hurricanes in a single year. These differences are reflected in the central lines and limits shown on the X chart in figure 3.

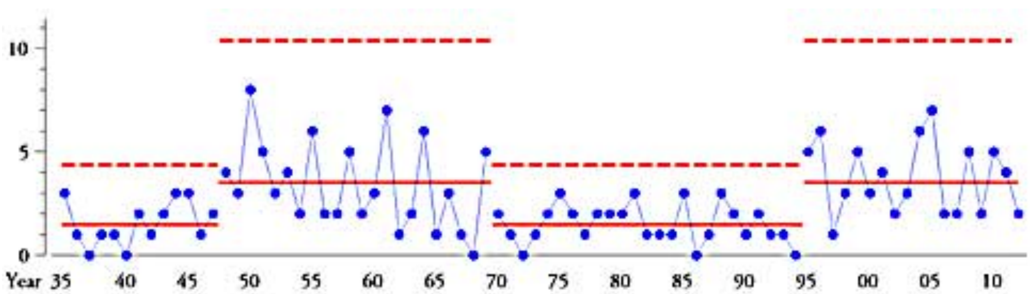


Figure 3: X chart for number of major North Atlantic hurricanes, 1935–2012. Click here for larger image.

To investigate this change in the weather patterns, I went to the NOAA website and found an article titled "NOAA Attributes Recent Increase in Hurricane Activity to Naturally Occurring Multi-Decadal Climate Variability" (Story 184). This article confirms the existence of the cycles shown in figure 3 and offers an explanation for this phenomenon. The histograms for these two levels of activity are shown in figure 4.

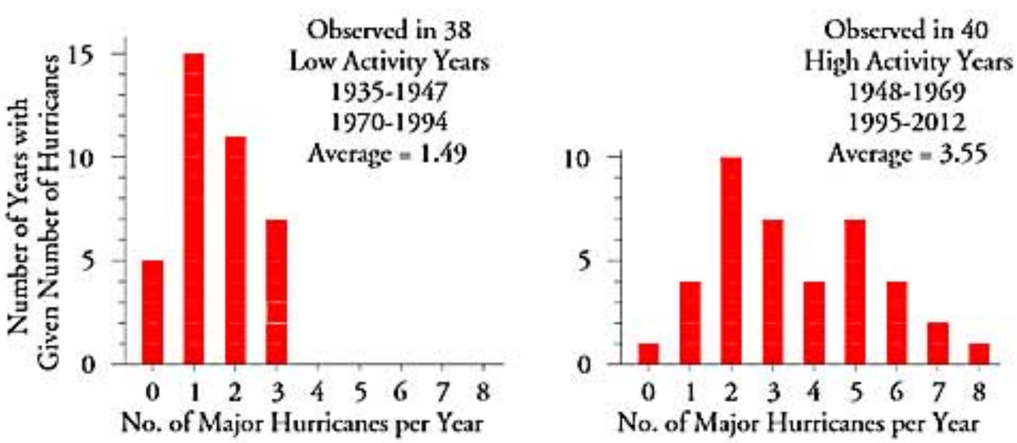


Figure 4: Histograms for the numbers of major North Atlantic hurricanes, 1935–2012

So what is wrong with the probability model used in figure 2? The problem is not with the choice of the model or with the mathematics, but rather with the assumption that the data were homogeneous. Anytime we compute a summary statistic, or fit a probability model, or do just about anything else in statistics, there is an implicit assumption that the data, on some level, are homogeneous. If this assumption of homogeneity is incorrect, then all of our computations, and all of our conclusions, are questionable. No single year between 1935 and 2012 is characterized by an average of 2.56 major hurricanes per year. Our 95-percent interval estimate of 2.2 to 3.0 major hurricanes per year simply does not apply to any year or any period of years. And our predictions as to how often certain numbers of major hurricanes will occur were completely off-base.

Yet this is what people are taught to do. The software makes it so easy to check for a lack of fit between a histogram and a probability model that many students have been taught elaborate systems for selecting and verifying a probability model as the first step in data analysis.

We collect data regarding all sorts of disasters: in addition to hurricanes, there are databases for tornadoes, floods, earthquakes, wildfires, etc. In each case we collect the observations into a single histogram, and use these observations to fit a probability model. It should be noted that when we do this, we are fitting the broad middle portion of the probability model to our data. Of course, this broad middle portion includes all of the commonly occurring events. If we fail to find a detectable lack of fit between our model and the observations, we will decide that we have the right model. Next we will then leave the broad middle of our probability model and move out into the extreme tails to make predictions about the frequency with which the more extreme events will occur: "According to the infinitesimal areas under the extreme tail of our assumed probability model, a flood of this magnitude will occur, on the average, only once in 100 years."

Of course, the model used for such a prediction is built on the assumption that the conditions that resulted in the more common events are the same conditions that result in the more extreme events. Since this assumption that the cause system remains the same is rarely correct, the data are rarely homogeneous. When the data are not homogeneous, the fitted model will be incorrect. When the fitted model is incorrect, our predictions will be erroneous. And when our predictions are erroneous, we end up having a 100-year flood about every 10 years or so.

This is why the primary question of statistical analysis is not, "What are the descriptive statistics?" Neither is it, "What probability model shall we use?" nor is it, "How can we estimate the parameters for a probability model?" The primary question of all data analysis is whether the data are reasonably homogeneous. When the data are not homogeneous, the whole statistical house of cards collapses. Hence, Shewhart's Second Rule for the presentation of data: "Whenever an average, range, or histogram is used to summarize the data, the summary should not mislead the user into taking any action that the user would not take if the data were presented as a time series."

The primary technique for examining a collection of data for homogeneity is the process behavior chart. Any analysis of observational data that does not begin with a process behavior chart is fundamentally flawed.

About The Author

Donald J. Wheeler

Dr. Donald J. Wheeler is a Fellow of both the American Statistical Association and the American Society for Quality, and is the recipient of the 2010 Deming Medal. As the author of 25 books and hundreds of articles, he is one of the leading authorities on statistical process control and applied data analysis. Contact him at www.spcpress.com.

Source URL (retrieved on *06/20/2013*): http://www.qualitydigest.com/inside/quality-insider-column/why-we-keep-having-100-year-floods.html

Links:
[1]
http://www.qualitydigest.com/IQedit/Images/Articles_and_Columns/2013/June_2013/Wheeler_June/3Lrg.jpg
[2] http://www.magazine.noaa.gov/stories/mag184.htm