Weather Normalization: Data Analysis, Not Voodoo Magic

A Primer

Lena M. Mantle, PE

Senior Analyst

Office of the Public Counsel

January 2026

Glossary of Terms

Annualization: adjustment to account for one year (365 days) of revenue or costs.

Base usage: usage that is not affected by fluctuations in weather. Also known as non-weather sensitive usage.

Billing determinants: the key factors like usage volume, peak demand, time-of-use periods, and fixed service charges, used by utilities to calculate bills, with including customer counts, and usage (kWh, CCF, gallons).

Billing cycle: a reoccurring 28 to 32 day time period over which usage is measured and recorded.

Billing month: The month that bills are provided, e.g. mailed, emailed, to customers. A utility's billing month consists of 20 to 22 billing cycles with different beginning and ending read dates.

Cost-of-Service: Normalized, annualized cost necessary to provide service to customers. Also known as revenue requirement.

Non-weather sensitive usage: usage that is not impacted by changes in the weather. Also known as base usage.

Normalization: adjustment to a cost or revenue for abnormal circumstances.

Revenue: Billing determinants multiplied by rates.

Revenue requirement: The total amount of revenue necessary to cover normal and annualized cost of the utility plus a return on the utilities investment. It is also known as the cost-of-service.

Unexplained usage: The usage that is not explained by regression coefficients. It is calculated as the difference between the actual usage and the usage predicted by the regression line.

Weather sensitive usage: usage that changes with changes in the weather.

Introduction

This principle is important for utilities and regulators in determining the accuracy of revenue collected and requested as well as setting rates to generate revenue to match the utility's cost to serve.

The objective of a rate case is to set rates to recover the normal costs of the utility over a year. Costs are reviewed and, if necessary, annualized[1] and normalized[2] to derive an appropriate estimate of the normal annual costs of the utility to provide safe and adequate service to customers at just and reasonable rates. If costs are normalized but revenue is not, then the assumption made is that the revenues the utility billed in the test year are representative of the revenue the utility will receive going forward. However, this is likely to create a mismatch between costs and revenue.

It is common knowledge that electricity and natural gas usage varies according to the weather and that weather is never "normal." When the weather is cold, heaters are turned on. If it is a mild winter, then usage was lower in that winter than it was in the previously colder winters. When the weather is hot, air conditioning is used. For a mild summer, the air conditioning usage is much less.

The objective of a rate case is to determine rates that will recover the normalized cost to serve customers. Revenue is the rates multiplied by billing determinants. The most basic billing determinants for utilities are customer numbers and usage. Since customer counts do not change with the weather, that means that the fluctuations in revenues due to weather are due to fluctuations in usage. It is important to normalize usage to calculate normalized revenue given current rates to know how much rates need to increase. Normalized billing determinants are also used to determine the rates once the new revenue requirement is known.

[1] Annualization adjustments are required when costs in the test year do not reflect of a full year of the cost or revenue.

[2] Normalization adjustments are required when the test year costs or revenues are abnormal.

Since rates are determined by dividing the revenue requirement by the billing determinants (usage of customers), it is important that the usage also be normalized for the extreme weather. For example, the non-fuel commodity revenues a natural gas utility incurs in a very cold test year are not indicative of the revenue the utility would see on an ongoing basis due to the higher than normal usage. While the cost of the natural gas commodity is passed directly passed through to customers, natural gas utilities typically have a volumetric charge to recover some of its non-fuel costs.[3] Excessive cold causes ratepayers to expend more natural gas to heat their homes, which results in the natural gas utility collecting more than normal revenues through its non-fuel volumetric charge. As a result, the test year revenues are normalized, *e.g.* reduced, to account for this higher usage. If revenue requirement was set using this higher billing, the rates will not recover the revenue requirement on a normal or warmer year the revenue, not because it is managed poorly, but because the billing determinants from a cold winter were used to set rates.

If usage for that extremely cold year is not normalized, to determine rates on a going forward basis, the lower "normalized" revenue requirement would be divided by high usage from the extremely cold test year resulting in lower rates and reducing the likelihood that the utility would earn a fair return going forward.

This paper provides the basics of weather normalization. While the methods change slightly from electric to natural gas, the principles of weather normalization are the same.

Usage – Basic Data Concepts

As described above, the billing determinant of usage is impacted by weather. Usage is typically supplied as the usage of a combined group of customers of a similar nature, *e.g.* a customer class, for a billing cycle or billing month. Customer counts usually do not vary by daily fluctuations in weather. Usage per customer does. To assure that only usage is being weather normalized, weather normalization analysis should therefore be conducted on usage per customer. Customer numbers should be annualized separately and applied to the normalized usage per

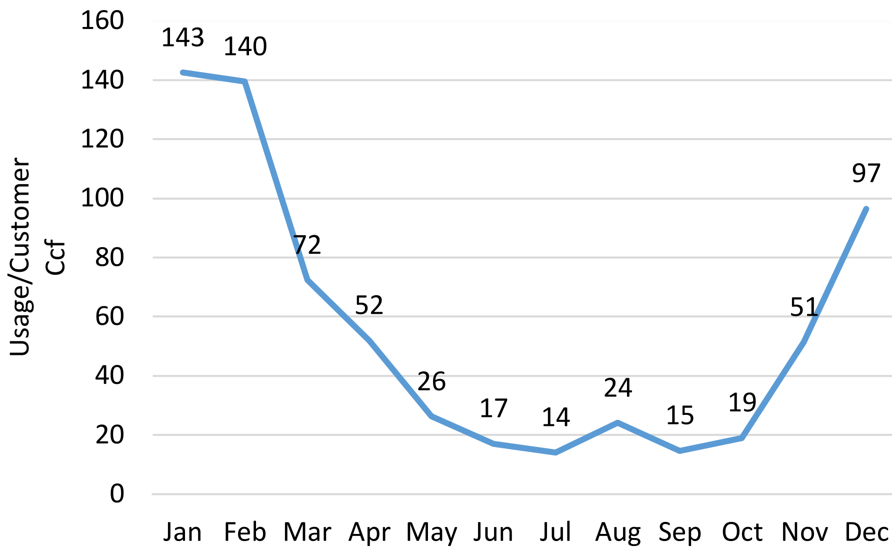
[3] The theory is that then larger users pay more of the non-fuel costs.

customer to determine normalized, annualized revenues to be used to determine normalized revenue of current rates and to set rates to achieve the total revenue requirement going forward.

As in any analysis, the results of the analysis are only as good as the data being analyzed. Therefore, the first step in weather normalization, as with any analysis, should always be an examination of the usage and weather data being used. While visual examination of the data can catch some problems such as negative usage, graphing the data is also a helpful tool to find data problems. For example, Chart 1 below of natural gas usage per residential customer by month shows that it is likely that there is a problem with the August usage per customer data.

Chart 1

Residential Billing Month Usage/Customer



In this graph, the August billing month usage of 24 Ccf stands out from the usage of the months around it. The magnitude of usage per customer for the months of January through July look realistic with the usage being the highest in January (which it typically the coldest month) with usage declining through the month of June with July's usage being a bit lower and about the same as September. But the usage of August looks to be high. It is unlikely that this representative of actual usage because there is very little, if any space heating occurring in June through October. The expectation is that the usage in these months would be stable (base usage) – most only water heater and cooking usage and not affected by fluctuations in weather. It is expected that usage starts increasing in November and December when it gets cold again.

The usage data of 24 Ccf for August should be investigated because it is inconsistent with the usages of 14 Ccf and 15 Ccf in July and September respectively. There could be several reasons for incorrect usage data mostly tied to the billing system. The most likely reason is that there was a large correction for incorrect billing readings that occurred in previous months. It could also be an erroneous bill that was corrected in a future month not in this test year. In either of these instances, the usage data for more than one month may need to be corrected.

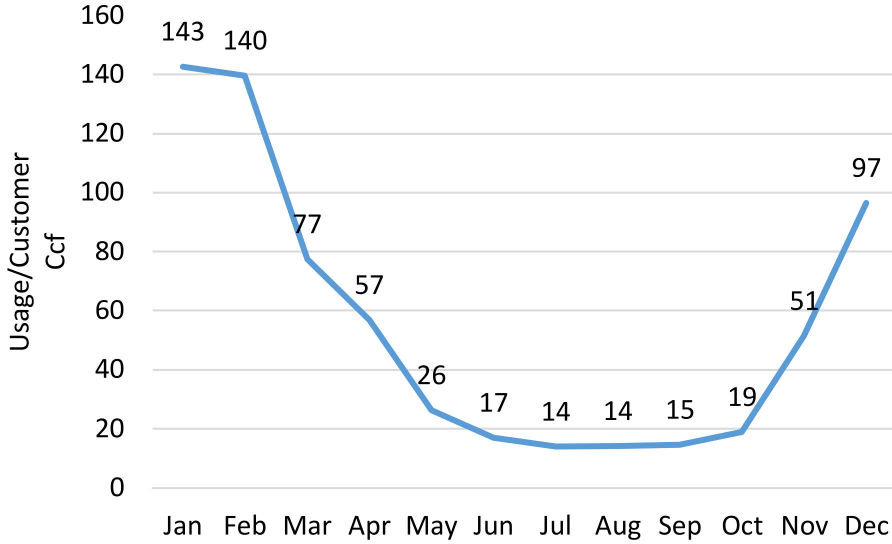
When suspicious data is found, then the data needs to be reviewed in a less aggregated form, *e.g*. billing cycle level or customer level, to see if any problem can be found. In the example above, it was determined that the usage in the March and April billing months had been misread and corrected by increasing the usage in the August billing month. To correct the data, the usage that was included in August for March and April's errors were moved to the correct months. The total usage for the year did not change. The usage was just moved to the month that it actually occurred in. The original and corrected monthly usages are shown in Table 1 below.

Table 1
Test Year Billing Month Usage

	Usage (Ccf)	
	Provided	Corrected
Jan	143	143
Feb	140	140
Mar	72	77
Apr	52	57
May	26	26
Jun	17	17
Jul	14	14
Aug	24	14
Sep	15	15
Oct	19	19
Nov	51	51
Dec	97	97
Total	670	670

Chart 2 below shows the corrected data plotted across time.

Chart 2
Residential Billing Month Usage/Customer



This graph passes the sanity check in that the months with no need for space heating have consistent usage. While there may still be problems with the data, due to the aggregation of data, there are no obvious problems in this usage data.

While it may seem odd to "correct" data, it is important to remember what the objective of the analysis is. The objective of this analysis is not to normalize the billing system.[4] The objective of the analysis is to get an estimate of how this group of customers respond to changes in weather. If a data base has many data points that need to be corrected and corrections are not made, analysis of the data will give inaccurate results. However, the correction of an obvious error such as this will increase the accuracy of the estimated response to weather.[5]

Measure of Weather

The typical measure of weather used depends on the type of usage that is being modeled. Both natural gas utilities and electric utilities use Heating Degree Days ("HDD") as a measure of

[4] This does not mean that the billing system problems should not be corrected. However, the corrections should occur as annualization adjustments outside of the weather normalization analysis.
[5] Later in this paper, there is a comparison of the results of the model with and without the data correction.

weather for explaining heating usage. It is calculated by first determining the average mean daily temperature ("MDT") by taking the daily high temperature, adding the daily low temperature and dividing this sum by 2. This is shown in the equation below.

$$\text{MDT} = \frac{\text{High Temperature} + \text{Low Temperature}}{2}$$

If the MDT is greater than or equal to 65 then HDD is zero. If the MDT is 64 or below, then HDD is 65 minus the MDT. This HDD is calculated with a base of 65. When HDD is used as a weather measure for weather normalization either for natural gas or electric usage, the analyst is assuming that the response to cold weather, *i.e.* heating response, begins when the MDT is less than 65 degrees.

A measure of warm or hot weather is Cooling Degree Days ("CDD"). CDD are calculated similar to HDD in that if the MDT is less than or equal to 65 then CDD is zero. If the MDT is 66 or above, then CDD is the MDT minus 65. This CDD is calculated with a base of 65. If used a measure of weather to model electric response to weather, the analyst is assuming that the response to hot weather, *i.e.*, cooling, begins when the MDT is greater than 65.

The base of 65 is not set in stone. HDD and CDD with a base of 65 is the measure of weather developed and published by the National Oceanic and Atmospheric Administration ("NOAA") National Weather Service ("NWS") for numerous sites around the United States of America because at the time the NWS started calculating degree days, it was a logical base.

The NWS provides daily high and low temperatures for many weather stations allowing the calculation of HDD and CDD with different bases.[6] Good analysts will test the usage data against measures of weather with different bases to determine what base is appropriate for the data that is being modeled. For example, the electrical usage of residential customers often has a deadband range where there is no cooling or heating instead of just one MDT of 65. While a base of 65 may be appropriate for cooling, a base of 58 may be more appropriate for heating for

[6] If the temperatures used in the analysis are not from an NWS Class A station, the daily high and low temperatures need to be reviewed for accuracy and missing data.

residential customers.[7] Commercial buildings often turn on their air conditioners long before the MDT is 65 due to the prevalence of heat generating computers and copy machines. A base as low as a 60 MDT may fit the usage better for cooling usage with heating beginning when the MDT is 58.[8]

Matching Weather to Usage

The usage provided above is for a billing month which consists of 21 billing cycles each typically covering 28 to 32 days. Attachment A provides a diagram of a September billing month with 20 billing cycles. The customers in billing cycle 1 are billed for usage from August 1 to August 31. Even though this billing cycle was for usage in August, it is considered part of the September billing month because the bill was provided to the customer in September. The customers in billing cycle 20 are billed for usage from August 27 through September 27.

The chart on the bottom of the attachment graphs the actual and normal CDD for each billing cycle. For the first billing cycle, the actual CDD of 358[9] were well below the normal of 444. For the last billing cycle, the actual CDD was 323 which was above the normal of 253.

The actual CDD for the September billing month billing cycles ranges from 386 to 295. The billing month average CDD for the September billing months shown in Atachment A is 331 CDD – the average of the CDD of the 20 billing cycles. The differences in this example of the billing month and calendar month actual and normal CDD are provided in Table 2 below.

---

[7] See Chart 5 later in this paper.
[8] See Chart 6 later in this paper.
[9] These were calculated by summing the actual and normal CDD for each day of the billing cycle. The daily actuals and normal are provided below the diagram. The sum for each billing cycle is shown on the right side of the diagram.

Table 2
Difference Between CDD of Calendar and Billing Months
From Example on Attachment A

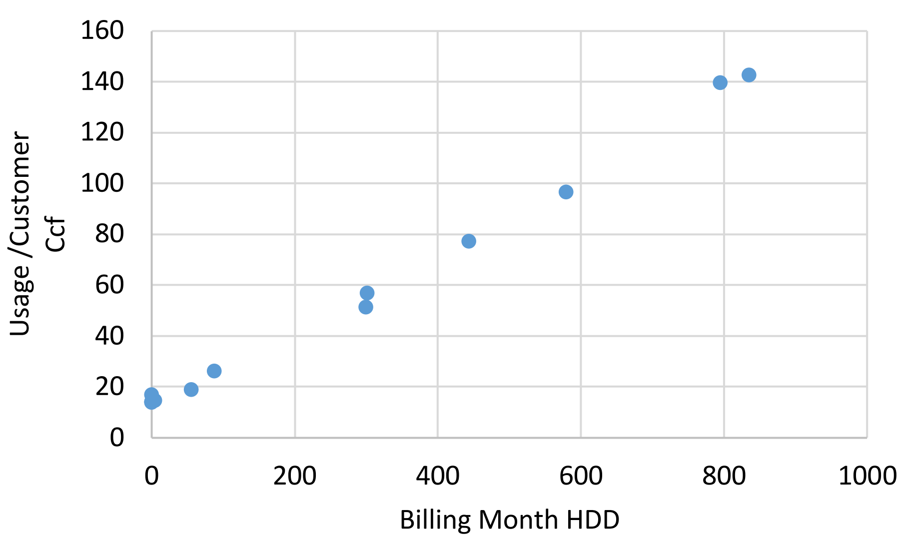
	Calendar Month	Billing Month
Actual CDD	307	331
Normal CDD	228	340
Weather Adj	-79	10

In this example, it was warmer than normal in the calendar month and cooler than normal in the billing month. This attachment demonstrates the mismatch between billing month usage data and calendar month weather data.

Check for Relationship

Intuitively, we know that natural gas and electric usage is affected by daily fluctuations in temperature. However, a check of the graphical representation of the weather to the usage is always appropriate and informative. Chart 3 below shows the usage data from Chart 2 plotted against the corresponding measure of weather for the time period the usage was tracked over.

Chart 3
HDD base 65 vs. Residential Usage/Customer



The data points in this graph lie in a tight, increasing line indicating a correlation between usage per customer and billing month HDD. The data points at the zero (0) billing month HDD indicate usage/customer of about 14 Ccf for months with no heating degree days.

Basic Analysis Concepts

The objective of weather normalization is to provide a measure of the usage that would have occurred under normal weather conditions. In its simplest form, this means estimating the change in usage associated with a change in weather, *i.e.* how much does usage change when the mean daily temperature changes by a degree.

Chart 3 above shows a linear relationship between HDD and usage/customer. Linear regression analysis is a statistical tool that provides an equation for a line that best fits a relationship between two variables – one of which is modeled as dependent upon the value of another, independent variable. Regressions by themselves only reveal relationships between two variables. Regressions analysis does not prove a relationship. In this instance, it is common knowledge that usage is dependent on the weather. Therefore, regression analysis is a tool that can be used to provide a measure of how much usage changes for the group of customers being modeled due to a change in weather.

Chart 3 shows that the relationship between this usage and weather data is linear. Consider the following common linear equation:

$$y = a + bx + \varepsilon$$

Where:

- y = dependent variable
- a = value of the dependent variable when the independent variable is zero (intercept)
- b = change in the dependent variable attributable to an increment in change in the independent variable (slope)
- x = independent variable
- ε = error

This simple linear equation can be used to model the effect of weather on usage. Where:

y = usage

a = non-weather sensitive usage

b = $\frac{\text{change in load}}{\text{change in weather}}$

x = measure of weather

ε = unexplained usage

The values of "a" and "b" are found through regression analysis of the weather (independent variable) against usage (dependent variable). These values are different for different types of customers, *e.g.* residential, small commercial, large commercial, and each utility. They also change across time as the base usage and weather sensitive usage changes. The unexplained usage ε is different each data point. It is the portion of the actual usage that cannot be explained by HDD.

For the data above, the resulting equation is:

$$y = 12.49 + 0.153x$$ [10]

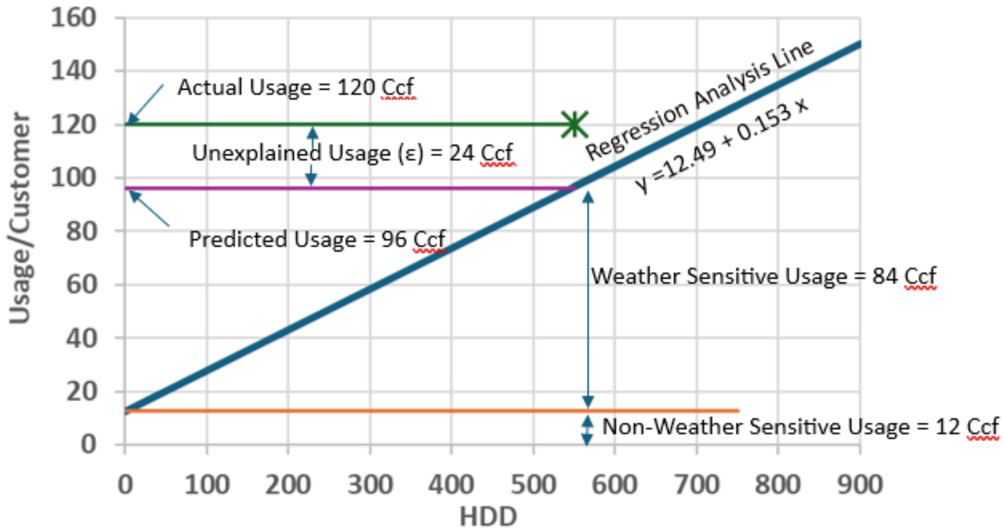
This equation predicts that, given the data in Chart 3 above, the usage when HDD is zero[11] is 12.49 Ccf and weather sensitive usage is 0.153 Ccf for each heating degree day change. Chart 4 below is a graphical representation of the information provided from this regression analysis and a data point with usage of 120 and a billing month HDD of 550.

---

[10] The regression minimizes the difference between the data points and the estimated line. The unexplained (ε) for all data points sum to zero.

[11] Commonly referred to as non-weather sensitive usage.

Chart 4
Components of Regression Results



The line labeled "Regression Analysis Line" shows the predicted billing month usage per customer for each of the HDDs on the horizontal axis. The actual usage (120 Ccf) versus the actual billing month HDD (550) is the asterisk on the graph. One of the first things to notice is that the actual usage is not on the regression line. The difference between the regression line and the actual usage (ε) is the usage that is not explained by the regression line. It is calculated as the difference between the observed data point of 120 Ccf at 550 HDD and 96 Ccf - the amount that the line predicts at 550.

The point where the regression line where HDD equals zero is the billing month the non-weather sensitive usage. This regression estimates that regardless of the magnitude of the HDD, the non-weather sensitive usage per customer is 12.49 Ccf. The difference between the regression line and the non-weather usage of 12.49 Ccf is the weather sensitive usage. The weather sensitive usage increases as HDD increases.

Weather Normalized Usage

The regression line provides a predicted amount for each HDD. The objective of weather normalization is to estimate what the usage would have been had normal weather occurred assuming all other factors, including the non-weather sensitive usage and the unexplained usage, stayed the same.

Finding the predicted usage for normal weather can be accomplished by inserting "normal" HDDs in the equation. However, this is not a weather normalized value. It is predicted value given normal weather. This calculated value does not take into account the actual usage incurred that cannot be explained by the regression line which is designated as ε.

A weather normalization adjustment is calculated as the difference between "normal" and "actual" weather. The formula to calculate the adjustment is:

$$\text{WN adjustment} = b \times (HDD_{Normal} - HDD_{Actual}).$$

Using the regression equation above, the weather normalization adjustment for each data point is:

$$\text{WN adjustment} = 0.153 \times (HDD_{Normal} - HDD_{Actual}).$$

Assuming that the normal HDD was 650 and the actual HDD was 550, the weather adjustment is calculated as:

$$\text{WN adjustment} = 0.153\ \text{Ccf/HDD} \times (650\ \text{HDD} - 550\ \text{HDD}) = 15.3\ \text{Ccf}$$

The weather normalized usage per customer is calculated as:

$$\text{Usage/Customer}_{Normalized} = \text{Usage/Customer}_{Actual} + \text{WN Adjustment}.$$

Using the actual usage of 120 Ccf in Chart 4 above and assuming that normal HDD is 650, the weather normalized usage for that data point is:

$$\begin{aligned}\text{Usage/Customer}_{Normalized} &= 120\ \text{Ccf} + 0.153\ \text{Ccf/HDD} \times (650\ \text{HDD} - 550\ \text{HDD}) \\ &= 120\ \text{Ccf} + 15.3\ \text{Ccf} \\ &= 135.3\ \text{Ccf}.\end{aligned}$$

Fit of Line

Linear regression minimizes the difference between a straight line and the data points. An often referred to statistical measure is the R^2. It is a measure of how much of the variation in the dependent variable, *i.e.* usage, can be explained by the independent variable, *i.e.* HDD. The closer R^2 is to one, the better fit of the data. The R^2 of the regression line above is 0.9943. While this is an excellent R^2, it does not mean that it is the best fit line.

An examination of the data shows that the average of the usage in billing months with zero HDD[12] is 15 Ccf. The regression analysis gives a lower base usage of 12.49 Ccf. While this does not seem significant, when multiplied by 100,000 customers it could be.[13] One option is to subtract the lowest billing month usage off of each of the billing month's usage and rerunning the regression but specifying that the intercept has to be zero.

This regression line gives a slightly lower weather response of 0.1488 per HDD. The R^2 of this line is also excellent at 0.9959. A weather adjustment for a 100 HDD difference using this regression, results in an adjustment of 14.88 Ccf where the previous regression estimated 15.3 Ccf.[14] Again, not a big difference but across 100,000 customers, it could be significant.

Both equations have good R^2. This is where the skill and knowledge of the analyst impact the results and determine the best fit line. Additional analysis may show that the R^2 increases if the base of HDD is moved from 65 to 62. The art of weather normalization is knowing how an analysis can be improved and when the result is good enough for the application.

Normal Weather

The normal MDT, HDD and CDD published by the NWS is calculated on thirty years of weather data. It has been the practice of the NWS to only update its measures of normal weather every ten years. The current normals are based on daily weather recorded for the years of 1991 through 2020. The next set of normals will not be calculated until after 2030. Before calculating

---

[12] June, July, and August billing months.
[13] For the current Spire West summer rate of $0.55916/Ccf, this would equate to $138,452 for 100,000 customers.
[14] For the current Spire West winter rate of $0.60966/Ccf, this would equate to $25,606 for 100,000 customers for a month with a difference between normal and actual HDD of 100.

daily normal, the NWS carefully reviews the daily high and low temperature data over the thirty years. Changes in measuring instruments are noted, and the data is adjusted if it shows that new instruments measured different from the old. Changes in the position of where the measurement instrument are also noted and modifications made to make the data consistent if the position of the instrument impacted the recorded data.

NWS makes no modification for trends in the weather, *e.g.* the weather getting warmer or colder. Therefore, before using the NWS normal, the weather history needs to be examined. If the data does show a trend in the weather, then there are a couple of ways to minimize the impact of the trend. First of all, more recent actual data can be used to create normal weather. This makes the trend in the weather end where the actual begins. However, to be accurate, the daily weather data needs to be examined for changes due to measurement instruments or placement of the instrument. If an inconsistency is found, the daily weather history then needs to be adjusted to be consistent with the actual weather being used. Then this adjusted history can then be used to calculate normal weather variables.

The other way to minimize a trend in the weather is to use less than thirty years of data. However, the statistical law of large numbers states that the greater number of years the more likely the results are accurate. A general rule of thumb is that the sample needs to be thirty or more. This is why the NWS uses a thirty-year history and is a concern with using a shorter time period than thirty years.

There are also different ways to calculate normal weather. One way is to take the MDT on January 1 for all of the years of the history and calculate an average of these values. Then the same is done for January 2 and so forth for each day of the year.

Another way to calculate daily normal weather is to rank the daily weather in each year – coldest to hottest MDT. Then calculate an average for each rank. This means that the coldest normal day is an average of all the coldest day in each year of the history used. The second coldest is an average of all the second coldest days, and so forth until there is a normal for each rank. The problem then becomes how to assign these 365 averages back to the days. One alternative is for the averages for each of the ranks be assigned to a day based on the rank of the weather in the test year. This minimizes the difference between normal and actual weather on each day. Once these ranked MDT have been assigned to a day, then HDD and CDD can be calculated with a base determined by the rate class that is being normalized.

Characteristics of a Good Weather Normalization Analysis

The analysis is only as good as the data. Garbage in, garbage out. The most important characteristic of a good weather normalization analysis is good data – good usage data, good actual weather data, and good weather history from which to calculate normal weather variables. Table 3 below compares the results of a linear regression of the initial data described earlier in this paper and the results using the corrected data.

Table 3
Comparison of Results

	Non-Weather Sensitive Usage	Weather Response	R^2
Initial Data	13.46	0.1496	0.9853
Corrected Data	12.49	0.1530	0.9943
Data with base usage removed	15.00	0.1488	0.9959

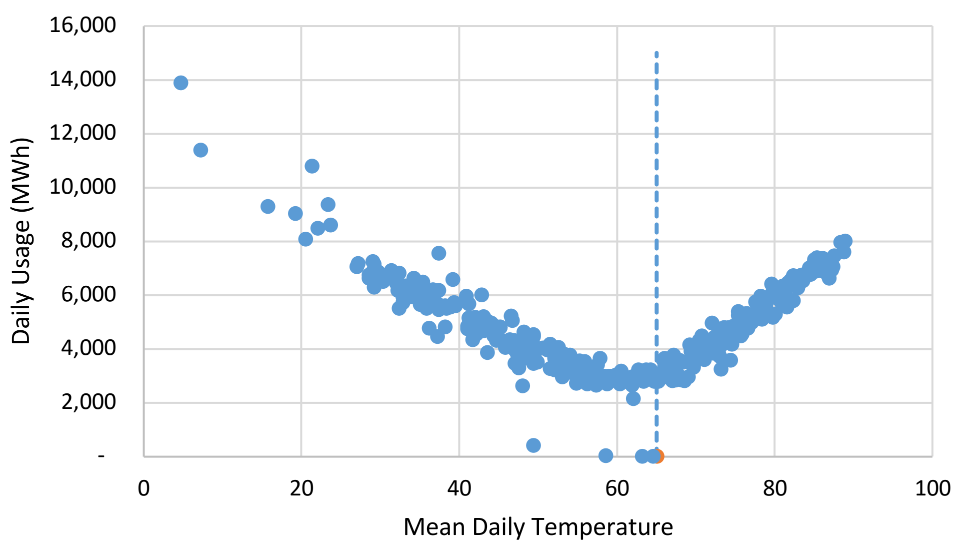
If a quick regression had be done on the initial data without looking at the data and only the R^2 was considered, an inexperienced analyst could have stopped since an R. However, as compared to the results of the analysis on the corrected data, the weather response from the initial data was as much as with the corrected data, i.e. a weather adjustment for 100 HDD would have been 14.96 Ccf when the correct result was 15.3 Ccf. Subtracting the average base usage, gives the best R^2 and the smallest response to weather. Again, over 100,000 customers, these differences could be significant. It also shows how the results can be manipulated by different modeling.

Impact Of Having More Data Points

Also important in regression analysis is the number of data points on which the analysis is conducted. Larger sample sizes increase the statistical power of the analysis, making it easier to detect the true relationship between weather and usage. The analysis of usage in this paper was conducted on twelve billing months – 12 data points. This is the bare minimum for weather normalizing usage for a year. These billing months were actually the sum of 21 billing cycles of data for each month. That means 252 data points (21 billing cycles over 12 months) were aggregated to 12 data points. Information was lost in this aggregation. A greater number of data points (252) would produce a more accurate result. With computers, regression analysis with 252 data points is not a difficult task.

The most accurate way to measure how a class of customers react to daily fluctuations in weather is to plot daily usage against daily MDT. With AMI meters, this information can be available.[15] For example, plotting electric utility residential class daily data against weather typically shows that residential response to weather is not linear. Chart 5 below is a graph of 365 days of a residential class usage of an electric utility plotted against MDT.

Chart 5
Residential Class Daily Usage vs. Daily MDT



An examination of the low data points shows a few data points where the usage is close to zero. An analysis with knowledge of the data knows that this is bad data. The residential usage would never fall to almost zero. If correct data can not be obtained, the analyst should remove these

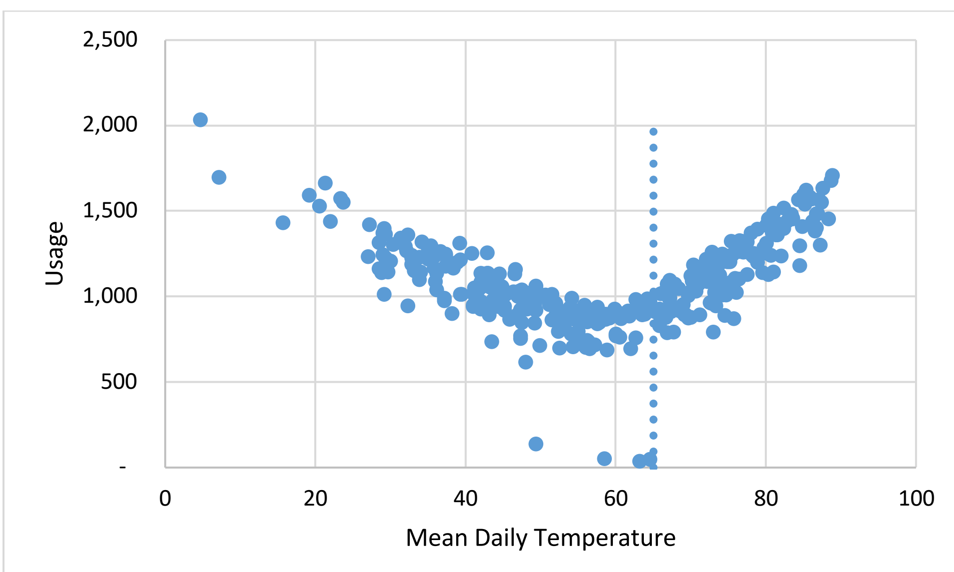
[15] Electric utilities have been estimating daily class loads for decades using the hourly loads of a sample of customers. This data is referred to as load research data.

data points because, again, the objective of the analysis is to measure response to weather and there will still be plenty of data with these points removed.

This chart shows that the residential response at this electric utility is not a "V" shape. It is more a flattened "U" shape with a range of MDT where there seems to be no response to weather. The dashed vertical line is the MDT of 65. What this graph shows is that with this customer class, response to hot weather (cooling) may start when the MDT is 63 or 64 instead of 65. The response to cool weather (heating) does not start until the MDT is below 58. Using HDD and CDD with a base of 65 as weather measures would assume that there is only one MDT with no response to weather – 65. This review of the data shows how this is not true for this set of customers. A better base for CDD would 63 and a better base for HDD would be 58. Regression analysis would show exactly the better base for both heating and cooling for this residential class.

Chart 6 below is a similar graph for the Small General Service Class.

Chart 6
Small General Service Daily Usage vs. Daily MDT



There are also points that are far below most of the data points that are bad data that should be removed from the analysis. The scatter of this usage is broader. Putting further detail in the graph that usage on weekend days is lower than the usage on weekdays but the response to weather is very similar.

This chart also shows that the response for this class is closer to a V than the residential class. It also shows that cooling response starts more at a MDT of 60 and heating starts at a MDT of 58. Using HDD with a base of 65 is not appropriate for this small general service class. Analysis should be conducted to determine the appropriate base for this class.

While daily data for classes for the electric utilities has been available as load research data for decades that can be utilized to measure the response to daily fluctuations in weather, this same information has not been available for gas utilities. However, with the advanced meter infrastructure meters ("AMI") that have been installed by many gas and water utilities, this information is now available for these utility types too allowing these utilities the ability to examine how their customer classes respond to weather.

When daily usage is not available, multiple years of monthly data may seem like a way to have more data points. However, using multiple years of monthly data can can be problematic because there are other factors over this longer period of time that can impact usage such as a recession or a pandemic. Long-term trends of general increases or decreases in usage also impact the measurement of the impact of weather on usage when more than one year of data is used. Some of this can be accounted for with trend or other regression variables. This is where the expertise of the analyst and a careful review of the data makes a difference.

A Caution

This methodology estimates the weather sensitivity of a class or group of customers. While the analysis is conducted on a usage per customer basis, it does not necessarily measure the weather sensitivity of any one customer. Each customer responds to weather differently based on their lifestyles and end-uses. If the weather sensitivity of a class of customers is used to determine an adjustment to an individual's usage, that customer's usage is being adjusted by the average response of the class. This will increase the usage of the customers who have weatherized their homes to be energy efficient. It will decrease the usage of customers who have not. Therefore, the use of a weather sensitivity adjustment of a class to customers on an individual basis results in a subsidization by customers of energy efficient dwellings of those whose dwellings have not been weatherized.

For an accurate weather adjustment to individual customer's usage, an analysis should be conducted on each customer's usage to capture the customer's usage habits and appliance efficiencies.

Normalizing Water Usage

The correct measure of weather to use for water usages is less straightforward. Water usage varies based on the precipitation. When it is dry, lawns and gardens are watered. Water usage is affected by both rainfall and temperature, but it does not respond as directly to daily weather natural gas usage and electric usage do. While both historical rainfall and temperatures values exist to match to usage and create normal weather variables, typically CDD is used as the weather variable in weather normalization analysis of water usage.

Conclusion

As the title to this paper state, weather normalization of energy usage is not voodoo magic. It is, however, statistical analysis informed by the analyst. The statistical analysis can be a simple straight-line regression analysis or it can be a complex daily model that takes into account time of the year and different weather response shapes.

Perhaps the most critical part of weather normalizing usage is that bad data will give bad results. Just as blood cannot be squeezed out of a turnip, the most skilled and experienced analyst cannot get accurate results from bad data. Significant effort should be used to review both the usage data and the weather data.

Weather normalization of usage in a utility rate case is important. It takes time and effort. Like most data analysis, weather and usage data can be manipulated to get different answers. This paper describes the most important aspects of weather normalization, not to encourage manipulation of the data but to present the most important aspects of a good weather normalization analysis.